Species Persistence with Hybridization in Toad-Headed Lizards Driven by Divergent Selection and Low Recombination

Wei Gao,[1,†] Chuan-Xin Yu,[1,2,†] Wei-Wei Zhou,[1,†] Bao-Lin Zhang,[1] E. Anne Chambers,[3,4] Hollis A. Dahn,[5] Jie-Qiong Jin,[1] Robert W. Murphy,[1,5,6] Ya-Ping Zhang,[1,7,*] and Jing Che[1,7,*]

[1]State Key Laboratory of Genetic Resources and Evolution & Yunnan Key Laboratory of Biodiversity and Ecological Security of Gaoligong Mountain, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China

[2]Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming, China

[3]Department of Integrative Biology and Biodiversity Center, University of Texas at Austin, Austin, TX, USA

[4]Department of Environmental Science, Policy, and Management, University of California, Berkeley, CA, USA

[5]Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON, Canada

[6]Centre for Biodiversity and Conservation Biology, Royal Ontario Museum, Toronto, ON, Canada

[7]Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, China

***Corresponding authors:** E-mails: chej@mail.kiz.ac.cn; zhangyp@mail.kiz.ac.cn.

[†]These authors contributed equally to this work.

Associate editor: Guang Yang

Abstract

Speciation plays a central role in evolutionary studies, and particularly how reproductive isolation (RI) evolves. The origins and persistence of RI are distinct processes that require separate evaluations. Treating them separately clarifies the drivers of speciation and then it is possible to link the processes to understand large-scale patterns of diversity. Recent genomic studies have focused predominantly on how species or RI originate. However, we know little about how species persist in face of gene flow. Here, we evaluate a contact zone of two closely related toad-headed lizards (*Phrynocephalus*) using a chromosome-level genome assembly and population genomics. To some extent, recent asymmetric introgression from *Phrynocephalus putjatai* to *P. vlangalii* reduces their genomic differences. However, their highly divergent regions (HDRs) have heterogeneous distributions across the genomes. Functional gene annotation indicates that many genes within HDRs are involved in reproduction and RI. Compared with allopatric populations, contact areas exhibit recent divergent selection on the HDRs and a lower population recombination rate. Taken together, this implies that divergent selection and low genetic recombination help maintain RI. This study provides insights into the genomic mechanisms that drive RI and two species persistence in the face of gene flow during the late stage of speciation.

Key words: **species persistence, divergent selection, recombination, gene flow, *Phrynocephalus*.**

Introduction

Speciation creates biodiversity. Ever since Darwin (1859), the question of how a species diverges into two has formed the basis of evolutionary biology (Mayr 1963; Coyne and Orr 2004). To understand speciation, it is essential to determine how reproductive isolation (RI) evolved between populations (Coyne and Orr 2004; Nosil and Feder 2012; Feder et al. 2013). Newly developed sequencing technologies have ushered in a new era in the study of speciation by offering an unprecedented opportunity to investigate the genetic architecture of RI across the entire genome (Seehausen et al. 2014; Wolf and Ellegren 2017). Recent empirical genomic studies have revealed heterogeneous genetic differentiation between species across the genome and have identified many highly divergent regions (HDRs) of the genome that are possibly involved in speciation (Poelstra et al. 2014; Burri et al. 2015; Wang et al. 2016; Han et al. 2017; Hirase et al. 2021).

