Parallel adaptations to an altitudinal gradient persist in tetraploid snowtrout (Cyprinidae: *Schizothorax*), despite extensive genomic exchange within adjacent Himalayan rivers

Tyler K. Chafin[1,2], Binod Regmi[1,3], Marlis R. Douglas[1], David R. Edds[4], Karma Wangchuk[1,5], Sonam Dorji[5], Pema Norbu[5], Sangay Norbu[5], Changlu Changlu[5], Gopal Prasad Khanal[5], Singye Tshering[5], and Michael E. Douglas[1]

[1]Department of Biological Sciences, University of Arkansas, Fayetteville, Arkansas USA 72701

[2]Department of Ecology and Evolutionary Biology, University of Colorado, Boulder USA 80309

[3]National Institute of Arthritis, Musculoskeletal & Skin Diseases (NIAMS), National Institutes of Health, Bethesda, Maryland, Maryland USA 20892

[4]Department of Biological Sciences, Emporia State University, Emporia, Kansas USA 66801

[5]National Research and Development Centre for Riverine & Lake Fisheries, Ministry of Agriculture & Forests, Royal Government of Bhutan, Haa Bhutan

Keywords: Adaptation, ecotype, convergent evolution, parallel evolution, polyploidy, introgression

Disclosure statement: Authors have nothing to disclose.

(200/200) Replicated evolutionary patterns are often attributed to recurrent emergence following parallel selective pressures. However, similar genetic patterns (e.g., ‘genomic islands’) can also emerge following extensive homogenization in secondary contact, as a by-product of heterogeneous introgression. For example, within Himalayan tributaries of the Ganges/Brahmaputra rivers, drainage-specific mtDNA clades of polyploid snowtrout (Cyprinidae: *Schizothorax*) are partitioned as co-occurring morphological ‘ecotypes,’ hypothesized to represent parallel divergence among adjacent streams. To evaluate this scenario, we utilized a reduced-representation genomic approach (*N*=35,319 *de-novo* and *N*=10,884 transcriptome-aligned SNPs) applied to high-altitude Nepali/Bhutanese snowtrout (*N*=48 each). We unambiguously quantified ploidy levels by first deriving genome-wide allelic depths followed by ploidy-aware Bayesian models that produced genotypes statistically consistent with diploid/tetraploid expectations. When genotyped SNPs were clustering within drainages, the convergence of eco-phenotypes was sustained. However, subsequent partitioned analyses of phylogeny and population admixture instead identified subsets of loci under selection which retained genealogical concordance with morphology, with apparent patterns of parallel ecotype emergence instead driven by widespread genomic homogenization. Here, prior isolation is effectively ‘masked’ by admixture occurring in secondary contact. We note two salient factors:1) Polyploidy has promoted homogenization in tetraploid Himalayan snowtrout; and 2) Homogenization varies across Himalayan tributaries, presumably in lockstep with extent of anthropogenic modification.

1. Introduction

Selection for local environmental conditions can drive rapid evolution, and occasionally does so within parallel systems [1]. This, in turn, can promote evolution of unique ecotypes, or even novel species [2–4]. Adaptations to ecological gradients can also occur, especially in the absence of geographic isolation, and are commonly attributed to selection acting upon existing variation [either ancestral/standing variation [5], or that acquired through hybridization [6,7]]. These represent but two of several scenarios through which such patterns can be generated [8–10].

Divergent selection in response to ecological gradients can also facilitate speciation when gene flow is ongoing [11–13]. A hallmark of this process is the formation of genomic 'islands of divergence,' with selection and its cumulative effects serving as a counterbalance for homogenizing gene flow [14,15]. These so-called 'islands' may then expand via a hitchhiking mechanism, such that divergence is also initiated within linked genomic regions [16–19]. However, heterogeneous genomic divergence can also arise via entirely different processes, to include those wholly unrelated to speciation. This, in turn, introduces an analytical dilemma, in that the signatures of one can either obfuscate or instead reflect that of the other [20–22].

The resulting emergence of parallel ecotypes is often manifested phylogenetically as clusters within study sites or regions, rather than among eco-phenotypes [23]. However, an alternative is that ecological adaptations instead evolve via isolation followed by subsequent dispersal, such that genes are now juxtaposed across both distributions and genomes [8,24,25]. 'Genomic islands' as well as phylogenetic patterns are then generated which are similar to those manifested by parallel divergence-with-gene-flow (per above). This can occur, for example, when introgression is constrained within those localized genomic regions that are associated with adaptation to an underlying ecological gradient [26]. The result is genome-wide homogenization,

with ancestral branching patterns (e.g., those uniting ecotypes) now restricted to genomic regions whose permeability to gene flow is reduced by selection and/or low recombination [27,28]. However, the singular origin of an adaptive allele spread selectively via localized introgression can also yield relatively similar genomic patterns [29].

Together, this implicates four distinct scenarios that could generate genomic landscapes compatible with those expected under parallel ecotype emergence. Yet only two of these necessitate divergence-with-gene-flow, e.g., with selection acting in a repetitive fashion on either 1) standing genetic variation or 2) independent emergence via *de novo* mutation.

In contrast, phylogenetic discord is produced in the latter two scenarios by the accumulation of adaptive variability while in isolation, subsequently followed by secondary gene flow among dispersing ecotypes. This then results in: 3) selective filtering against a background of genomic homogenization; or 4) selective introgression of adaptive alleles into novel populations. Thus, while all scenarios are effectively operating in ‘parallel’ (e.g., in separate river drainages or habitat patches), their relationship to the speciation process differ substantially.

(a) Can parallel emergence be discriminated from parallel introgression?

We posit it is possible for these factors to be discriminated by predicting the ancestries for replicated “pairs” of ecotypes distributed among sites (although patterns may also depend upon migration rates among populations [30]). For hypotheses of ‘independent origin,’ we would predict that genealogies with regions encapsulating targets of divergent selection [i.e., scenarios (1) or (2)], would lack a shared ancestry among ecotypes. In a similar vein, there would be a failure to correlate with unlinked targets of selection [8].

Thus, in the case of repeated selection upon standing genetic variation, adaptive alleles

will be identical-by-descent, but with flanking regions excluded in that populations will lack identical, independently-fixed haplotypes [9,10]. Further, in a scenario of independent *de novo* emergence, mutations underlying adaptation may occur at different loci altogether.

By contrast, a "divergence first" model with subsequent secondary contact [scenarios (3) or (4)] would imply that ancestries are both correlated with and shared among ecotypes at those loci targeted for selection. However, genomic patterns can be difficult to disentangle from those expected under a scenario of adaptive introgression (scenario 4), in that gene flow during secondary contact may effectively 'swamp' divergence developed in isolation [8,31]. In the absence of genomic resources (a frequent scenario for non-model organisms), biogeographic context can often be used to provide additional clarification.

(b) A case study involving *Schizothorax*

We here investigate the source of adaptive genetic variation among differentially adapted pairs of rheophilic freshwater fishes occupying altitudinal gradients in Himalayan tributaries of the Ganges and Brahmaputra rivers. Our study group (snowtrout; Cyprinidae: *Schizothorax* spp.) is of interest in that it displays a legacy of whole-genome duplication (WGD), with variable adaptations to high altitude habitat (both phenotypic and life historic) [32–35]. Yet, patterns of convergence in previous studies are also broadly apparent [36,37]. Additionally, the rapid formation of multiple independent 'species flocks' has occurred within unique and isolated alpine lake habitat [38–42], with an emerging consensus indicating recurrent ecotypic formation along riverine gradients [38,43]. Here, two Himalayan species are of particular interest: *Schizothorax progastus* and *S. richardsonii*. Both are distributed along an elevational gradient, with generalized morphologies within tributaries of the Ganges and Brahmaputra rivers

converging upon ‘blunt-nosed’ (*S. richardsonii*) in upstream reaches, and ‘pointed-nosed’ (*S. progastus*) in downstream reaches [39,44–46].

Here, the ‘pointed-nosed’ eco-phenotype displays a more terete, streamlined shape, and gradually replaces the ‘blunt-nosed’ variety longitudinally along an altitudinal gradient, a pattern seemingly replicated in each of several collateral, southward-flowing Nepali tributaries, suggesting the presence of ecological non-exchangeability [47]. Recent phylogenetic analyses based on mtDNA also failed to support the nominal species, with relationships occurring instead at the drainage level [43]. Despite this, morphologies are consistent across drainages [38], suggesting the independent emergence of ‘blunt-nosed’ phenotypes (i.e., those associated with *S. richardsonii*) within each highland drainage.

Li and Guo [48] hypothesized a role for ploidy in modulating macroevolutionary trends in cyprinid fishes, with a rapid, positive shift in net diversification rates occurring within three predominantly polyploid subfamilies: Torinae, Schizopygopsinae, and Schizothoracinae, with the most profound effect in the latter. Of particular interest, all three subfamilies are exclusively within the Himalayan and Qinghai-Tibetan Plateau (QTP) regions, and generally display specializations for life at elevated altitudes [49]. Many historic whole genome duplications (WGDs) in plants also overlap with periods of dramatic global change [50–52]. Thus, diversification of polyploid cyprinids on the QTP and adjacent regions is seemingly linked to the QTP uplift followed by periods of marked climate change [48,53–55].

However, gauging the manner by which polyploidy has driven evolutionary patterns, such as those potentially involved in the adaptation by *Schizothorax* to elevational gradients, is currently limited by population genomic methods. Studies involving non-model polyploid species such as snowtrout must employ methodological paradigms that assume diploidy that

employ fundamentally divergent theoretical expectations [56,57]. An additional complication involves the varying degrees of divergence found among ohnologs (i.e., gene duplicates originating from WGD) that potentially occur in those species at intermediate stages of re-diploidization. One positive is that models suitable for genotyping polyploid or mixed-ploidy data [58,59] now incorporate short-read sub-genomic methods (e.g., RADseq and related methods).

There are two stumbling blocks with regard to the identification of genomic adaptations to elevation in *Schizothorax*: Drivers of morphological and phylogenetic discordance are not only numerous, but also constrained by polyploidy. Here, we combine novel statistical models and robust genotyping of non-model polyploids with expectations regarding the manner by which ancestries should be distributed across the genome. In so doing, we demonstrate that high-altitude adaptations can be maintained in isolated populations via parallel selection, despite the deletion of historic ancestry across a large portion the genome due to introgression. In this sense, the latter serves to essentially camouflage parallel divergence previously derived. Our results also emphasize the importance of alternative hypotheses when considering the origin of genomic 'islands of divergence.'

2. Methods

(a) Sampling and DNA preparation

Tissue samples (*N*=96) represent six *Schizothorax* species distributed throughout tributaries of the Brahmaputra and Ganges rivers (Bhutan and Nepal, respectively) (figure 1*a-c*). Of these, *N*=48 Nepali specimens were obtained from the University of Kansas Natural History Museum (KUNHM) (figure 1*b;* table S1) and represent two riverine species ['high elevation' *S.*

richardsonii (*N*=16) and 'low elevation' *S. progastus* (*N*=8)]. These were sampled from each of the three major Nepali drainages (Kali Gandaki, Koshi, and Karnali), with the exception of *S. progastus* from Karnali. The remaining Nepali samples (*N*=24) represent a land-locked lacustrine radiation endemic to Lake Rara [40]. These are: *S. macrophthalmus* (*N*=8), *S. raraensis* (*N*=8), and *S. nepalensis* (*N*=8).

Bhutanese *Schizothorax* (*N*=48) were sampled from three major Bhutanese drainages (Wang Chhu, Punatsang Chhu, and Mangde Chhu: figure 1*c*; table S1) and identified as *S. progastus* and *S. richardsonii* based upon morphological diagnoses of vouchered specimens [39]. However, subsequent mtDNA sequence analysis has putatively identified Bhutanese *S. richardsonii* as another convergent high-altitude specialist, *S. oconnori* [43], a species typically found on the Qinghai-Tibetan Plateau (QTP [60,61]) within high elevation tributaries of the upper Brahmaputra River [herein referred to as the Yarlung-Tsangpo River (YLTR)]. They are subsequently referred to herein as *S. cf. oconnori* (figure 1*c*).

Tissues were processed using the Qiagen DNeasy Blood and Tissue Kit (Qiagen, Inc.), following manufacturer's protocols. Extracts were evaluated for presence of high-molecular weight DNA using gel electrophoresis (2% agarose) and quantified at 2µl per sample in 200µl assays using Qubit broad-range DNA fluorometry assays (Thermo Fisher Scientific). DNA (500-1000ng of DNA per sample in 50µl reactions) was then fragmented using a restriction double-digest [62] with *Pst*I (5'-CTGCAG-3') and MspI (5'-CCGG-3'). Digests were subsequently visualized on 2% agarose gels, purified using AMPure XP beads (Beckman Coulter, Inc.), and again quantified via Qubit fluorometer.

Samples were then standardized at 100ng of digested DNA and ligated in 30µl reactions using T4 DNA ligase (New England Biolabs, Inc.) following manufacturer's protocols. Barcoded

oligonucleotide adapters were designed and employed following Peterson *et al.* [62]. After a second AMPure XP purification, samples were multiplexed in groups of *N*=48 and size-selected at 350-400bp (excluding adapters), using a Pippin Prep automated system (Sage Sciences). Adapters for Illumina sequencing were subsequently extended via a 10-cycle PCR with Phusion high-fidelity DNA polymerase (New England Biolabs, Inc.). Final reactions were purified via AMPure XP beads and standardized per submission requirements of the DNA Core Facility (University of Oregon Genomics & Cell Characterization Facility, Eugene, OR U.S.A.). Additional quality control at the core facility included fragment size analysis (to confirm successful amplification and the absence of artifacts) and qPCR (to assess the proportion of sequenceable library). Sequencing pooled two *N*=48 multiplexed libraries into a single lane of 1x100 sequencing on the Illumina HiSeq 4000.

(b) Data filtering and ploidy-aware assembly

Raw reads were demultiplexed and filtered for initial assembly using IPYRAD [63]. Reads having >0 barcode mismatches or >5 low-quality bases were discarded, and adapter sequences trimmed using the 'strict' option. Two assemblies were performed: one *de novo* (at an 85% identity threshold), and one using bwa to align against assembled transcriptome data [64]. Because the genotyping model in IPYRAD assumes diploidy [65], we used relaxed settings so as to retain assembled paralogs (e.g., electing for more stringent filtering following ploidy-aware genotyping; see below). These included a minimum depth threshold of only 6, allowing 20 heterozygous sites per consensus locus, with up to 4 alleles at a given locus within an individual. This yielded *N*=48,350 and 14,301 SNPs for *de novo* and transcriptome-guided assemblies, respectively.

To explore potential ploidy variation in our samples, we first employed BWA [66] to realign raw reads against the candidate locus catalog generated in our *de novo* assembly. For ploidy model selection, we then computed allelic read depths for bi-allelic sites using a de-noising procedure (NQUIRE; [67] with expectations that ploidies represent allelic depths as follows: At ~50% in a diploid bi-allelic heterozygote (i.e., an approximate 50:50 representation of each allele in an AB heterozygote); at ~33% or ~66% for triploids (=ABB or AAB possible genotypes); and ~25%, 50%, or 75% for tetraploids (=ABBB, AABB, AAAB genotypes, respectively). Log-likelihoods of the observed allelic depth distributions were extracted for Gaussian expectations under each fixed model (e.g., 2n, 3n, or 4n), then normalized by that under a model of freely variable allele depths ($=log\mathrm{L}_{\mathrm{Fixed}}/log\mathrm{L}_{\mathrm{Free}}$), with chosen models representing the greatest normalized log-likelihood. Results were employed to generate prior expectations for subsequent ploidy-aware genotyping and downstream filtering.

Formatted SNPs (as .vcf) were genotyped for both *de novo* and transcriptome-guided assemblies using the Bayesian variant calling pipeline in POLYRAD [58]. We first computed per-locus $H_{\mathrm{IND}}/H_{\mathrm{E}}$ values, a statistic representing the probability of sampling reads from two different alleles at a given locus in an individual [59]. We then simulated expected $H_{\mathrm{IND}}/H_{\mathrm{E}}$ values given the sample size and read depth distribution from the observed data, under expectations of either diploidy or tetraploidy, so as to define threshold values under which markers are statistically consistent with Mendelian behavior as a 95% upper-bound of the simulated distribution. Because expected values differ by ploidy, we used $H_{\mathrm{IND}}/H_{\mathrm{E}}$ values averaged across loci to segregate those statistically consistent with a diploid genotype model from those consistent with a tetraploid model.