The origin and persistence of RI are conceptually distinct processes. The origin of RI involves the accumulation of loci related to it, whereas persistence encompasses the maintenance of RI in the face of gene flow. This distinction is particularly important for allopatric speciation, a geographic model of speciation that is thought to occur most commonly, in which geographical barriers prevent two populations from exchanging genes (Mayr 1963). During geographic isolation, genomic differentiation can occur through both background selection and selective sweeps (Noor and Bennett 2009; Cruickshank and Hahn

© The Author(s) 2022. 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-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Open Access

2014; Burri et al. 2015). Most occurrences of HDRs likely reflect sweeps of adaptive alleles or differentiation by drift, and loci related to RI may or not be involved. Due to species range expansions or the disappearance of geographical isolation, secondary contact can occur among populations or species that have not yet developed complete prezygotic or postzygotic RI. This situation allows for the exchange of genetic components (Abbott et al. 2013; Feder et al. 2013; Vijay et al. 2016), whereupon either the accumulated differentiation is reduced, or selection acts to complete the process of speciation. In the early phases of speciation, genetic exchange results in population fusion (Wu 2001), and in the later phases, species persist (Wu 2001; Abbott et al. 2013) because gene flow erodes genetic differentiation only in parts of the genome where barrier loci do not occur (Ravinet et al. 2017). Thus, natural hybridization via contact provides us with a model for identifying reproductive barrier loci (Harrison and Larson 2014). In contact zones, HDRs of the genome may reflect the maintenance of RI during the late phase of speciation, but not its buildup (Mallet et al. 2009; Feder et al. 2013; Wagner and Mandeville 2017). Such situations facilitate studies of speciation via elucidating how barriers loci promote species' persistence in the face of gene flow (Matute 2010). Many macroevolutionary studies have investigated the role of species' persistence in the discontinuity of speciation rates at different scales (Rosenblum et al. 2012; Dynesius and Jansson 2014; Etienne et al. 2014). Treating the origins and persistence of RI as distinct processes can bridge the mechanisms of speciation and the patterns of large-scale biodiversity (Wagner and Mandeville 2017). Most genome-wide studies have examined the origins. However, genomic mechanisms of their persistence, including what factors maintain the differentiation of genomic barriers to gene flow, need further exploration based on the whole-genome data.

Divergent selection, one of the main driving forces during speciation (Schluter 2000; Rundle and Nosil 2005), results in differentiation either by limiting genetic exchange or directly affecting loci and linked loci (Nosil et al. 2009). Divergent selection against gene flow generates peaks of elevated genetic differentiation during speciation (Rundle et al. 2000; Nosil et al. 2008; Kautt et al. 2020; Hirase et al. 2021; Turbek et al. 2021). Most such studies have focused on isolation by ecology: differentiation due to ecological adaptation. Given that genetic divergence between two allopatric populations accumulates more easily than in the presence of gene flow, the loci involved in RI should diverge at a higher rate in allopatry, especially over longer periods of isolation. When secondary contact occurs, we hypothesize that divergent selection on such loci will maintain divergence. Accordingly, RI between sympatric taxa will increase due to natural selection against hybridization (Howard 1993), which is often considered a final step in the process of speciation (Coughlan and Matute 2020). Thus, we hypothesize at contact zones, selection will act to further promote speciation.

Recombination is another important factor in considering the process of speciation. Genomic regions of restricted recombination in hybrids are expected to be associated with maintaining species despite gene flow (Butlin 2005; Ortiz-Barrientos et al. 2016). This may involve producing linkage disequilibrium (LD) along large swaths of the genome, including alleles conferring barriers to gene flow. HDRs of the genome between populations or species usually exhibit low recombination, whereas introgression always involves regions with high recombination (Burri et al. 2015; Wang et al. 2016; Han et al. 2017; Martin et al. 2019). Other factors, such as background selection, mutation rate variation, and evolutionary history, also cause elevated genomic divergence, especially for relative measures such as F_{ST} (Ravinet et al. 2017). It is necessary to either distinguish these factors or choose more suitable measures when identifying HDRs related to RI.