(c) Geometric morphometrics of Nepali fishes

We first examined patterns of morphological convergence and divergence using geometric morphometric data [38] derived from 528 images captured from museum specimens representing all five Nepali species. Our analyses were restricted to the Nepali sub-tree in that voucher specimens for Bhutanese samples were not available. Briefly, 18 landmarks were targeted, focusing on head morphology (i.e., snout elongation, head depth, eye/nostril placement), and body shape (origin/insertion of fins, upper/lower bases of the caudal fin, and urostyle as most posterior point). Coordinates were then superimposed across samples using Generalized Procrustes Analysis implemented in GEOMORPH [68]. Full details on geometric morphometric data processing, including bias and sensitivity analysis, can be found in Regmi *et al.* [38].

Procrustes-aligned coordinates were summarized using a principal component analysis (PCA) in GEOMORPH [68]. We then conducted a linear discriminant analysis (LDA) using the MASS R package [69], maximizing discriminant capacity with 80% of samples as a training set, and group classifications specified as species X drainage.

(d) Phylogenetic relationships

Given that most phylogenomic methods cannot take into account polyploid SNP genotypes [70], our initial phylogenies were instead built using an alignment-free method that computes distances from the intersection of *k*-mer distributions in unaligned reads [71,72]. Matrices of *k*-mer distances were then used to infer a phylogenetic tree with branch lengths using FASTME [73].

The resulting topology was additionally contrasted with that inferred under a polymorphism-aware (POMO) model inferred in IQ-TREE [74,75], as evaluated across 1,000

ultrafast bootstraps, with a GTR substitution model, gamma-distributed rates (*N*=4 categories), and a virtual population size of 19.

(e) Population structure and molecular clustering

Most widely used clustering or 'assignment test' methods that examine population structure are either inappropriate for polyploid data or cannot handle mixed-ploidy genotypes [56]. However, STRUCTURE is robust in both situations [76,77], though more computationally intensive than alternatives. We thus coded our mixed-ploidy genotypes as input for STRUCTURE, following recommendations of Meirmans *et al.* [56] and Stift *et al.* [77], using 10 replicates each of *K* values (= # sub-populations) ranging from *K*=1 to *K*=20. Analyses were additionally replicated across *de novo* and transcriptome-guided assemblies, using a total MCMC length of 100,000 iterations following a 50,000 iteration burn-in. Input files were generated using *polyVCFtoStructure.py* (github.com/tkchafin/polyrad_scripts), and results were aggregated/visualized using the CLUMPAK pipeline [78], with selection of optimal *K* following the Evanno method [79]. Final visualizations aligned the ancestry proportion bar plots aligned to our SKMER phylogeny using code from Martin *et al.* [80].

Results were contrasted with ordination via discriminant analysis of principal components (DAPC), performed in ADEGENET [81,82]. Samples were stratified according to species X basin assignment (e.g., *S. progastus* from Wang Chhu), and with analyses performed in three ways: 1) globally; 2) Nepal-only; and 3) Bhutan-only. The number of principal components (PCs) retained in each case was determined using the *xvalDapc* function as a cross-validation approach, with the optimal number of the root-mean-square-error (RMSE) of assignment derived from a 10% subset (with the remaining 90% serving as a training set). This

was accomplished across 30 replicates per level of PC retention, up to a maximum of 300 retained PCs, resulting in 20 (globally-), 10 (Bhutanese-), and 5 (Nepali-) retained PCs.

(f) Modeling population mixture

We assessed hybridization within Nepali and Bhutanese sub-trees using TREEMIX [83], with a global search across numbers of migration events (*m*) ranging from 0–5. Because markers are assumed diploid, migration analyses were only performed on markers statistically-fitting diploid expectations, as designated by POLYRAD [58]. Optimal values of *m* for each subtree were determined from rates of log-likelihood change as computed in OPTM [84], which generates an *ad hoc* statistic (ΔM) representing the second-order rate of change weighted by the standard deviation (e.g., the 'Evanno' method; Evanno et al. 2005). Independent replicates on the full dataset yielded identical likelihoods for some *m*-values (i.e., yielding an undefined ΔM) and, given this, we assessed variability using 100 bootstrap pseudoreplicates per migration model. To additionally discriminate among divergence scenarios among Nepali *S. richardsonii* and *S. progastus*, we calculated a 4-taxon Patterson's *D* statistic and admixture fractions (f_4-ratio) [85,86]. Both are formulated to test an enrichment of shared-derived site patterns between either component of a lineage pair (P1 or P2), and a third lineage (P3) relative to an outgroup. The assumed phylogenetic structure would be: (((P1, P2), P3), P4). Our interest was in shared patterns, both at the level of "conspecifics" among drainages and "heterospecifics" within drainage, and our tests were conducted such that P1 and P2 were defined by river [e.g., (*rich*X, *prog*X), *rich*Y), Outgroup), where X and Y represent different drainages], and where P1 and P2 were defined per taxon-assignment [e.g., (*rich*X, *rich*Y), *prog*X), Outgroup)]. The outgroup in all cases was *S. cf. oconnori* from Bhutan.

(g) Testing patterns of selection and locus-wise differentiation

Here we sought loci strongly associating with axes of population differentiation, and therefore employed the program PCADAPT [87] to identify loci putatively associated with differentiation of *S. richardsonii* and *S. progastus*. Detecting loci under selection is a persistent methodological gap in those studies examining polyploid SNP data [56] and, given this, we restricted our analysis to transcriptome-mapped loci that could appropriately be genotyped as diploid [58]. Analyses were performed by partitioning the data so as to target specific divergence events: *S. richardsonii* X *S. progastus* (Gandaki); *S. richardsonii* X *S. progastus* (Koshi); *S. richardsonii* X Lake Rara (*S. macrophthalmus*, *S. raraensis*, and *S. nepalensis*); *S. nepalensis* X *S. macrophthalmus*+*raraensis*; and *S. progastus* (Bhutan) X *S. cf. oconnori*. Loci were additionally restricted to those with a minor allele frequency (MAF) >0.05, with significance assessed using α=0.01 adjusted via Bonferroni correction (where *N* tests = *N* SNPs). Transcripts encapsulating outlier SNPs were additionally annotated with gene ontology (GO) terms for biological functions using the BLAST2GO pipeline [88], searching against the SWISS-PROT and InterPro databases [89,90]. Redundancy was reduced in GO-term annotations by measuring semantic similarity using RRVGO [91] and the *Danio rerio* organismal database.

We additionally contrasted population pairs (delineated as above), using a ploidy-appropriate F_{ST} measure . We calculated gene diversity as: $H_s = 1 - \Sigma p_i^2$, where p_i is the frequency of allele *i* [92]. Although in diploids this measure is often referred to as 'expected heterozygosity,' it does not have the same relationship with heterozygosity across ploidies [56]. Finally, we calculated genetic distance as Jost's *D* [93], as well as absolute distance (D_{XY}) for each SNP based on allele frequencies and summarized across loci as an arithmetic mean [94].

(h) Partitioned analysis across subsets of loci

As a final test of the hypothesis that divergence between pairs of *S. richardsonii-S. progastus* represents isolation prior to secondary hybridization, we performed a partitioned analysis across subsets of our data. Our prediction was as follows: If differential adaptation reflects a retention of minor genomic regions from secondary introgression (i.e., as opposed to arising independently), then disparate populations of *S. richardsonii* would be more similar at loci for which drainage pairs (*richardsonii-progastus*) are highly diverged. We diagnosed such loci in two ways: First, we computed a ratio of D_{XY} between *richardsonii-progastus* pairs within Koshi and Gandaki rivers, and D_{XY} of *S. richardsonii* populations among drainages. Here, a value less than 1.0 indicates a locus for which *richardsonii-progastus* pairs are more similar than *richardsonii-richardsonii* comparisons. A value >1.0 indicates a locus for which *richardsonii-progastus* are more highly diverged. If loci with a D_{XY} ratio >1.0 originated from hybridization, genetic relationships among those loci should reflect a conspecific phylogeny. To assess the latter, we replicated TREEMIX analyses (as above) for loci >1.0 and <1.0. As a secondary test, we also partitioned locus-wise *richardsonii-progastus* F_{ST} into four groups, each receiving the same analytical treatment as the D_{XY} ratio partitions: 1) F_{ST} = 0.0-0.249; 2) 0.25-0.49; 3) 0.50-0.749; and 4) 0.75-1.0.

3. Results

(a) Sequence assembly and genotyping

Allele depths for biallelic sites were overwhelmingly trimodal, with NQUIRE likelihoods suggesting tetraploidy across all study taxa (figure 2). Raw outputs from *de novo* and

transcriptome-aligned assemblies resulted in 48,350 and 14,301 loci, respectively (table S2). Three samples failed to yield genotype data for >=50% of loci and were removed. Simulated expectations in POLYRAD yielded H_{IND}/H_E thresholds of 0.52 (diploid) and 0.65 (tetraploid). Using these thresholds and custom Python code (github.com/tkchafin/polyrad_scripts/filterPolyVF.py), we partitioned genotypes for both assembled datasets into loci statistically consistent with tetraploid and diploid expectations. An additional requirement was that >50% of individuals be genotyped at an average mean per-individual depth of 20X. Post-filtering datasets consisted of 5,601 (2n) and 5,284 (4n) transcriptome-mapped, and 15,547 (2n) and 19,772 *de novo* SNPs (table S2). For each assembly method, these were divided into files containing all loci as well as subsets containing only those statistically consistent with a diploid model. Scripts for formatting datafiles and outputs can be found at: github.com/tkchafin/polyrad_scripts.

(b) Contrasting molecular and geometric morphometric results

Morphological DAPC showed a marked convergence of *S. richardsonii* and *S. progastus* body shapes, regardless of origin, with Lake Rara taxa (*S. macropthalmus*, *S. nepalensis*, and *S. raraensis*) having weak or no differentiation (figure 3). Among the latter, *S. nepalensis* was most intermediate between a 'Lake Rara' cluster, and that for *S. richardsonii* (and to a lesser extent, *S. progastus*). This contrasted sharply with the clustering of Nepali specimens based solely on ddRAD data (figure 4*a*,*c*), where the predominant relationship was by drainage (e.g., with Koshi and Gandaki *S. progastus* and *S. richardsonii* grouping together).

The latter was mirrored in our phylogenetic analyses, with pairs of *S. progastus*-*richardsonii* from the Koshi and Gandaki rivers being monophyletic in the alignment-free

phylogeny (figure 5*a*) and agreeing approximately with STRUCTURE-inferred assignment probabilities (figure 5*b*). Optimal Evanno-derived *K* in the latter identified peaks in *DeltaK* at *K*=3 and *K*=7. The former yielded a homogenous Nepali sub-tree, with only the Koshi River samples somewhat distinct (figure 5*a*,*b*), whereas *K*=7 reveals weak differentiation, with mixed assignment spanning basins.

Inferred edges in TREEMIX suggested migration between *S. progastus* of the Koshi and Gandaki rivers, as well as between *S. progastus* of the Koshi with *S. raraensis* and *S. macrophthalmus* of Lake Rara (figure 5*c*). However, the latter may be misascribed, particularly given the absence of sampling for *S. progastus* from the Karnali River. Furthermore, migration edges in TREEMIX may also explain the mixed inter-drainage assignment seen in STRUCTURE, as well as the non-monophyly of Koshi River *richardsonii-progastus* in our POMO analysis (figure S1).

Bhutanese *Schizothorax* spp. differed in that no clear delineations were found between high- and low-elevation clades in *S. progastus*, and with clustering instead reflecting species-level assignments (figure 4, 5*b*). STRUCTURE results suggested mixed assignment of *S. progastus* and *S. cf. oconnori*, particularly in the Wang Chhu where both *de novo* and transcriptome-guided results agreed (figure 5*b*). TREEMIX results also revealed evidence in multiple sub-basins for exchange across species boundaries (figure 5*c* and S2). Of note, mixed assignment was also observed between *S. progastus* (Bhutan) with those from Nepal (figure 5*c*). We could not test whether this reflects retained ancestral variation (i.e., prior to divergence of Ganges and Brahmaputra rivers), or more contemporary mixture. Patterson's *D*-statistic likewise supported this mixed ancestry, both at the level of conspecifics among drainages and heterospecifics within drainages (table S3).

(c) Locus-wise differentiation and outlier analysis

The greatest number of significant SNPs following Bonferroni-correction occurred in outlier analysis involving Koshi and Gandaki *S. richardsonii-progastus* pairs, and from the comparison of *S. progastus* and *S. cf. oconnori* from Bhutan (figure S3*a*). Of these, the most substantial number of overlapping SNPs was found among Koshi and Gandaki comparisons. Outliers in both cases had higher F_{ST} than did the genome-wide distribution involving *de novo* or transcriptome-guided assemblies. This pattern was not repeated in the same loci for Bhutan (figure S3*b*), although outlier loci in Bhutan did share a negative relationship of hene diversity (H_E), as compared to *S. richardsonii-progastus* F_{ST} (figure S3*c*). Outlier F_{ST} for *S. richardsonii-progastus* was correlated in the Gandaki *versus* Koshi rivers, but with no discernible relationship with the same measure when compared to Karnali River *S. richardsonii* and Lake Rara species, Bhutanese *S. progastus-oconnori*, nor among Bhutanese populations of only *S. progastus* (figure S4). Relationships among different locus-differentiation statistics otherwise occurred per theoretical expectations (figure S5). Hence only comparisons (above) are presented for F_{ST}. Relating D_{XY} ratios to D_{XO} showed that high values were likely driven by near-zero distances among *S. richardsonii* pairs for a subset of loci (figure S6), in line with evidence from the *D*-statistics indicating shared inter-drainage ancestry among conspecifics (table S3).

Although ~75% of outlier transcripts could not be assigned candidate protein products, gene ontology (GO) term enrichment for shared outliers revealed a number of related biological processes, including anatomical structure formation (GO: 0048646), up-regulation of phosphoprotein phosphatase activity (GO: 0032516), and up-regulation of G protein coupled receptor signaling (GO: 0045745)(figure S3*d*). The few outlier transcripts that could be assigned

candidate proteins in the overlapping Gandaki-Koshi set were: Aurora Kinase A (*AURKA*), Zinc Finger SWIM Domain Containig Protein 6 (*ZSWIM6*), Adhesion G protein-coupled receptor A3 (*ADGRA3*), and Caspase B (*CASPB*).

(d) Partitioned ancestry and mixture analyses

Loci for which *richardsonii-richardsonii* divergence was greater than *richardsonii-progastus* (i.e., D_{XY} ratio < 1.0; figure 6*a*) occurred in drainages independent of ecomorphological taxon-assignment and corresponded to ~68%. The remaining ~32% yielded groupings consistent with species-level taxonomy, but with migration among species suggested within the Gandaki and Koshi rivers (figure 6*a*). Repeating the same analysis across loci binned by F_{ST} yielded a transition from drainage-level at low values (with migration among mid-highland *S. progastus*) towards species-level at high values (with migration within drainages) (figure 6*b*). This pattern was consistent when loci were binned with F_{ST} estimates derived either from *S. richardsonii-progastus* comparisons in the Gandaki or the Koshi, although the transition towards species-level grouping was slightly protracted in the latter (figure 6*b*).

4. Discussion

Using genome-wide SNP data and ploidy-aware genotyping, we showed that contrasting patterns of genetic and morphological differentiation in *Schizothorax* species from southern-slope Himalayan tributaries point to parallel selection against secondary gene flow, rather than to the parallel emergence of altitudinal forms. Highland specialists in Nepal converged strongly onto a singular phenotype, with this relationship recapitulated in phylogenetic patterns occurring at loci under selection. However, neutral variation showed signs of homogenization among altitudinal

pairs (*S. richardsonii* and *S. progastus*) within each basin, suggesting a breakdown of reproductive isolation. In Bhutan, altitudinal pairs (*S. progastus* and *S. cf. oconnori*) showed signs of low-level mixture, suggesting either greater reproductive fidelity or that altitudinal pairs are at an early stage of homogenization via secondary gene flow.