Here, we examine two species of toad-headed agama that readily hybridize, *Phrynocephalus vlangalii* and *P. putjatai*, which mainly occur across the northeastern Qinghai-Tibetan Plateau. A long-time disruption of gene flow dates to about 3.79–5.06 Ma due to geological movements and climatic change (Guo and Wang 2007; Jin et al. 2008; Jin and Brown 2013). Following the recent disappearance of the ancient Lake Gonghe ~150,000 years ago (Jin and Liu 2008), a contact zone was formed in the Gonghe Basin. Although the species exhibit some morphological differences in the contact zone, the genetic exchange has occurred (Jin and Liu 2008; Noble et al. 2010). This long-term divergence paired with a more recent admixture indicates that *P. vlangalii* and *P. putjatai* may be in the late stage of speciation, thus providing a model for studying how RI is maintained upon contact.

Given the clear population structure within these two species (Jin et al. 2008; Wang et al. 2009), we focus on populations in the contact zone. Although *P. vlangalii* has a scaffold-level reference genome based on Illumina short-read sequencing (Gao et al. 2019), a high-connectivity genome is necessary to explore the genomic landscape and driving forces of species persistence (Wolf and Ellegren 2017). We provide a high-quality chromosome-level genome for *P. vlangalii* and analyze whole-genome resequencing data from 41 individuals in the contact zone, 45 samples from allopatric populations of the two species, and one individual of *P. forsythii* as the outgroup (OG). We investigate the genomic patterns of introgression and HDRs of the genome between species in the contact zone. Further, we explore the genomic landscape and possible driving factors that promote the maintenance of species divergence and RI in the face of gene flow.

Results

Genome Sequencing, Assembly, and Annotation of *P. vlangalii*

In total, 184.32 Gb of PacBio long-read data were used for the *de novo* assembly of our genome using Falcon (Chin et al. 2016). A high-quality contig-level genome with a total size of 1.84 Gb and contig N50 length of 1.33 Mb (supplementary table S1, Supplementary Material online) was obtained after polishing the preliminary assembly using PacBio long-read Illumina short-read data. This genome size was close to the length estimated from the *k*-mer analysis (supplementary fig. S1 and table S2, Supplementary Material online). We located 3,959 contigs (88.31% of all contigs) on 24 pseudochromosomes using data from the Hi-C library (fig. 1, supplementary table S3, Supplementary Material online) and the anchored rate was 99.73% (supplementary table S1, Supplementary Material online). Finally, our chromosome-level genome was 1.84 Gb in size and had a scaffold N50 length of 94.69 Mb (supplementary table S1, Supplementary Material online). Compared with the previous assembly (Gao et al. 2019), our assembly yielded an almost 40-fold improvement in both contigs and scaffolds (contig N50: 1.33 Mb vs. 31.2 Kb, scaffold N50: 94.69 vs. 2.39 Mb). A total of 894.81 Mb of repetitive sequences was identified, comprising ∼48.64% of the genome (supplementary table S4, Supplementary Material online). Transposable elements

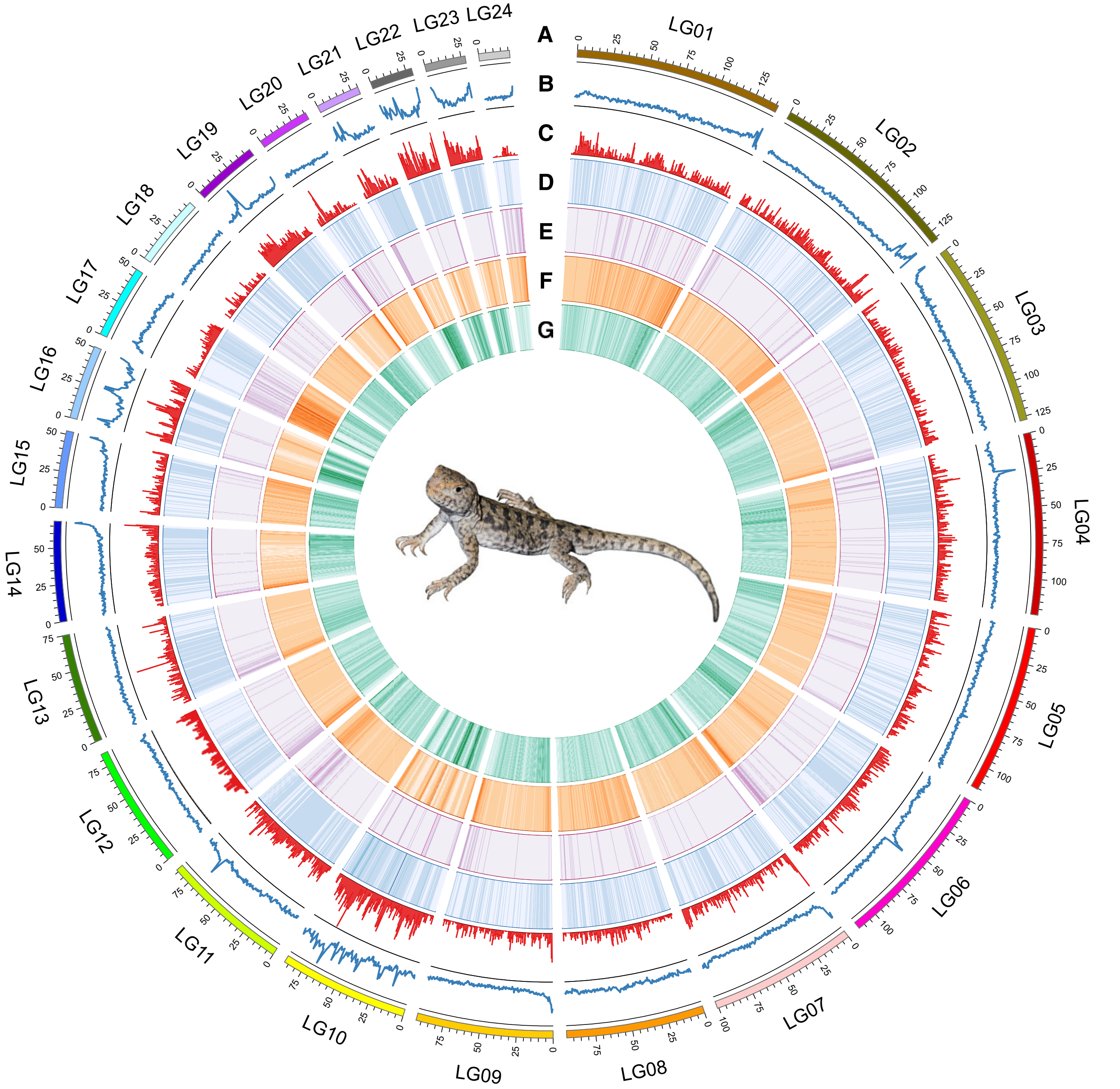

Figure 1. Chromosome-level genome and characteristics of *Phrynocephalus vlangalii*. (*A*) Lines represent 24 chromosomes; *x*-axis represents the position of each chromosome. (*B*) Distribution of the gene coding density. (*C*) Distribution of the GC content. (*D*) Distribution of DNA transposons (DNA). (*E*) Distribution of long terminal repeats (LTR). (*F*) Distribution of long interspersed nuclear elements (LINE). (*G*) Distribution of short interspersed nuclear elements (SINE).

(TEs) accounted for most of the repetitive sequences (864.02 Mb) and were found in 46.97% of the assembly (supplementary fig. S2 and table S5, Supplementary Material online). The Maker-P annotation pipeline (Cantarel et al. 2008) predicted 22,438 protein-coding genes, with an average transcript length of 36.36 Kb (fig. 1, supplementary fig. S3 and table S6, Supplementary Material online), which was much longer than the previous assembly (19.41 Kb; supplementary table S6, Supplementary Material online). More than 83% of the protein-coding genes were functionally annotated to at least one of Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO), SwissProt, TrEMBL, and nonredundant (NR) databases (supplementary fig. S4 and table S7, Supplementary Material online). In an assessment using the BUSCO database (Simão et al. 2015), our assembly retrieved more expected vertebrate genes (90%) than the previous assembly (85%), especially on the complete orthologue genes (79% vs. 61%; supplementary table S8, Supplementary Material online). Producing a high-quality genome facilitated our downstream analyses of genome-wide distribution patterns of HDRs, yielded an accurate recombination map, and enabled us to examine the genomic landscape of introgression.