(a) Parallel divergence versus parallel hybridization in Nepali snowtrout

Genome-wide relationships among Nepali altitudinal pairs (*S. richardsonii* and *S. progastus*) give the appearance of recurrently emerging ecotypes, with repeated colonization and subsequent adaptation to highland habitats from a mid-highland progenitor occurring independently in each drainage (figure 4, 5). This seemingly fits the narrative of whole-genome duplications (WGDs) as promoting the adaptive potential of schizothoracine fishes, as well as the broader pattern of convergence among high-altitude specialists [37,48]. Indeed, WGD-derived adaptive potential may have played a role in the parallel colonization of highland habitats by *S. richardsonii* in the Ganges tributaries of Nepal, and *S. cf. oconnori* in the Brahmaputra tributaries of Bhutan, although we note that the current study lacks geometric morphometric data for the latter.

In the case of altitudinal *Schizothorax* pairs in Nepal, we argue that the shift of dominant ancestries from grouping at the drainage-level (independent of eco-morphology) at lowly differentiated/“neutral” loci towards the conspecific-level (with gene flow within drainages) at strongly differentiated/selected loci suggests instead a singular origin of adaptive alleles in the highland specialists, with subsequent homogenization of much of the genome (figure 7). However, this raises an additional question: How are adaptive phenotypes retained despite a majority of the genome having rampant introgression? Here, one potential is that an allopolyploid origin of WGD in *Schizothorax* may cause a conflated signal of hybridization.

However, Xiao *et al.* [95] found a lack of sub-genome divergence in a study of *S. oconnori*, as would be expected under a scenario of allopolyploidy.

It is also possible that polyploidy facilitates a greater degree of genomic exchange than may be typical of reproductively isolated diploid species [96]. For example, ploidy has been suggested to promote introgression in diploid-tetraploid crosses of *Arabidopsis*, due to a circumvention of dosage-mediated postzygotic isolation [97–99]. Increased introgression has also been supported among tetraploid-tetraploid crosses (as herein), with one potential mechanism being an increase in local recombination rates, and thus a relaxation of linkage as a component of purifying selection, as ploidy increases [100,101]. Similarly, increased dosage may 'mask' deleterious loads, especially in young polyploid species [102,103]. Given a young estimated age of WGD in *S. oconnori* [95], it is possible that these predictions manifest as a vastly more porous genomic landscape than might be typical in their diploid counterparts [99]. This may also be an underappreciated mechanism promoting the hypothesized increased adaptability of polyploids to stressful or novel environments [104,105].

(b) Varying rates of hybridization among drainages

In contrast to the substantially permeable genomic landscape seen among *S. richardsonii* and *S. progastus* pairs in Nepal, we saw a much smaller signature of gene flow spanning altitudinal forms in Bhutan (*S. progastus* and *S. cf. oconnori*). In addition to gene flow obscuring species boundaries [80,106], parallels drawn between Himalayan streams and another region characterized by extreme freshwater habitats (e.g., the American Southwest [106–108]) suggest that the potential for anthropogenically mediated hybridization should not be ignored. Here, 'extinction vortices' may be driven by a coupling of potentially maladaptive hybridization and

declining population sizes. This creates a concerning backdrop against which to contrast climatic expectations of shrinking habitats for vulnerable highland species such as *S. richardsonii* [109].

If altitudinal pairs in Nepal are indeed on such a trajectory (e.g., culminating in extinction via hybridization [110,111]), then the relatively lower (albeit present) levels of interspecific gene flow in Bhutan may be an indicator of further homogenization. Such a scenario is reasonable, especially given a contrast to the degree of hydrologic modification in the two countries. For example, while both countries are rich in hydropower potential, but there is a greater power demand and larger number of dams in Nepal [112]. Given the potential for seasonal migratory behavior in *Schizothorax* [61,113], dams may disrupt movement of species, inadvertently promoting interspecific contact. Here, we again emphasize parallels with the heavily modified and regulated rivers of the American Southwest, where water usage policy and damming promotes hybridization [107,114], defines habitat suitability [115], and alters environmental cues [116]. Though Bhutan has a high percentage of protected areas, our results suggest that anthropogenically-mediated hybridization may be one additional dimension to consider when balancing freshwater conservation planning as hydropower development continues to expand in both countries [112,117].

(c) High-altitude adaptation

Though our results do not support parallel independent divergence within altitudinal gradients within Nepali rivers, we do note that convergent colonization of highland habitats is both apparent and prevalent within schizothoracine fishes [36], to include *S. richardsonii* and *S. cf. oconnori* examined herein. Here, elevational distributions of putative ‘high’ and ‘low’ forms of Bhutanese and Nepali *Schizothorax* were concordant (figure 1*d-e*), save *S. cf. oconnori* from the

lower Mangde Chhu near the confluence with the Dangme Chhu (figure 1*c*, *e*). Of note, the latter were collected during February–May, a period that overlaps with elevated springtime flows during the pre-monsoon/early monsoon season in the eastern Himalaya, a period when co-occurring cyprinid fishes such Golden mahseer (*Tor putitora*) migrate to upstream spawning sites [118]. The remainder of our Bhutanese samples were collected during December/early January. Thus, the discrepancy with regard to the collection of lower Mangde Chhu *S. cf. oconnori* may represent a response to seasonal flow variation and thus reflect migratory behavior (Qadri and Yousuf 1983; Mohan 2005).

Though we found relatively little overlap in putative targets of selection (figure S3*a*), we note that our reduced-representation approach only surveyed a small percentage of transcripts, with a poor frequency of successful annotation (e.g., with the majority of transcripts not being assigned gene ontology terms) further obscuring results. Molecular convergence among high-altitude populations has been observed in disparate human populations [120], and even among humans and their domesticates [121]. Other studies in *Schizothorax* have shown an array of candidate genes displaying signatures of adaptation relating to hypoxic and ultraviolet conditions associated with elevation [122]. Though we cannot corroborate these results with our reduced-representation approach, we can speculate with regards to the morphological underpinnings, in that these showed strong fidelity among conspecifics, even in drainages with high levels of introgression.

Schizothoracine fishes in general have been characterized as positioning similarly to trout species (Salmonidae) in torrential flows, with muscular bodies and cylindrical shapes facilitating position maintenance near the stream bed in rapid currents [123]. A longitudinal survey of fish assemblages in Nepal showed a replacement of *S. richardsonii* by *S. progastus* at decreasing

altitudes [47], suggesting that factors associated with elevation (e.g., dissolved oxygen, flow rates [124]) play a role in driving ecological non-exchangeability of *Schizothorax* in altitudinal gradients. Other studies of Himalayan taxa have likewise found turnover and narrow ranges associated with elevation, suggesting it as a general driver of metacommunity structure [125,126]. The specialized sucker-like adaptation present in *S. richardsonii* but absent in *S. progastus* [47], and the higher species richness of *Schizothorax* species at mid-elevation [127], seemingly corroborates altitude (or variables associated with altitude) as a constraint driving adaptation in *Schizothorax*.

An interesting case of adaptation in *Schizothorax* which we did not explore fully herein is that of the putative ecological radiation of species in Lake Rara. Though we found species therein to overlap in body shape (figure 2), they exhibit differences in mouth morphology, gill raker shape, spawning microhabitat choice, and diet (with *S. raraensis* insectivorous, *S. nepalensis* herbivorous, and *S. macrophthalmus* planktivorous) [40]. As with previous studies [41,43], we found them genetically indistinguishable across several analyses, suggesting the nature of this radiation is either recent and rapid, or reflects plasticity. Because of this, we suggest that further analyses be done with a larger percentage of the genome (e.g., using a skimming approach), from which more robust genotype-phenotype associations, if they indeed exist, could be assessed [128]. Such a whole-genomic study may also be more explanatory with regards to the potential of polyploidy in supplying variation necessary for adaptive radiation [105,129].

5. Conclusions

We show that, in light of recent methodological advancements, population genomics in non-

model polyploid species can be done in a statistically appropriate and analytically rigorous manner, with relatively minor adjustments to a traditional molecular ecology workflow. However, several limitations (such as the lack of ploidy-aware methods for outlier detection) were encountered, signaling a need for further development. Despite these limitations, we were able to disentangle the nature of parallel evolution in *Schizothorax* species in Himalayan tributaries of the Ganges and Brahmaputra rivers, and with implications of recurrent processes at differing timescales: convergent adaptation towards high-altitude environments among major drainages (e.g., Ganges *versus* Brahmaputra), and parallel selection against introgressive homogenization within drainages. With regards to the latter, we found patterns suggesting heterogeneous levels of 'progression' towards potential merger of populations, with species boundaries in Bhutan being potentially more stable. This points to the potential for environmental instability or anthropogenic modification in driving hybridization in Himalayan drainages, though we suggest that more work is needed.

Ethics. All methods were performed in accordance with relevant guidelines and regulations. Bhutanese collecting permits were in conjunction with the National Research & Development Centre for Riverine and Lake Fisheries (NRDCR&LF), Ministry of Agriculture & Forests (MoAF), Royal Government of Bhutan. The export of fin clips was authorized through a Material Transfer Agreement (MTA) provided by the National Biodiversity Centre (NBC), with additional approvals provided by the Department of Forests & Park Services (DoFPs) and Bhutan Agricultural and Food Regulatory Authority (BAFRA). Sampling protocols were approved by the University of Arkansas Institutional Animal Care and Use Committee (UA_IACUC_17064). The study was also carried out in compliance with ARRIVE guidelines

(https://arriveguidelines.org).

Data and code accessibility. Raw sequence data will be made available via the NCBI Short Read Archive (SRA) pending acceptance, under BioProject PRJA########. Assembled datasets are available in the Open Science Framework repository (doi: 10.17605/OSF.IO/TMJFK). All custom codes used to for the manipulation of data files are available on GitHub (github.com/tkchafin/polyrad_scripts and github.com/tkchafin/scripts). Additional analytical or accessory scripts are available via the Open Science Framework (doi: 10.17605/OSF.IO/TMJFK).

Authors' contributions. All contributed to conceptualization and study design. D.R.E. collected samples in Nepal. S. T., K. W., S. D., P. N., S. N., C. L., and G. P. K. coordinated and supervised field work in Bhutan. S .T., K.W., M.E.D., M.R.D, S. D., P. N., S. N., C. L., G. P. K., and T.K.C. collected samples in Bhutan. B.R. collected and curated geometric morphometric data. M.E.D. and M.R.D. generated molecular data. T.K.C. wrote necessary codes and performed bioinformatic work. B.R. and T.K.C. analyzed data. T.K.C. generated figures and drafted the manuscript. All authors contributed to revising the final product and subsequently approved its submission.

Competing interests. We declare we have no competing interests.

Funding. The National Research Centre for Riverine and Lake Fisheries (NRCRLF, Bhutan) provided logistic support and personnel with which to sample fishes. Analytical resources were

provided by the Arkansas Economic Development Commission (Arkansas Settlement Proceeds Act of 2000) and the Arkansas High Performance Computing Center (AHPCC). Computational support was also provided by the U.S. National Science Foundation (NSF) funded XSEDE Jetstream cloud (award #TG-BIO200074 to T.K.C.). This research was made possible through generous endowments to the University of Arkansas: The Bruker Professorship in Life Sciences (M.R.D.), the 21st Century Chair in Global Change Biology (M.E.D.), Distinguished Doctoral Fellowship award (T.K.C.), and Graduate Teaching Assistantships (K.W., T.K.C.), all of which provided salaries and/or research funds to complete this study. T. K. C. received additional support via an NSF Postdoctoral Fellowship in Biology under Grant No. DBI: 2010774. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the funding agencies nor affiliated organizations.

Acknowledgments. Museum specimens and tissue samples for Nepal were obtained from the University of Kansas Natural History Museum (KUNHM; https://biodiversity.ku.edu/ichthyology/collections). We thank Curator Leo Smith and Collection Manager Andrew Bentley for tissue loans and use of imaging facilities at KUNHM. Nepali samples were originally collected by D.R.E., as funded by a Fulbright Scholarship, the Explorers Club, and the National Geographic Society (NGEO) Committee for Research.

References

1. Schluter D, Nagel LM. 1995 Parallel speciation by natural selection. *Am. Nat.* **146**, 292–301.
2. Schluter D, Conte GL. 2009 Genetics and ecological speciation. *Proc. Natl. Acad. Sci.* **106**, 9955–9962. (doi:10.1073/pnas.0901264106)

3. Rundle HD, Nosil P. 2005 Ecological speciation. *Ecol. Lett.* **8**, 336–352. (doi:10.1111/j.1461-0248.2004.00715.x)

4. Hendry AP, Nosil P, Rieseberg LH. 2007 The speed of ecological speciation. *Funct. Ecol.* **21**, 455–464. (doi:10.1111/j.1365-2435.2007.01240.x)

5. Barrett RDH, Schluter D. 2008 Adaptation from standing genetic variation. *Trends Ecol. Evol.* **23**, 38–44. (doi:10.1016/j.tree.2007.09.008)

6. 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. (doi:10.1038/ncomms14363)

7. Meier JI *et al.* 2019 The coincidence of ecological opportunity with hybridization explains rapid adaptive radiation in Lake Mweru cichlid fishes. *Nat. Commun.* **10**, 5391. (doi:10.1038/s41467-019-13278-z)

8. Van Belleghem SM, Vangestel C, De Wolf K, De Corte Z, Möst M, Rastas P, De Meester L, Hendrickx F. 2018 Evolution at two time frames: Polymorphisms from an ancient singular divergence event fuel contemporary parallel evolution. *PLoS Genet.* **14**, e1007796. (doi:10.1101/255554)

9. Welch JJ, Jiggins CD. 2014 Standing and flowing: The complex origins of adaptive variation. *Mol. Ecol.* **23**, 3935–3937. (doi:10.1111/mec.12859)

10. Roesti M, Gavrilets S, Hendry AP, Salzburger W, Berner D. 2014 The genomic signature of parallel adaptation from shared genetic variation. *Mol. Ecol.* **23**, 3944–3956. (doi:10.1111/mec.12720)

11. Nosil P, Funk DJ, Ortiz-Barrientos D. 2009 Divergent selection and heterogeneous genomic divergence. *Mol. Ecol.* **18**, 375–402. (doi:10.1111/j.1365-294X.2008.03946.x)

12. Via S. 2009 Natural selection in action during speciation. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 9939–9946. (doi:10.1073/pnas.0901397106)

13. Nosil P, Feder JL. 2012 Genomic divergence during speciation: causes and consequences. *Philos. Trans. R. Soc. B Biol. Sci.* **367**, 332–342. (doi:10.1098/rstb.2011.0263)

14. Nosil P, Parchman TL, Feder JL, Gompert Z. 2012 Do highly divergent loci reside in genomic regions affecting reproductive isolation? A test using next-generation sequence data in *Timema* stick insects. *BMC Evol. Biol.* **12**, 164. (doi:10.1186/1471-2148-12-164)

15. Yeaman S, Whitlock MC. 2011 The genetic architecture of adaptation under migration-selection balance. *Evolution* **65**, 1897–1911. (doi:10.1111/j.1558-5646.2011.01269.x)

16. Feder JL, Flaxman SM, Egan SP, Nosil P. 2013 Hybridization and the build-up of genomic divergence during speciation. *J. Evol. Biol.* **26**, 261–266. (doi:10.1111/jeb.12009)

17. Flaxman SM, Feder JL, Nosil P. 2013 Genetic hitchhiking and the dynamic buildup of genomic divergence during speciation with gene flow. *Evolution* **67**, 2577–2591. (doi:10.1111/evo.12055)

18. Via S. 2012 Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. *Philos. Trans. R. Soc. B Biol. Sci.* **367**, 451–460.

(doi:10.1098/rstb.2011.0260)

19. Via S, West J. 2008 The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Mol. Ecol.* **17**, 4334–4345. (doi:10.1111/j.1365-294X.2008.03921.x)

20. Cruickshank TE, Hahn MW. 2014 Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol.* **23**, 3133–3157. (doi:10.1111/mec.12796)

21. Noor MAF, Bennett SM. 2009 Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity* **103**, 439–444. (doi:10.1038/hdy.2009.151.Islands)

22. Guerrero RF, Hahn MW. 2017 Speciation as a sieve for ancestral polymorphism. *Mol. Ecol.* **26**, 5362–5368. (doi:10.1111/mec.14290)

23. Quesada H, Posada D, Caballero A, Morán P, Rolán-Alvarez E. 2007 Phylogenetic evidence for multiple sympatric ecological diversification in a marine snail. *Evolution* **61**, 1600–1612. (doi:10.1111/j.1558-5646.2007.00135.x)

24. Bierne N, Gagnaire PA, David P. 2013 The geography of introgression in a patchy environment and the thorn in the side of ecological speciation. *Curr. Zool.* **59**, 72–86.

25. Roux C, Fraïsse C, Castric V, Vekemans X, Pogson GH, Bierne N. 2014 Can we continue to neglect genomic variation in introgression rates when inferring the history of speciation? A case study in a *Mytilus* hybrid zone. *J. Evol. Biol.* **27**, 1662–1675. (doi:10.1111/jeb.12425)

26. Martin BT, Douglas MR, Chafin TK, Placyk JS, Birkhead RD, Phillips CA, Douglas ME. 2019 Contrasting signatures of introgression in North American box turtle (*Terrapene* spp.) contact zones. *Mol. Ecol.* **29**, 4186–4202. (doi:10.1101/752196)

27. Payseur BA, Rieseberg LH. 2016 A genomic perspective on hybridization and speciation. *Mol. Ecol.* **25**, 2337–2360. (doi:10.1111/mec.13557)

28. Runemark A, Fernández LP, Eroukhmanoff F, Sætre GP. 2018 Genomic contingencies and the potential for local adaptation in a hybrid species. *Am. Nat.* **192**, 10–22. (doi:10.1086/697563)

29. Edelaar P, Siepielski AM, Clobert J. 2008 Matching habitat choice causes directed gene flow: A neglected dimension in evolution and ecology. *Evolution* **62**, 2462–2472. (doi:10.1111/j.1558-5646.2008.00459.x)

30. Pérez-Pereira N, Quesada H, Caballero A. 2017 Can parallel ecological speciation be detected with phylogenetic analyses? *Mol. Phylogenet. Evol.* **116**, 149–156. (doi:10.1016/j.ympev.2017.08.019)

31. Le Moan A, Gagnaire PA, Bonhomme F. 2016 Parallel genetic divergence among coastal-marine ecotype pairs of European anchovy explained by differential introgression after secondary contact. *Mol. Ecol.* **25**, 3187–3202. (doi:10.1111/mec.13627)

32. Qi D, Li T, Zhao X, Guo S, Li J. 2006 Mitochondrial cytochrome b sequence variation and phylogenetics of the highly specialized schizothoracine fishes (Teleostei: Cyprinidae) in the Qinghai-Tibet plateau. *Biochem. Genet.* **44**, 270–285. (doi:10.1007/s10528-006-

9022-5)

33. Zhang D *et al.* 2017 Genetic adaptation of schizothoracine fish to the phased uplifting of the Qinghai-Tibetan Plateau. *G3 Genes, Genomes, Genet.* **7**, 1267–1276. (doi:10.1534/g3.116.038406)

34. Zhang C *et al.* 2018 Adaptive evolution of the Eda gene and scales loss in schizothoracine fishes in response to uplift of the Tibetan Plateau. *Int. J. Mol. Sci.* **19**, 1–14. (doi:10.3390/ijms19102953)

35. Guan L, Chi W, Xiao W, Chen L, He S. 2014 Analysis of hypoxia-inducible factor alpha polyploidization reveals adaptation to Tibetan Plateau in the evolution of schizothoracine fish. *BMC Evol. Biol.* **14**, 17–21. (doi:10.1186/s12862-014-0192-1)

36. Tang Y, Li C, Wanghe K, Feng C, Tong C, Tian F, Zhao K. 2019 Convergent evolution misled taxonomy in schizothoracine fishes (Cypriniformes: Cyprinidae). *Mol. Phylogenet. Evol.* **134**, 323–337. (doi:10.1016/j.ympev.2019.01.008)

37. Qi D, Chao Y, Guo S, Zhao L, Li T, Wei F, Zhao X. 2012 Convergent, parallel and correlated evolution of trophic morphologies in the subfamily schizothoracinae from the Qinghai-Tibetan Plateau. *PLoS One* **7**, 1–10. (doi:10.1371/journal.pone.0034070)

38. Regmi B, Douglas MR, Edds DR, Douglas ME. 2020 Geometric morphometric analyses define riverine and lacustrine species-flocks of Himalayan snowtrout (Cyprinidae: Schizothorax) in Nepal. *Aquat. Biol.* **30**, 19–31. (doi:10.3354/ab00737)

39. Tshering S, Wangchuk K, Dorji S, Norbu P, Philipp DP, Claussen JE, Douglas ME, Douglas MR. 2017 *Field Guide to Fishes of Western Bhutan*. Thimphu, Bhutan: Kuensel Corporation, Ltd.

40. Terashima A. 1984 Three new species of the cyprinid genus *Schizothorax* from Lake Rara, northwestern Nepal. *Japanese J. Ichthyol.* **31**, 122–135. (doi:10.11369/jji1950.31.122)

41. Dimmick WW, Edds DR. 2002 Evolutionary genetics of the endemic schizorathicine (Cypriniformes: Cyprinidae) fishes of Lake Rara, Nepal. *Biochem. Syst. Ecol.* **30**, 919–929. (doi:10.1016/S0305-1978(02)00030-3)

42. Qiao J, Hu J, Xia Q, Zhu R, Chen K, Zhao J, Yan Y, Chu L, He D. 2020 Pelagic–benthic resource polymorphism in *Schizopygopsis thermalis* Herzenstein 1891 (Pisces, Cyprinidae) in a headwater lake in the Salween River system on the Tibetan Plateau. *Ecol. Evol.* **10**, 7431–7444. (doi:10.1002/ece3.6470)

43. Regmi B *et al.* 2020 The Himalayan uplift and the evolution of aquatic biodiversity across Asia□: Snowtrout (Cyprininae: *Schizothorax*) as a test case. *bioRxiv* (doi:10.1101/2020.10.12.336149)

44. Shrestha J. 1999 Coldwater fish and fisheries of Nepal. In *Fish and fisheries at higher altitudes: Asia. FAO Fisheries Technical Paper No. 385* (ed T Petr), p. 304. Rome.

45. Sehgal KL. 1999 Coldwater fish and fisheries in the Indian Himalayas: Rivers and streams. In *Fish and fisheries at higher altitudes: Asia. FAO Fisheries Technical Paper No. 385* (ed T Petr), p. 304. Rome.

46. Petr T. 1999 Coldwater fish and fisheries in Bhutan. In *Fish and fisheries at higher altitudes: Asia. FAO Fisheries Technical Paper No. 385* (ed T Petr), p. 304. Rome.

47. Edds DR. 1993 Fish assemblage structure and environmental correlates in Nepal's Gandaki River. *Copeia* **1993**, 48–60.

48. Li X, Guo B. 2020 Substantially adaptive potential in polyploid cyprinid fishes: Evidence from biogeographic, phylogenetic and genomic studies. *Proc. R. Soc. B Biol. Sci.* **287**. (doi:10.1098/rspb.2019.3008)

49. Tong C, Tian F, Zhao K. 2017 Genomic signature of highland adaptation in fish: A case study in Tibetan Schizothoracinae species. *BMC Genomics* **18**, 948. (doi:10.1186/s12864-017-4352-8)

50. Levin DA, Soltis DE. 2018 Factors promoting polyploid persistence and diversification and limiting diploid speciation during the K–Pg interlude. *Curr. Opin. Plant Biol.* **42**, 1–7. (doi:10.1016/j.pbi.2017.09.010)

51. Cai L, Xi Z, Amorim AM, Sugumaran M, Rest JS, Liu L, Davis CC. 2019 Widespread ancient whole-genome duplications in *Malpighiales* coincide with Eocene global climatic upheaval. *New Phytol.* **221**, 565–576. (doi:10.1111/nph.15357)

52. Van de Peer Y, Ashman T-L, Soltis PS, Soltis DE. 2020 Polyploidy: an evolutionary and ecological force in stressful times. *Plant Cell* **33**, 11–26. (doi:10.1093/plcell/koaa015)

53. Favre A, Päckert M, Pauls SU, Jähnig 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. *Biol. Rev. Camb. Philos. Soc.* **90**, 236–253. (doi:10.1111/brv.12107)

54. Herbert TD, Lawrence KT, Tzanova A, Peterson LC, Caballero-Gill R, Kelly CS. 2016 Late Miocene global cooling and the rise of modern ecosystems. *Nat. Geosci.* **9**, 843–847. (doi:10.1038/ngeo2813)

55. He D, Sui X, Sun H, Tao J, Ding C, Chen Y, Chen Y. 2020 Diversity, pattern and ecological drivers of freshwater fish in China and adjacent areas. *Rev. Fish Biol. Fish.* **30**, 387–404. (doi:10.1007/s11160-020-09600-4)

56. Meirmans PG, Liu S, Van Tienderen PH. 2018 The analysis of polyploid genetic data. *J. Hered.* **109**, 283–296. (doi:10.1093/jhered/esy006)

57. Dufresne F, Stift M, Vergilino R, Mable BK. 2014 Recent progress and challenges in population genetics of polyploid organisms: An overview of current state-of-the-art molecular and statistical tools. *Mol. Ecol.* **23**, 40–69. (doi:10.1111/mec.12581)

58. Clark L V., Lipka AE, Sacks EJ. 2019 polyRAD: Genotype calling with uncertainty from sequencing data in polyploids and diploids. *G3 Genes, Genomes, Genet.* **9**, 663–673. (doi:10.1534/g3.118.200913)

59. Clark L V., Mays W, Lipka AE, Sacks EJ. 2020 A population-level statistic for assessing Mendelian behavior of genotyping-by-sequencing data from highly duplicated genomes. *bioRxiv* (doi:10.1101/2020.01.11.902890)

60. Guo XZ, Zhang GR, Wei KJ, Yan RJ, Ji W, Yang R Bin, Wei QW, Gardner JPA. 2016 Phylogeography and population genetics of *Schizothorax o'connori*: Strong subdivision in

the Yarlung Tsangpo River inferred from mtDNA and microsatellite markers. *Sci. Rep.* **6**, 1–15. (doi:10.1038/srep29821)

61. He D, Chen Y. 2009 Phylogeography of *Schizothorax o'connori* (Cyprinidae: Schizothoracinae) in the Yarlung Tsangpo River, Tibet. *Hydrobiologia* **635**, 251–262. (doi:10.1007/s10750-009-9918-2)

62. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012 Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS One* **7**, e37135.

63. Eaton DAR, Overcast I. 2020 ipyrad: Interactive assembly and analysis of RADseq datasets. *Bioinformatics* **36**, 2592–2594.

64. Zhou C *et al*. 2020 Comprehensive transcriptome data for endemic schizothoracinae fish in the Tibetan Plateau. *Sci. Data* **7**, 1–7. (doi:10.1038/s41597-020-0361-6)

65. Li H, Ruan J, Durbin R. 2008 Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res.* **18**, 1851–1858.

66. Li H, Durbin R. 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760. (doi:10.1093/bioinformatics/btp324)

67. Weiß CL, Pais M, Cano LM, Kamoun S, Burbano HA. 2018 nQuire: a statistical framework for ploidy estimation using next generation sequencing. *BMC Bioinformatics* **19**, 122. (doi:10.1186/s12859-018-2128-z)

68. Adams DC, Otárola-Castillo E. 2013 Geomorph: An r package for the collection and analysis of geometric morphometric shape data. *Methods Ecol. Evol.* **4**, 393–399. (doi:10.1111/2041-210X.12035)

69. Venables WN, Ripley BDB. 2002 *Modern Applied Statistics with S*. Fourth. New York: Springer.

70. Smith ML, Hahn MW. 2021 New approaches for inferring phylogenies in the presence of paralogs. *Trends Genet.* **37**, 174–187. (doi:10.1016/j.tig.2020.08.012)

71. Sarmashghi S, Bohmann K, Thomas P Gilbert M, Bafna V, Mirarab S. 2019 Skmer: Assembly-free and alignment-free sample identification using genome skims. *Genome Biol.* **20**, 34. (doi:10.1101/230409)

72. Balaban M, Sarmashghi S, Mirarab S. 2020 APPLES: Scalable distance-based phylogenetic placement with or without alignments. *Syst. Biol.* **69**, 566–578. (doi:10.1093/sysbio/syz063)

73. Lefort V, Desper R, Gascuel O. 2015 FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol. Biol. Evol.* **32**, 2798–2800. (doi:10.1093/molbev/msv150)

74. Nguyen L-TL-T, Schmidt HA, von Haeseler A, Minh BQ. 2014 IQ-TREE: A fast and effective stochastic algorithm for estimating Maximum-Likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274. (doi:10.1093/molbev/msu300)

75. De Maio N, Schrempf D, Kosiol C. 2015 PoMo: An allele frequency-based approach for species tree estimation. *Syst. Biol.* **64**, 1018–1031. (doi:10.1093/sysbio/syv048)

76. Pritchard JK, Stephens M, Donnelly P. 2000 Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959. (doi:10.1111/j.1471-8286.2007.01758.x)

77. Stift M, Kolář F, Meirmans PG. 2019 Structure is more robust than other clustering methods in simulated mixed-ploidy populations. *Heredity* **123**, 429–441. (doi:10.1038/s41437-019-0247-6)

78. Kopelman NM, Mayzel J, Jakobsson M, Rosenberg NA, Mayrose I. 2015 Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* **15**, 1179–1191.

79. Evanno G, Regnaut S, Goudet J. 2005 Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* **14**, 2611–2620. (doi:10.1111/j.1365-294X.2005.02553.x)

80. Martin BT, Chafin TK, Douglas MR, Placyk Jr. JS, Birkhead RD, Phillips CA, Douglas ME. 2021 The choices we make and the impacts we have: Machine learning and species delimitation in North American box turtles (Terrapene spp.). *Mol. Ecol. Resour.* **Early View**.

81. Jombart T. 2008 Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405. (doi:10.1093/bioinformatics/btn129)

82. Jombart T, Devillard SS, Balloux FF. 2010 Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet.* **11**, 94. (doi: 10.1186/1471-2156-11-94)

83. Pickrell JK, Pritchard JK. 2012 Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967. (doi:10.1371/journal.pgen.1002967)

84. Fitak RR. 2019 optM: an R package to optimize the number of migration edges using threshold models. *https://CRAN.R-project.org/package=dplyr*

85. Malinsky M, Matschiner M, Svardal H. 2021 Dsuite - Fast D-statistics and related admixture evidence from VCF files. *Mol. Ecol. Resour.* **21**, 584–595. (doi:10.1111/1755-0998.13265)

86. 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. (doi:10.1534/genetics.112.145037)

87. Luu K, Bazin E, Blum MGB. 2017 pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Mol. Ecol. Resour.* **17**, 67–77. (doi:10.1111/1755-0998.12592)

88. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. 2005 Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676. (doi:10.1093/bioinformatics/bti610)

89. Hunter S *et al.* 2009 InterPro: The integrative protein signature database. *Nucleic Acids Res.* **37**, 211–215. (doi:10.1093/nar/gkn785)

90. Boeckmann B *et al.* 2003 The SWISS-PROT protein knowledgebase and its supplement

TrEMBL in 2003. *Nucleic Acids Res.* **31**, 365–370. (doi:10.1093/nar/gkg095)

91. Sayols S. 2020 rrvgo: a Bioconductor package to reduce and visualize gene ontology terms. *https://ssayols.github.io/rrvgo*.