Population Structure

To explore the genomic mechanisms of species persistence when facing gene flow, we sampled two species at the contact zone, including four representative sites of *P. putjatai* (GHput), and three sites of *P. vlangalii* (GHvla) near *P. putjatai*. To assess introgression in the contact zone, we also used samples from other regions according to previous phylogeographic studies (Jin and Liu 2008; Jin et al. 2014) to estimate the time, size, and genomic distribution of introgressed genes, and to discern the driver(s) of RI in the face of gene flow. For this, the samples of *P. putjatai* were collected from Guide (GD) and the samples of *P. vlangalii* were from Aksay (AKS), Madoi (MD), and Qaidam (CDM), where the relationship among them and GHvla was not clear (details in supplementary table S9, Supplementary Material online).

We resequenced 55 genomes of *P. vlangalii*, 31 samples of *P. putjatai*, and a single individual of *P. forsythii*. This included 16 individuals from GHvla (three sites) and 25 from GHput (four sites) at the contact zone (fig. 2*A* and *B*, supplementary table S9, Supplementary Material online). The average sequencing coverage was 12.99 ± 3.86× (mean ± SD; supplementary table S9, Supplementary Material online). After mapping all data to the chromosome-level genome of *P. vlangalii*, 51.13 million high-quality single-nucleotide polymorphisms (SNPs) were obtained for downstream analyses.

The phylogeny was reconstructed with 1,000 neutral regions using both coalescent-based species tree and concatenated methods. Both focal species clustered into distinct clades. Coalescent analysis obtained four clades of *P. vlangalii* (GHvla, AKS, CDM, MD), and two of *P. putjatai* (GHput, GD) (fig. 2*C*, supplementary fig. S5, Supplementary Material online). The topology of the concatenated ML tree was very similar to that of the species tree, except for one sample from GHput (supplementary fig. S6, Supplementary Material online). Generally, GHput and GD showed the least differentiation.

Our principal component analysis (PCA) separated *P. vlangalii* and *P. putjatai* along the first eigenvector, and explained 17.2% of total genetic variance, with the second and third eigenvector identifying different clades within each species explaining 10.2% and 8.8% of the variance, respectively (fig. 2*D* and *E*). Specifically, GHvla was closer to *P. putjatai* along the first eigenvector than the other clades of *P. vlangalii*, indicating possible admixture within the contact zone.

The results of the population genetic structure analyses using ADMIXTURE and FRAPPE were consistent with the phylogeny and the PCA (fig. 2*F*, supplementary figs. S7 and S8, Supplementary Material online). The optimal number of genetic clusters was four for all the samples of *P. vlangalii* ($K = 4$; GHvla, AKS, CDM, MD), and two for *P. putjatai* ($K = 2$; GHput, GD) (supplementary fig. S8, Supplementary Material online). Our phylogeny and genetic structure analyses indicated that samples within each region showed no obvious divergence, thus we subsequently treated samples from different regions as different populations (GHvla, AKS, CDM, MD, GHput, GD). There was also evident genetic contribution of *P. putjatai* within GHvla, indicating a recent asymmetric gene flow from *P. putjatai* to *P. vlangalii* (fig. 2*F*, supplementary fig. S7, Supplementary Material online).