92. Nei M. 1987 *Molecular evolutionary genetics*. Columbia University Press.

93. Jost L. 2008 GST and its relatives do not measure differentiation. *Mol. Ecol.* **17**, 4015–4026. (doi:10.1111/j.1365-294X.2008.03887.x)

94. Smith J, Kronforst MR. 2013 Do *Heliconius* butterfly species exchange mimicry alleles? *Biol. Lett.* **9**, 20130503. (doi:10.1098/rsbl.2013.0503)

95. Xiao S *et al.* 2020 Genome of tetraploid fish *Schizothorax o'connori* provides insights into early re-diploidization and high-altitude adaptation. *iScience* **23**, 101497. (doi:10.1016/j.isci.2020.101497)

96. Novikova PY *et al.* 2020 Polyploidy breaks speciation barriers in Australian burrowing frogs *Neobatrachus*. *PLoS Genet.* **16**, 1–24. (doi:10.1371/journal.pgen.1008769)

97. Povilus RA, Diggle PK, Friedman WE. 2018 Evidence for parent-of-origin effects and interparental conflict in seeds of an ancient flowering plant lineage. *Proc. R. Soc. B Biol. Sci.* **285**, 20172491. (doi:10.1098/rspb.2017.2491)

98. Lafon-Placette C *et al.* 2017 Endosperm-based hybridization barriers explain the pattern of gene flow between *Arabidopsis lyrata* and *Arabidopsis arenosa* in Central Europe. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E1027–E1035. (doi:10.1073/pnas.1615123114)

99. Schmickl R, Yant L. 2021 Adaptive introgression: how polyploidy reshapes gene flow landscapes. *New Phytol.* **230**, 457–461. (doi:10.1111/nph.17204)

100. Marburger S *et al.* 2019 Interspecific introgression mediates adaptation to whole genome duplication. *Nat. Commun.* **10**, 1–11. (doi:10.1038/s41467-019-13159-5)

101. Pecinka A, Fang W, Rehmsmeier M, Levy AA, Mittelsten Scheid O. 2011 Polyploidization increases meiotic recombination frequency in *Arabidopsis*. *BMC Biol.* **9**, 24. (doi:10.1186/1741-7007-10-33)

102. Monnahan P *et al.* 2019 Pervasive population genomic consequences of genome duplication in *Arabidopsis arenosa*. *Nat. Ecol. Evol.* **3**, 457–468. (doi:10.1038/s41559-019-0807-4)

103. Otto SP. 2007 The evolutionary consequences of polyploidy. *Cell* **131**, 452–462. (doi:10.1016/j.cell.2007.10.022)

104. Baniaga AE, Marx HE, Arrigo N, Barker MS. 2020 Polyploid plants have faster rates of multivariate niche differentiation than their diploid relatives. *Ecol. Lett.* **23**, 68–78. (doi:10.1111/ele.13402)

105. Van De Peer Y, Mizrachi E, Marchal K. 2017 The evolutionary significance of polyploidy. *Nat. Rev. Genet.* **18**, 411–424. (doi:10.1038/nrg.2017.26)

106. Bangs MR, Douglas MR, Chafin TK, Douglas ME. 2020 Gene flow and species delimitation in fishes of Western North America: Flannelmouth (*Catostomus latipinnis*) and Bluehead sucker (*C. Pantosteus discobolus*). *Ecol. Evol.* **10**, 6477–6493.

107. Chafin TK, Douglas MR, Martin BT, Douglas ME. 2019 Hybridization drives genetic erosion in sympatric desert fishes of western North America. *Heredity* **123**, 759–773.

108. Bangs MR, Douglas MR, Brunner PC, Douglas ME. 2020 Reticulate evolution as a management challenge: Patterns of admixture with phylogenetic distance in endemic fishes of western North America. *Evol. Appl.* **13**, 1400–1419. (doi:10.1111/eva.13042)

109. Sharma A, Dubey VK, Johnson JA, Rawal YK, Sivakumar K. 2021 Is there always space at the top? Ensemble modeling reveals climate-driven high-altitude squeeze for the vulnerable snow trout *Schizothorax richardsonii* in Himalaya. *Ecol. Indic.* **120**, 106900. (doi:10.1016/j.ecolind.2020.106900)

110. Todesco M *et al.* 2016 Hybridization and extinction. *Evol. Appl.* **9**, 892–908. (doi:10.1111/eva.12367)

111. Seehausen O, Takimoto G, Roy D, Jokela J. 2008 Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Mol. Ecol.* **17**, 30–44. (doi:10.1111/j.1365-294X.2007.03529.x)

112. Ogino K, Dash SK, Nakayama M. 2019 Change to hydropower development in Bhutan and Nepal. *Energy Sustain. Dev.* **50**, 1–17. (doi:10.1016/j.esd.2019.02.005)

113. Mohan M. 2005 Spawning biology of snow trout, *Schizothorax richardsonii* (Gray) from River Gaula (Kumaon, Himalaya). *Indian J. Fish.* **52**, 451–457.

114. Bangs MR, Douglas MR, Thompson P, Douglas ME. 2017 Anthropogenic impacts facilitate native fish hybridization in the Bonneville Basin of Western North America. *Trans. Am. Fish. Soc.* **146**, 16–21. (doi:10.1080/00028487.2016.1235611)

115. Dibble KL, Yackulic CB, Schmidt JC, Kennedy TA, Bestgen KR. 2020 Water storage decisions will determine the distribution and persistence of imperiled river fishes. *Ecol. Appl.* (doi:10.1002/eap.2279)

116. Douglas MR, Douglas ME. 2000 Late season reproduction by big-river Catostomidae in Grand Canyon (Arizona). *Copeia* **2000**, 238–244. (doi:10.1643/0045-8511(2000)2000[0238:LSRBBR]2.0.CO;2)

117. Dorji T, Linke S, Sheldon F. 2020 Freshwater conservation planning in the context of nature needs half and protected area dynamism in Bhutan. *Biol. Conserv.* **251**, 108785.

118. Bhatt JP, Pandit MK. 2016 Endangered Golden mahseer *Tor putitora* Hamilton: a review of natural history. *Rev. Fish Biol. Fish.* (doi:10.1007/s11160-015-9409-7)

119. Qadri MY, Yousuf AR. 1983 Breeding biology of *Schizothorax richardsonii* Gray and Hard. *J. Indian Inst. Sci.* **64**, 73–81.

120. Foll M, Gaggiotti OE, Daub JT, Vatsiou A, Excoffier L. 2014 Widespread signals of convergent adaptation to high altitude in Asia and America. *Am. J. Hum. Genet.* **95**, 394–407. (doi:10.1016/j.ajhg.2014.09.002)

121. Witt KE, Huerta-Sánchez E. 2019 Convergent evolution in human and domesticate adaptation to high-altitude environments. *Philos. Trans. R. Soc. B Biol. Sci.* **374**. (doi:10.1098/rstb.2018.0235)

122. Chi W, Ma X, Niu J, Zou M. 2017 Genome-wide identification of genes probably relevant

to the adaptation of schizothoracins (Teleostei: Cypriniformes) to the uplift of the Qinghai-Tibet Plateau. *BMC Genomics* **18**, 1–8. (doi:10.1186/s12864-017-3703-9)

123. Menon AGK. 1949 Notes on fishes in the Indian Museum. XLIV. Fishes from the Kosi Kimalayas, Nepal. *Rec. Indian Museum* **47**, 231–237.

124. Jacobsen D. 2008 Low oxygen pressure as a driving factor for the altitudinal decline in taxon richness of stream macroinvertebrates. *Oecologia* **154**, 795–807. (doi:10.1007/s00442-007-0877-x)

125. Hoppeler F, Tachamo Shah RD, Shah DN, Jähnig SC, Tonkin JD, Sharma S, Pauls SU. 2016 Environmental and spatial characterisation of an unknown fauna using DNA sequencing – an example with Himalayan Hydropsychidae (Insecta: Trichoptera). *Freshw. Biol.* **61**, 1905–1920. (doi:10.1111/fwb.12824)

126. Tonkin JD, Tachamo Shah RD, Shah DN, Hoppeler F, J□hnig SC, Pauls SU. 2017 Metacommunity structuring in Himalayan streams over large elevational gradients: the role of dispersal routes and niche characteristics. *J. Biogeogr.* **44**, 62–74. (doi:10.1111/jbi.12895)

127. Li J, He Q, Hua X, Zhou J, Xu H, Chen J, Fu C. 2009 Climate and history explain the species richness peak at mid-elevation for Schizothorax fishes (Cypriniformes: Cyprinidae) distributed in the Tibetan Plateau and its adjacent regions. *Glob. Ecol. Biogeogr.* **18**, 264–272. (doi:10.1111/j.1466-8238.2008.00430.x)

128. Rudman SM *et al.* 2017 What genomic data can reveal about eco-evolutionary dynamics. *Nat. Ecol. Evol.* (doi:10.1038/s41559-017-0385-2)

129. Otto SP, Whitton J. 2000 Polyploid incidence and evolution. *Annu. Rev. Genet.* **34**, 401–437.

Tables and Figures

Figure 1: Sampling locality information for *N*=96 *Schizothorax* spp. selected for ddRAD sequencing. Sites represent major Himalayan tributaries of the Brahmaputra and Ganges rivers from Nepal and Bhutan (A). Samples for five species (*S. macrophthalmus*, *S. nepalensis*, *S. progastus*, *S. raraensis*, and *S. richardsonii*) were collected from 11 sites in the major drainages of Nepal (Karnali, Gandaki, and Koshi; B). Major drainages of Bhutan (Wang Chhu, Punatsang Chhu, and Mangde Chhu) were represented by two species (*S. progastus* and *S. cf. oconnori*) collected from 11 localities (C). *Schizothorax progastus* from both Nepal (D) and Bhutan (E) were collected from relatively lower elevations (~2000-4000 ft), whereas *S. richardsonii* (Nepal; D) and *S. cf. oconnori* (Bhutan; E) were generally found >4000 ft.

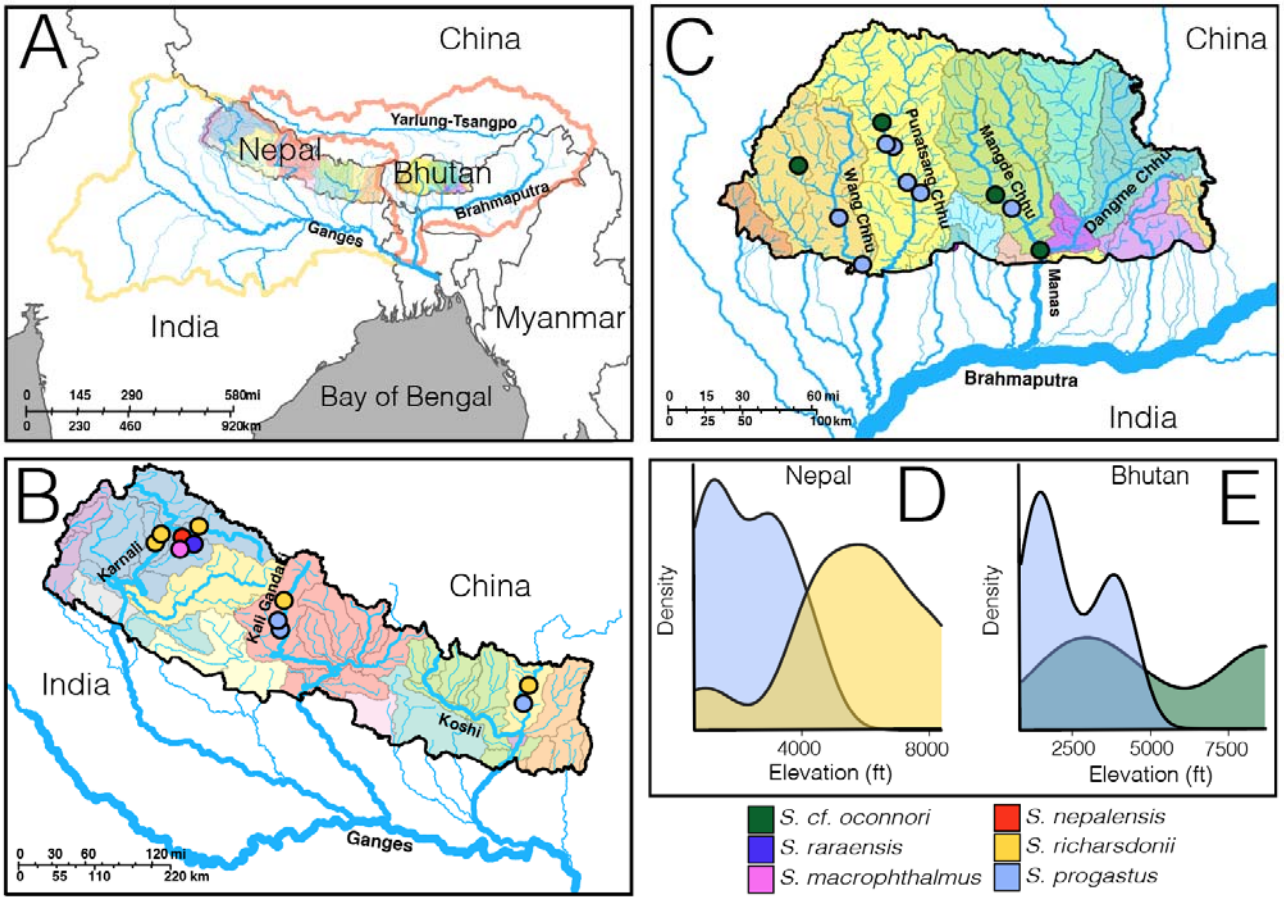


Figure 2: Evaluation of ploidy from *N*=48,350 ddRAD loci for six *Schizothorax* species. (Left) Distributions of base frequencies among raw reads representing bi-allelic SNPs for six *Schizothorax* spp. collected from Nepal [*S. macrohpthalmus*, *S. nepalensis*, *S. progastus* (solid line), *S. raraensis*, and *S. richardsonii*] and Bhutan [*S. cf. oconnori* and *S. progastus* (dashed line)]. (Right) Likelihoods from Gaussian Mixture Models of fixed ploidy (e.g., diploid, triploid, tetraploid), divided by the likelihood under a model of freely variable base frequencies (=logL_{Ploidy}/logL_{Free}).

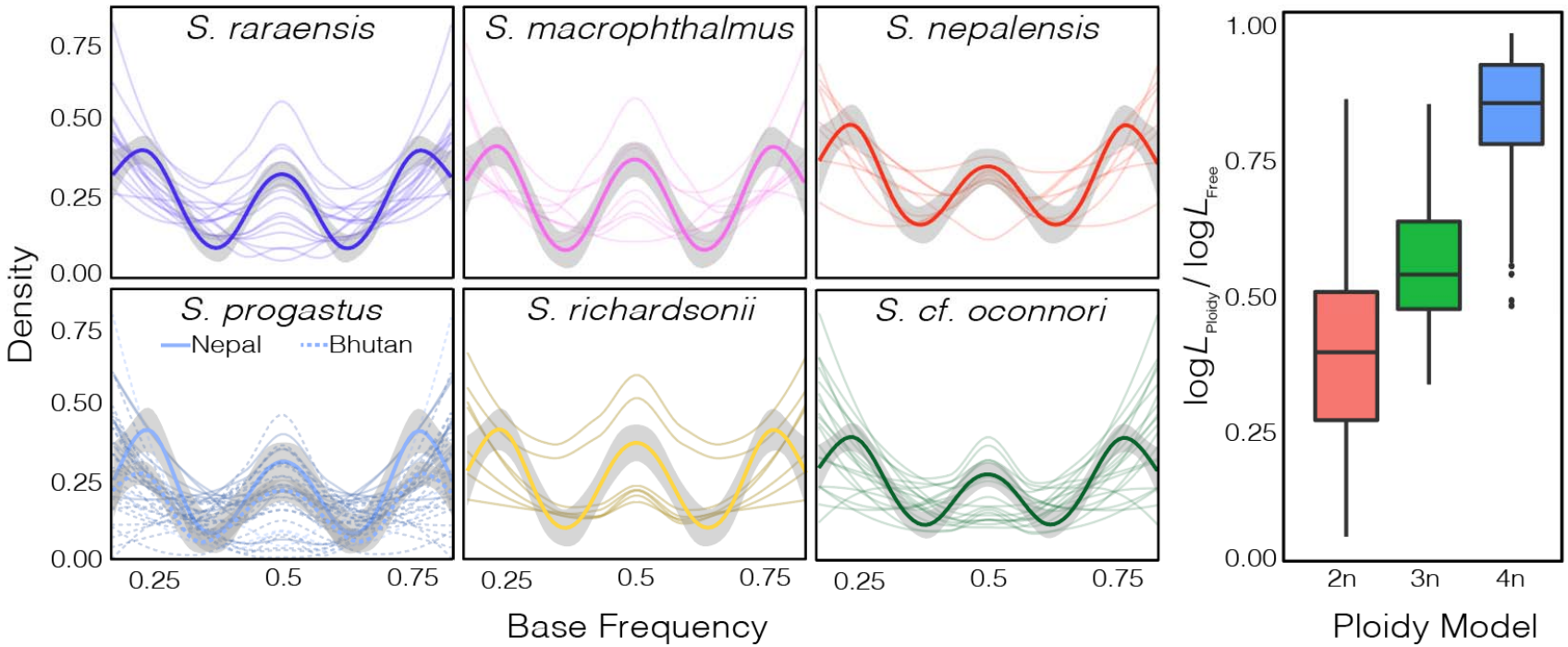


Figure 3: Discriminant analysis of principal components (DAPC) of morphological shape variation among five *Schizothorax* species. Results represent *N*=528 individuals of *S. macrophthalmus*, *S. nepalensis*, *S. progastus*, *S. raraensis*, and *S. richardsonii* collected from the Gandaki, Karnali (to include Lake Rara), and Koshi drainages of Nepal, and Procrustes-aligned two-dimensional Cartesian coordinates for 18 anatomical landmarks.

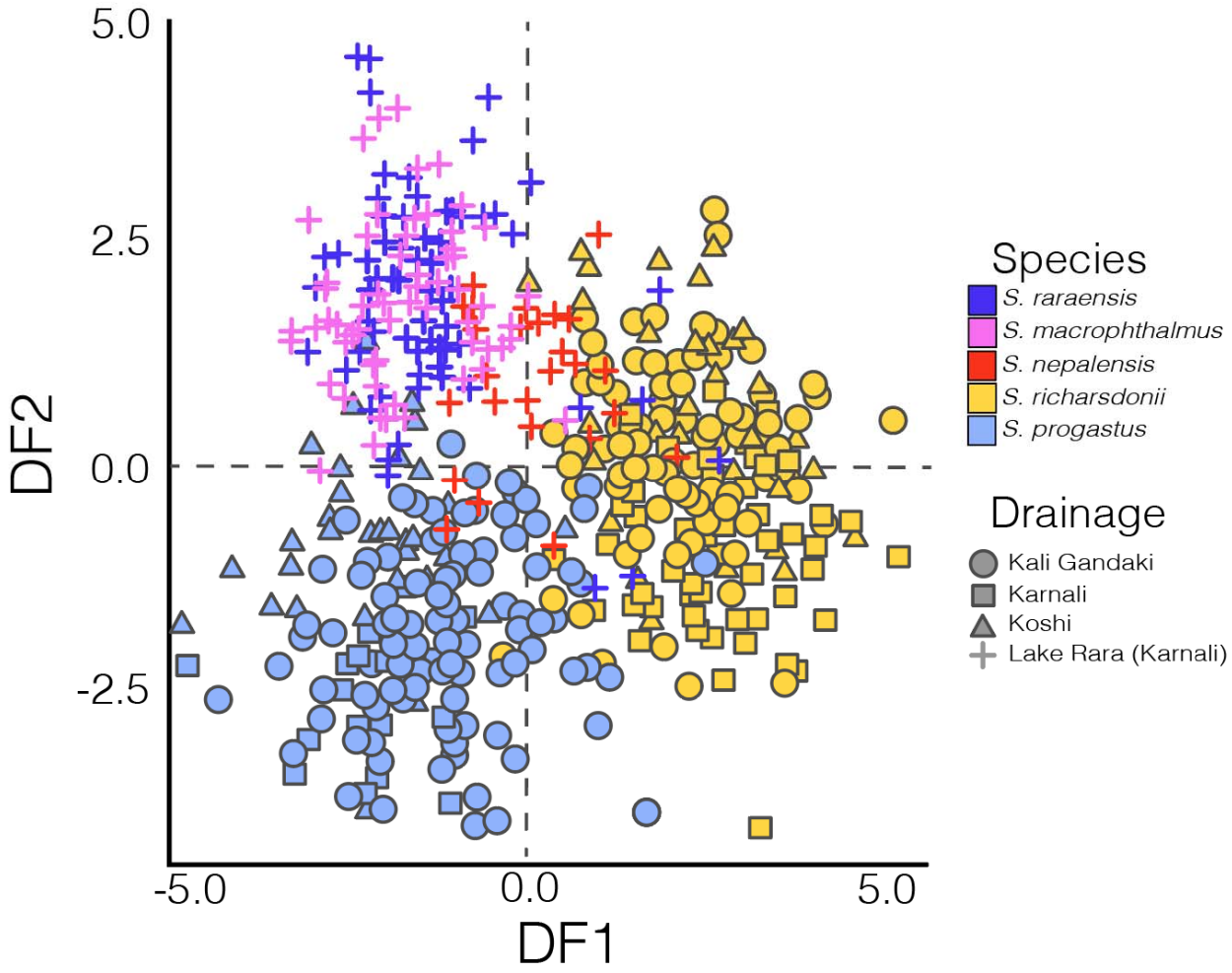


Figure 4: Visualization of genetic variation among six *Schizothorax* species, as assessed across *N*=15,547 statistically diploid SNPs. Visualizations represent the first two discriminant functions from a discriminant analysis of principal components (DAPC) computed across (A) all samples from Bhutan (*S. cf. oconnori* and *S. progastus*) and Nepal (*S. macrophthalmus*, *S. nepalensis*, *S. progastus*, *S. raranesis*, and *S. richardsonii*); (B) Only those samples from Bhutan, additionally partitioned by major drainage encompassing the samples locality (i.e., Wang Chhu, Mangde Chhu, or Punatsang Chhu); and (C) Only individuals samples in Nepal. Nepali samples are partitioned by major drainage (i.e., Gandaki, Koshi, and Karnali), with species from Lake Rara (*S. macrophthalmus*, *S. nepalensis*, and *S. raraensis*) additionally separated.

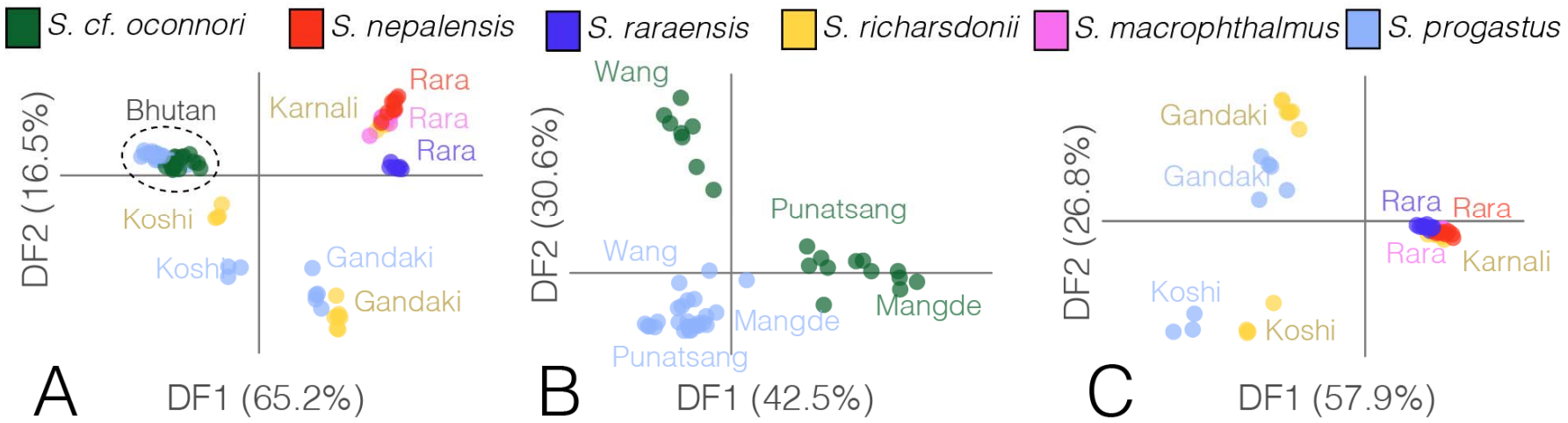


Figure 5: Phylogenetic and population structure of *N*=93 *Schizothorax* spp. collected in major drainages of Bhutan and Nepal. Results represent (A) an unrooted alignment/assembly free phylogeny inferred from k-mer distances, computed from raw ddRAD reads. Clades are grouped by drainage, and with colored bars denoting tips grouped by taxon (*S. macrohpthalmus*, *S. nepalensis*, *S. cf. oconnori*, *S. progastus*, *S. raraensis*, and *S. richardsonii*). (B) Bar plots are provided for population assignments at *K*=3 and *K*=7 for (left) 10,884 transcriptome-aligned SNPs and (right) 35,319 *de novo* assembled SNPs. (C) TREEMIX analyses of population mixtures, with colored arrows representing inferred migration weights.

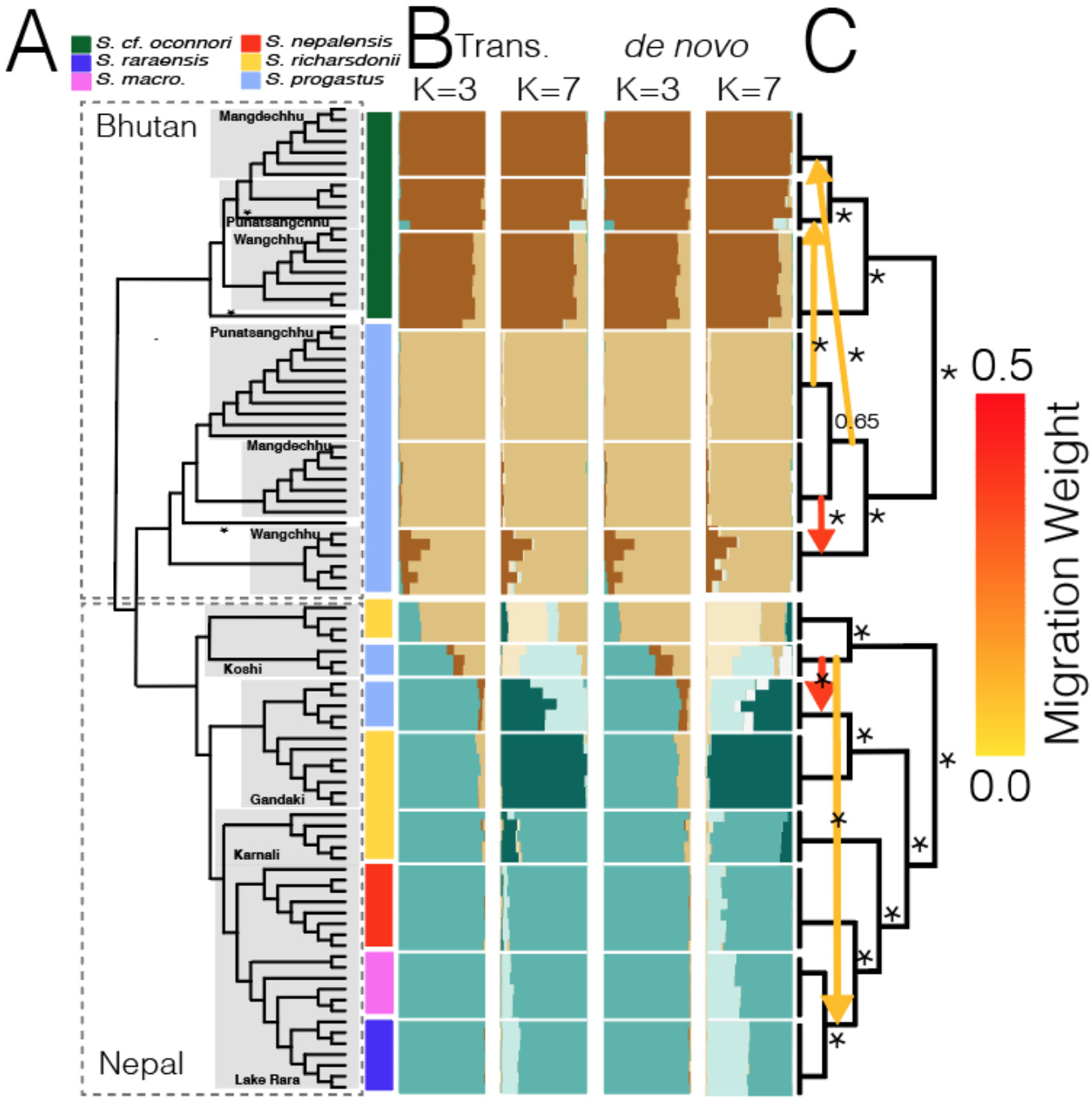
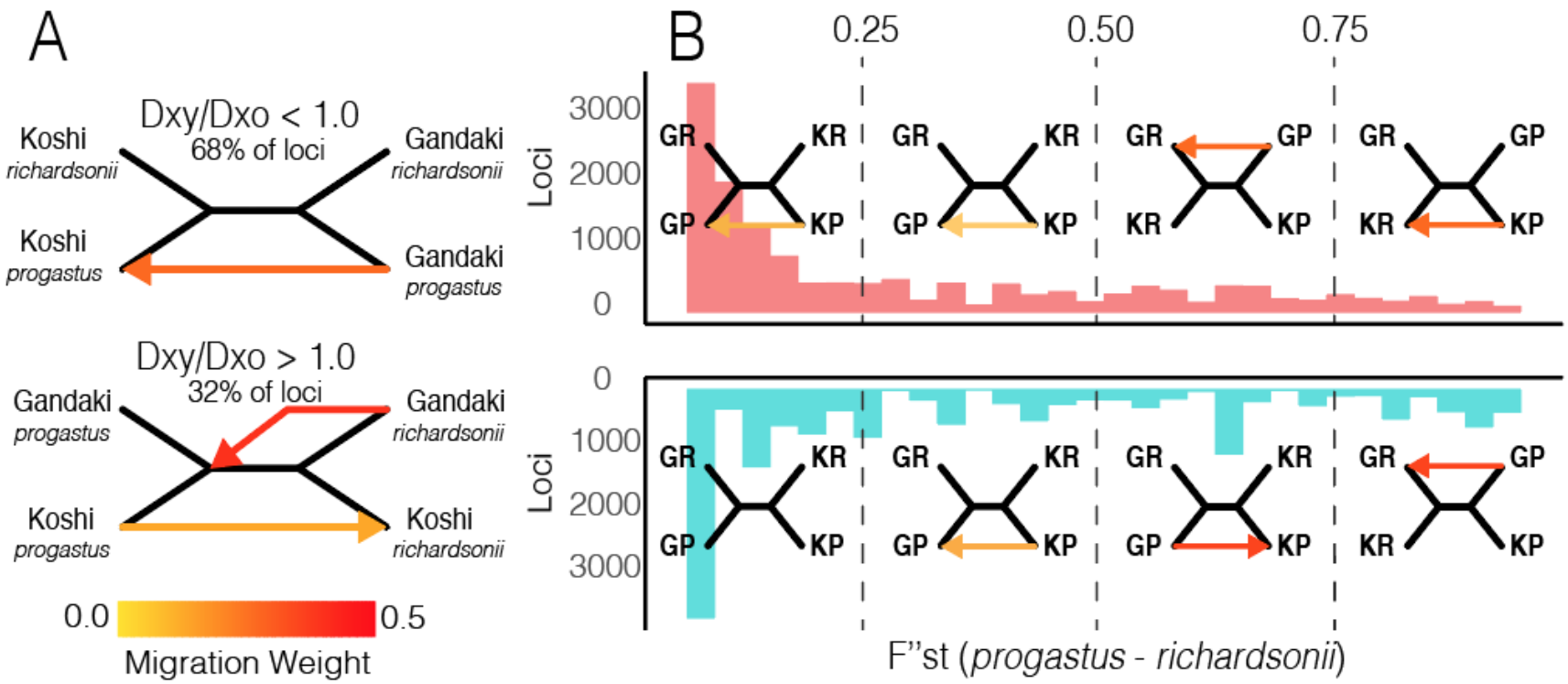


Figure 7: TREEMIX analyses of population mixtures among *S. progastus* and *S. richardsonii* pairs collected from the Gandaki and Koshi rivers of Nepal, based on partitioning different subsets of loci. (A) Loci were grouped by ratios of D_{XY} (computed between heterospecifics within each drainage) and D_{XO} (computed between conspecific *S. richardsonii*), with those loci having a D_{XY} ratio > 1.0 in both Gandaki and Koshi comparisons (i.e., greater distances among *S. richardsonii* and *S. progastus* within drainage than *S. richardsonii* among drainages) analyzed separately. (B) Loci were also partitioned into four bins defined by per-locus F_{ST} computed in Gandaki (above; red) and Koshi (below; blue) pairs. Colored arrows indicate inferred migrations within each partitioned analysis, with the color indicating the migration weight.