Population Divergence, Characteristics, and Demographic History

Genome-wide divergence among six populations was measured using both absolute (Dxy) and relative (F_{ST}) methods. Interspecific average values of Dxy were slightly greater than those within species (supplementary figs. S9 and S10, table S10, Supplementary Material online). F_{ST} values showed a similar pattern (average F_{ST} of intraspecific vs. interspecific comparison: 0.23 vs. 0.32), but the maximum relative divergence occurred between MD and AKS of *P. vlangalii* (0.40; supplementary figs. S9 and S10, table S11, Supplementary Material online). This may have been due to the high values of F_{ST}, which can occur when a population has low genetic diversity, rather than large differentiation between populations. Regardless, lower divergence occurred between GHput and GHvla than between GHput and populations of *P. vlangalii* sampled from outside the contact zone using both methods (supplementary fig. S9, Supplementary Material online).

Mean nucleotide diversity (π) showed significant differences among six populations (Tukey HSD test, P-value < 0.0001, supplementary fig. S9 and table S12, Supplementary Material online), of which population GHvla exhibited the highest average nucleotide diversity (0.00344 ± 0.00127), followed by CDM of *P. vlangalii*

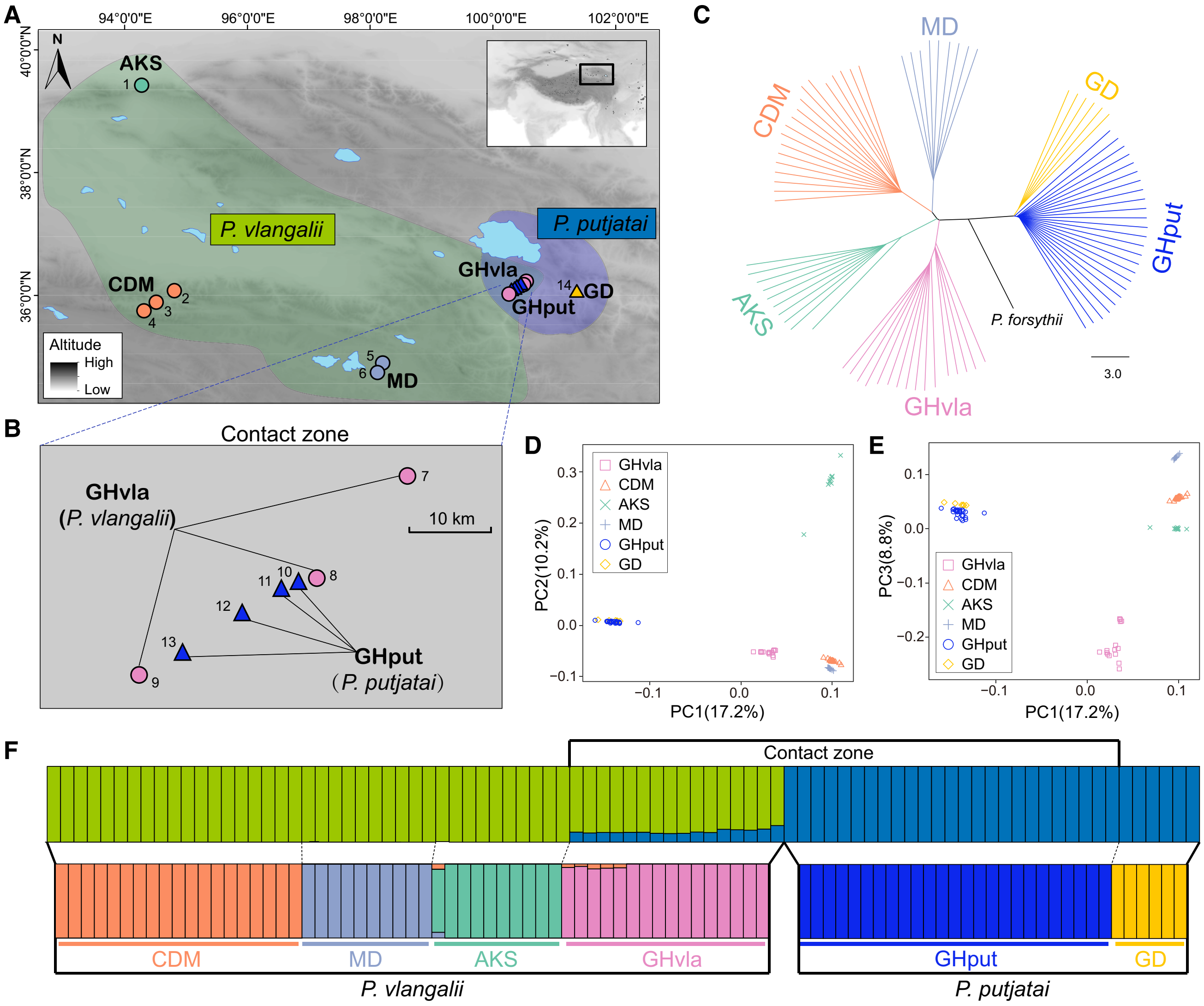

Figure 2. Geographic distribution and genetic structure of *Phrynocephalus vlangalii* and *P. putjatai*. (*A*) Geographic distribution of each species and sampling localities. Circles represent populations of *P. vlangalii*, and triangles *P. putjatai*. Number next to the dot is the sampling site in supplementary table S9, Supplementary Material online. (*B*) Enlargement of the contact zone. (*C*) Species tree constructed by MP-EST. (*D*) Distribution of PC1 and PC2 from the PCA analysis. (*E*) Distribution of PC1 and PC3 from the PCA analysis. (*F*) Population structure calculated using ADMIXTURE. Top panel represents the genetic structure of all samples at *K* = 2, which corresponds exactly to two species, and the bottom panel shows the genetic structure of the best *K* value (4 and 2, respectively) for each species. Colors denote species or populations.