Supplementary Materials

Table S1: Sampling locality information for *Schizothorax* samples included in ddRAD analyses. Listed are: Species name; Source (NRCR&LF = National Resource Centre for Riverine and Lake Fisheries, Ministry of Agriculture, Royal Government of Bhutan; KU=University of Kansas Biodiversity Institute and Natural History Museum); Country; Drainage; Site; Latitude (Lat); Longitude (Long); and *N* (sample size per site).

Species	Source	Country	Drainage	Site	Lat	Long	N
oconnori	NRCR&LF	Bhutan	Wang Chhu	Haa Chhu	27.416	90.692	3
oconnori	NRCR&LF	Bhutan	Wang Chhu	Nya Chhu	27.374	89.291	4
oconnori	NRCR&LF	Bhutan	Punatsang Chhu	Pho Chhu	27.592	89.872	4
oconnori	NRCR&LF	Bhutan	Mangde Chhu	Bibigang Chhu	27.156	90.668	4
oconnori	NRCR&LF	Bhutan	Mangde Chhu	Rindigang Chhu	27.005	90.836	4
progastus	NRCR&LF	Bhutan	Wang Chhu	Chukha	27.742	89.738	3
progastus	NRCR&LF	Bhutan	Wang Chhu	Kalikhola	27.052	89.573	3
progastus	NRCR&LF	Bhutan	Punatsang Chhu	Dik Chhu	26.883	90.271	2
progastus	NRCR&LF	Bhutan	Punatsang Chhu	Shengarong Chhu/ Toebrang Chhu	27.526	89.872	5
progastus	NRCR&LF	Bhutan	Punatsang Chhu	Dang Chhu	27.480	89.910	2
progastus	NRCR&LF	Bhutan	Punatsang Chhu	Kame Chhu	27.269	89.036	3
progastus	NRCR&LF	Bhutan	Punatsang Chhu	Tinku Chhu	27.221	90.152	2
progastus	NRCR&LF	Bhutan	Mangde Chhu	Mainstem	26.967	90.863	2
progastus	NRCR&LF	Bhutan	Mangde Chhu	Dakpai Chhu	27.146	90.692	4
Macroph	KU	Nepal	Karnali	Lake Rara	29.539	82.077	8
nepalensis	KU	Nepal	Karnali	Lake Rara	29.533	82.067	8
progastus	KU	Nepal	Gandaki	Kali Gandaki@Beni	28.340	83.568	2
progastus	KU	Nepal	Gandaki	Rahugat River	28.399	83.567	3
progastus	KU	Nepal	Koshi	Tumlingtar	27.340	87.188	3
raraensis	KU	Nepal	Karnali	Lake Rara	29.538	82.076	7
richardsonii	KU	Nepal	Karnali	Srikot	29.522	81.930	2
richardsonii	KU	Nepal	Karnali	Jhugala	29.522	81.780	1
richardsonii	KU	Nepal	Karnali	Gumgarh River	29.553	81.170	2
richardsonii	KU	Nepal	Gandaki	Kali Gandaki	28.684	81.930	7
richardsonii	KU	Nepal	Koshi	Kotka River	27.552	87.302	4

Table S2: Dataset size for pre- and post-filtering double-digest restriction-site associated DNA sequencing (ddRADseq) for *N*=93 *Schizothorax* spp. retained following data assembly. Shown are, for each assembly method (=*de novo* or transcriptome-guided): the number of post-assembly unfiltered single-nucleotide polymorphisms (SNPs); the number of SNPs retained with mean depth of coverage >20X across individuals (>20X mean depth); the number of SNPs retained with genotypes present in greater than 50% of individuals (>50% individuals); the number of loci retained that were statistically tetraploid or diploid (<0.65 Hind/He); and the number of loci which were genotyped as statistically diploid (<0.52 Hind/He).

Assembly method	**Unfiltered SNPs**	**>20X mean depth**	**>50% individuals**	**<0.65 Hind/He**	**<0.52 Hind/He**
Transcriptome-guided	14,301	11,617	11,582	10,884	5,601
De novo	48,350	36,233	36,202	35,319	15,547

Table S3: Results of four-taxon Patterson's *D*-statistic and admixture ratio (f_4) calculated among *S. richardsonii* and *S. progastus* riverine pairs sampled in the Gandaki and Koshi rivers of Nepal.

P1	**P2**	**P3**	***D***	***Z***	***p*-value**	**f_4 ratio**	**BBAA**	**ABBA**	**BABA**
Gandaki *progastus*	Gandaki *richardsonii*	Karnali *richardsonii*	0.411	29.20	<0.000001	0.4300	982.3	756.2	315.6
Gandaki *richardsonii*	Gandaki *progastus*	Koshi *progastus*	0.326	26.91	<0.000001	0.4017	1288.3	709.8	360.7
Gandaki *progastus*	Gandaki *richardsonii*	Koshi *richardsonii*	0.491	38.51	<0.000001	0.3370	1176.9	675.6	230.9
Koshi *richardsonii*	Koshi *progastus*	Gandaki *progastus*	0.284	17.6	<0.000001	0.3191	835.6	830.0	462.5
Koshi *progastus*	Koshi *richardsonii*	Karnali *richardsonii*	0.368	19.39	<0.000001	0.3021	767.6	756.7	349.2
Karnali *richardsonii*	Gandaki *richardsonii*	Koshi *richardsonii*	0.104	3.83	0.000006	0.0857	714.7	439.4	356.9
Gandaki *richardsonii*	Koshi *richardsonii*	Koshi *progastus*	0.326	20.20	<0.000001	0.4176	792.1	720.5	365.9

Figure S1: Population tree generated for *N*=93 *Schizothorax* spp. divided into 13 populations using a polymorphism-aware model applied to data for 15,543 ddRAD SNPs. Nodal values represent bootstrap support expressed as percentage of 1,000 ultra-fast bootstraps supporting each node.

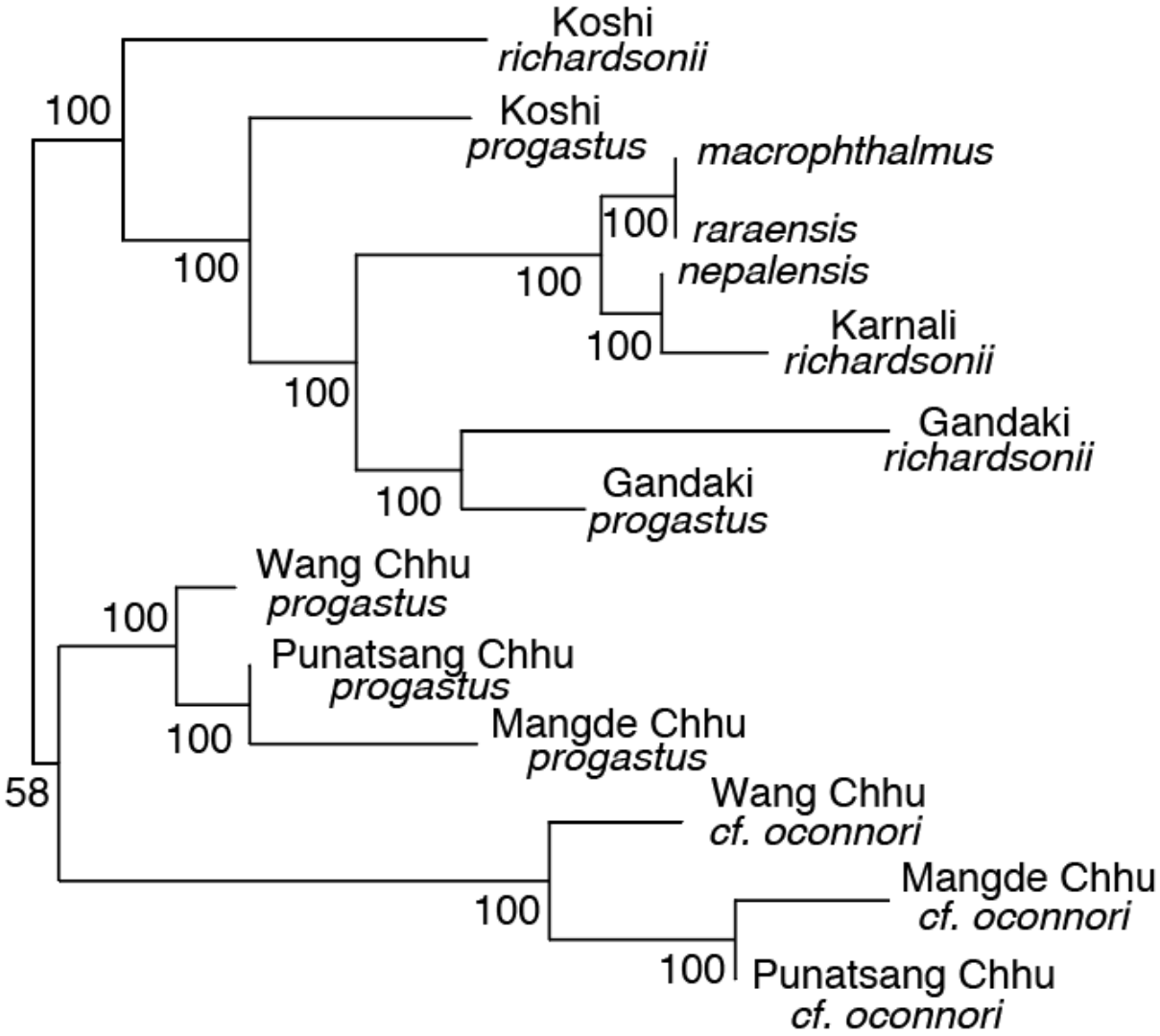


Figure S2: Optimal number of migration edges for TreeMix analyses of Bhutanese (upper) and Nepali (lower) *Schizothorax* species, as assessed across 100 replicates per value of *m* (migration edges; x-axis). Optimal values of *m* for each run were selected using the Evanno method (y-axis).

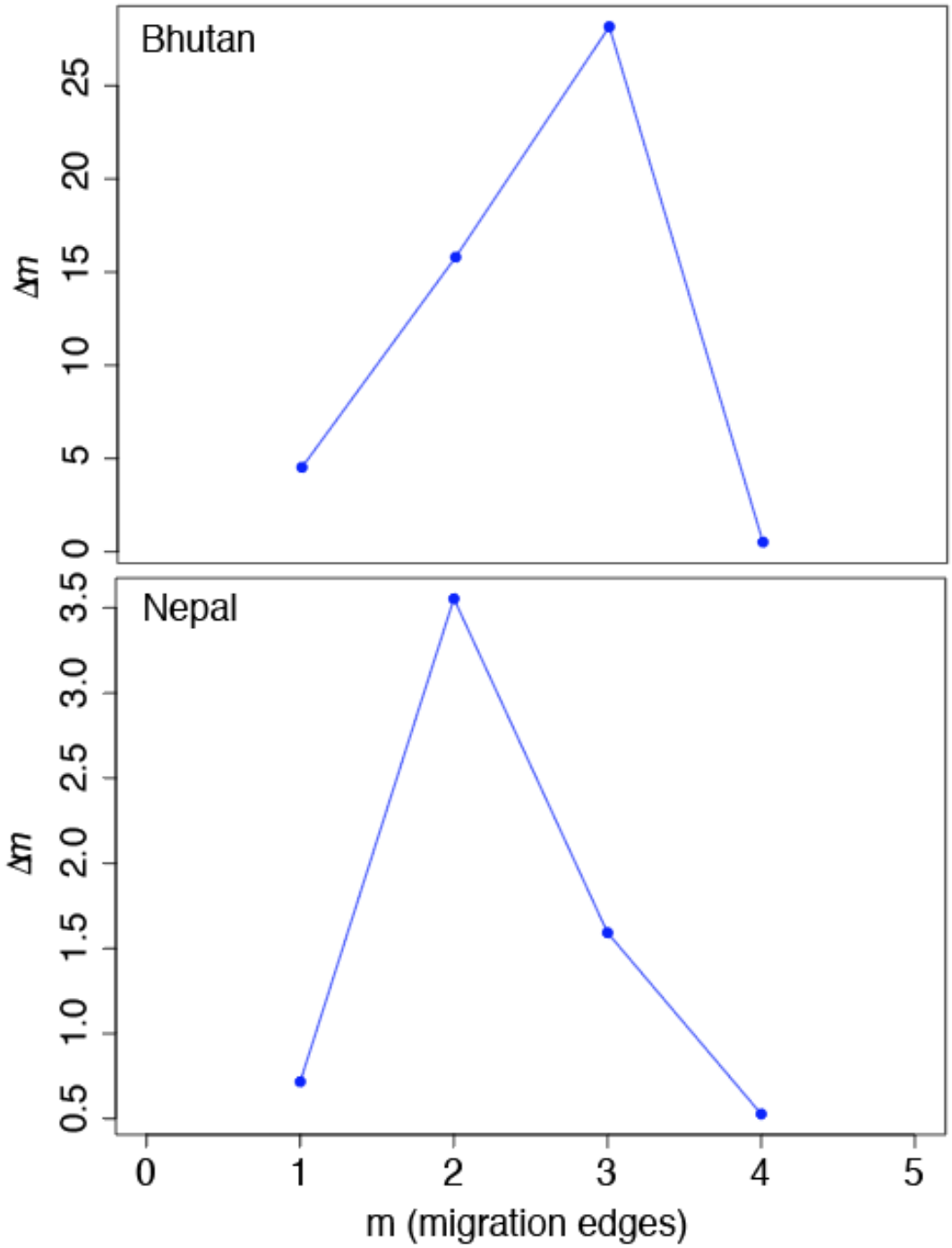


Figure S3: Summary of locus-wise differentiation and outliers identified among pairwise comparisons of *Schizothorax* spp. collected from major drainages in Bhutan and Nepal. (A) Five-way Venn diagram of shared and unique single-nucleotide polymorphisms (SNPs) detected as outliers by PCADAPT within pair-wise comparisons of *S. richardsonii* and *S. progastus* from the Gandaki and Koshi rivers of Nepal; *S. richardsonii* and the Lake Rara 'species complex' (*S. macrophthalmus*, *S. nepalensis*, and *S. raraensis*) in the Karnali River (Karnali-Rara); among Lake Rara species only ("Rara"), and between *S. progastus* and *S. cf. oconnori* from Bhutan. (B) Distribution of site-wise F_{ST}, partitioned by *de novo*, transcriptome-aligned, and outlier SNPs. (C) F_{ST} for SNPs from Gandaki (triangles) and Koshi (circles) river *S. richardsonii* and *S. progastus* comparisons *versus* gene diversity (H_E) for the same loci computed among *S. progastus* for Bhutan for N=31 outlier SNPs (above) and for all *de novo* SNPs (below). (D) Relationship of *richardsonii-progastus* F_{ST} with the ratio of D_{XY} computed within Gandaki (triangle) and Koshi (circle) rivers and D_{XO} computed between conspecific *S. richardsonii* ("D_{XY} Ratio"; where values >1 indicate $D_{XY} > D_{XO}$).

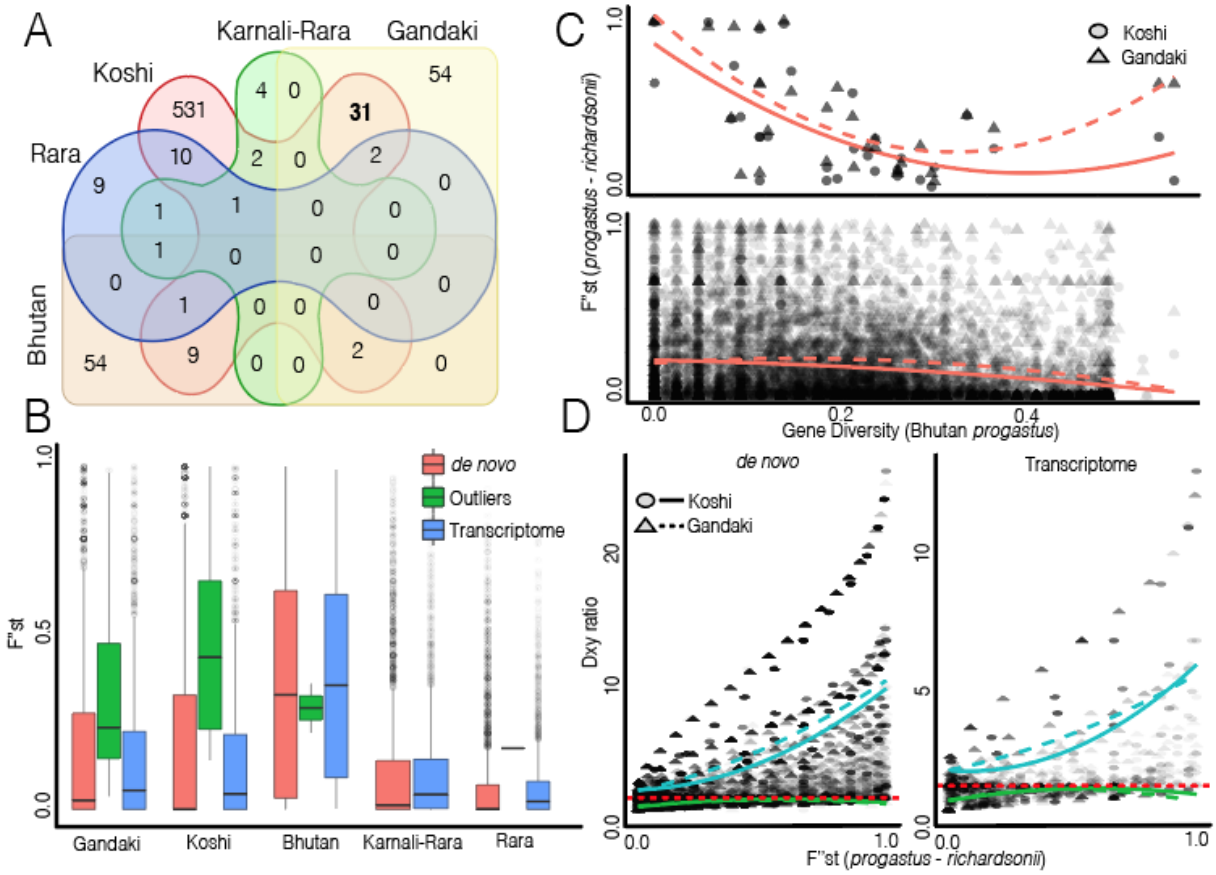


Figure S4: Relationship of F_{ST} values for N=31 outlier SNPs shared among comparisons of *S. richardsonii* and *S. progastus* pairs in the Gandaki and Koshi rivers of Nepal. Shown is the F_{ST} for *richardsonii-progastus* pairs for each locus (y-axis) for the Gandaki (upper-left;x-axis); and for the Karnali *S. richardsonii* X Lake Rara complex (*S. macrophthalmus*, *S. nepalensis*, *S. raraensis*; upper-right, x-axis) *versus* Koshi (circles) and Gandaki (triangles). The latter comparison is also shown for the Bhutanese *S. progastus* X *S. cf. oconnori* (lower-left); and for F_{st} calculated only among Bhutanese *S. progastus* (lower-right). Note that F_{st} for outlier SNPs was only clearly correlated among the two Gandaki and Koshi *richardsonii-progastus* pairs.

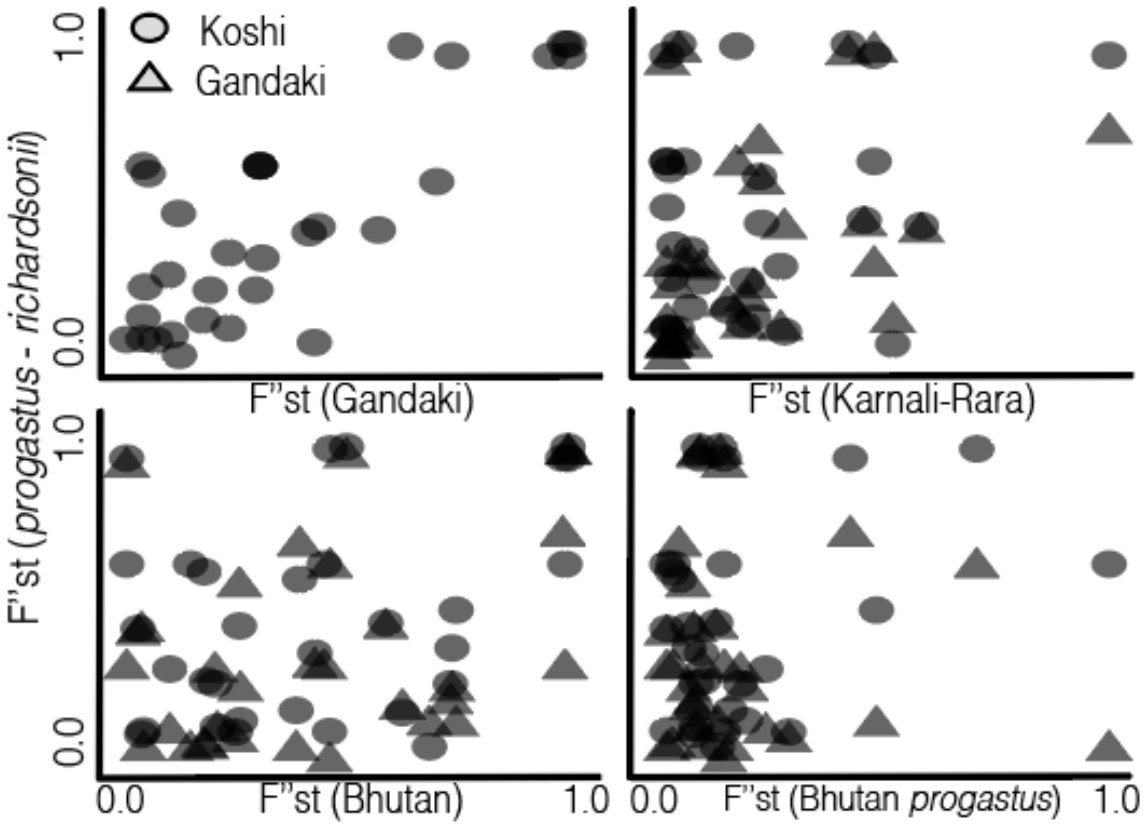


Figure S5: Relationship among locus differentiation statistics for *Schizothorax richardsonii* and *S. progastus* from the Gandaki (above) and Koshi (below) river basins.

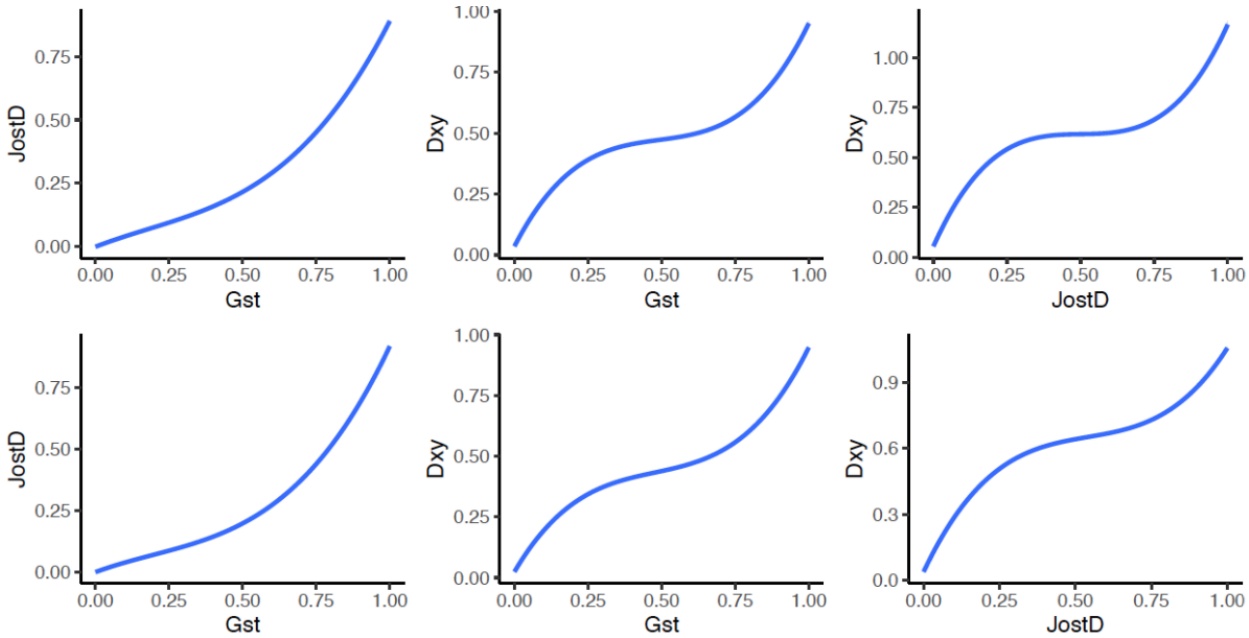


Figure S6: Relationship of D_{xo} (=genetic distance computed from *S. richardsonii* between drainages) and D_{xy}/D_{xo} (where D_{xy} = genetic distance computed between *S. richardsonii* and *S. progastus* within-drainage) for the Gandaki (left) and Koshi (right) rivers.

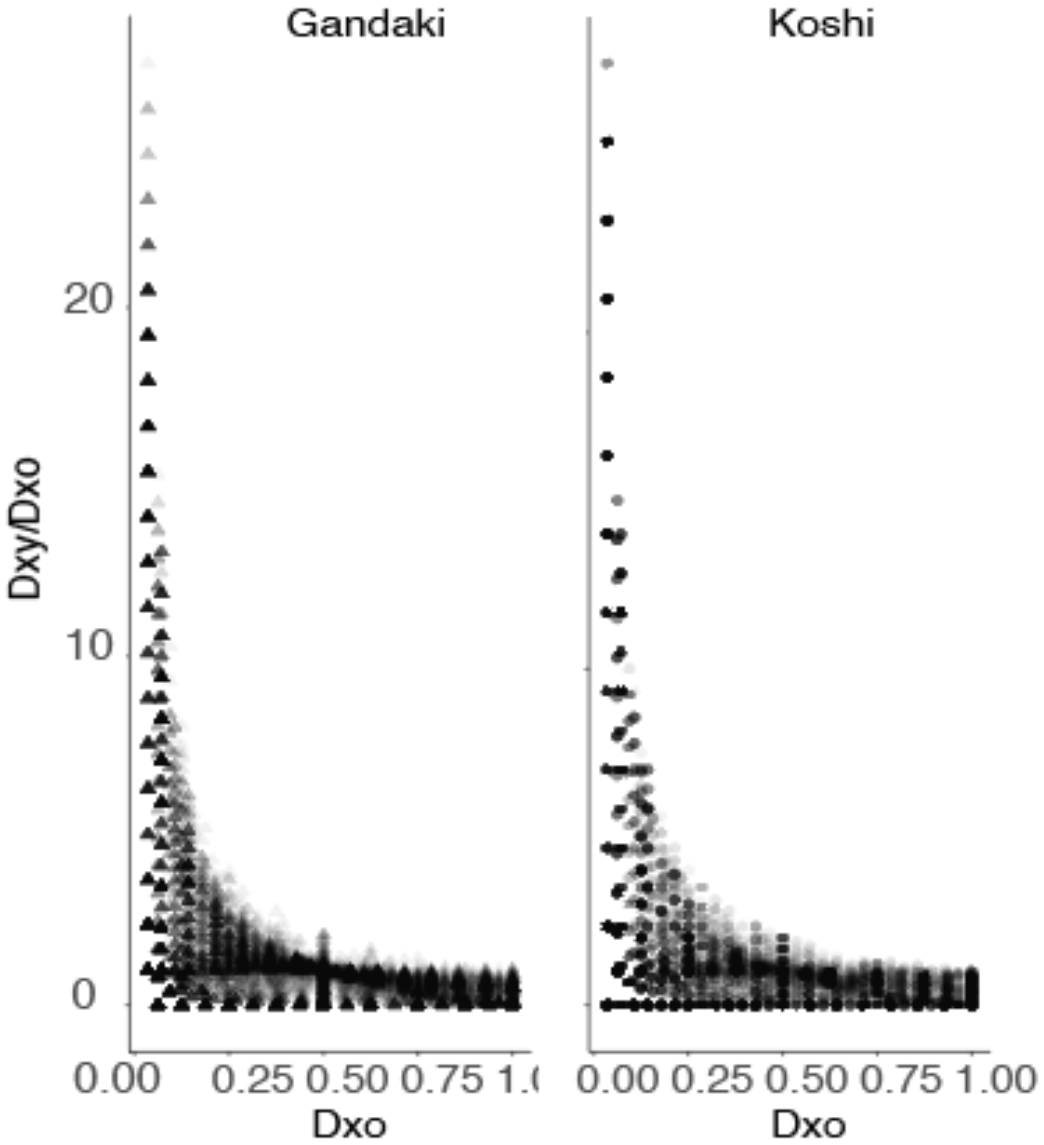
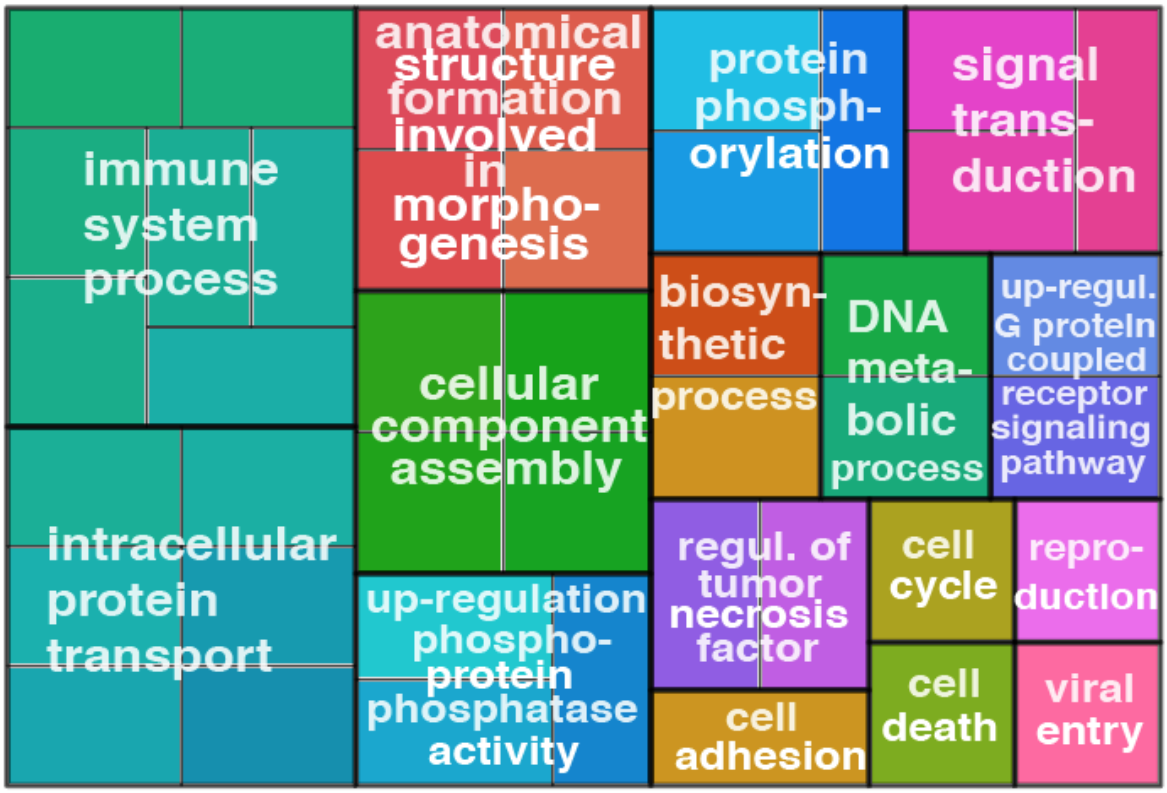



Figure S7: Visualization of gene ontology (GO) terms for outliers identified among *S. richardsonii* and *S. progastus* pairs in the Koshi and Gandaki rivers of Nepal, derived from transcriptome-mapped ddRADseq data and a genome-wide annotation database for Zebrafish (*Danio rerio*).