(0.00290 ± 0.00181) and GHput of *P. putjatai* (0.00217 ± 0.00086). The average nucleotide diversity of population MD (0.00137 ± 0.00101) was the lowest (supplementary fig. S9 and table S12, Supplementary Material online). Both populations in the contact zone (GHvla and GHput) had negative mean Tajima's *D*-values, whereas the other populations had positive values (supplementary fig. S11, Supplementary Material online). The LD decay curve rates varied greatly among populations (supplementary fig. S9, Supplementary Material online). Population GHvla had the fastest decay rate and shortest decay distance, followed by population GHput (supplementary fig. S9, Supplementary Material online). Population MD had the longest decay distance, which exceeded 200 Kb.

Results from the PSMC analysis showed no obvious changes in the effective population size of GHput and GD over time (supplementary fig. S12, Supplementary Material online). There was a significant decrease in the effective population size of GHvla during the Kunlun Glaciation and minor fluctuations thereafter (supplementary fig. S12, Supplementary Material online). The effective population size of the other populations (AKS, CDM, and MD) of *P. vlangalii* also began to decline dramatically during the Kunlun Glaciation. A slight expansion occurred after the Zhonglianggan Glaciation in population AKS, followed by a decrease since the Early Stage of the Last Glaciation (supplementary fig. S12, Supplementary Material online).

Phylogenetic Discordance Implied Introgression upon Contact Zone

We explored population relationships across the genome using TWISST (Martin and Van Belleghem 2017), which quantifies the frequency or weighting of alternative topological relationships in sliding windows. Individuals from four populations were selected for each analysis, including GHput and GHvla, one allopatric population of *P. vlangalii* (AKS, CDM, and MD), and *P. forsythii* as the OG. We tested three topologies, including (OG, (GHput, (GHvla, AKS))), (OG, (GHvla, (GHput, AKS))), and (OG, (AKS, (GHvla, GHput))) when AKS was used as the allopatric parent population, and replaced AKS with the corresponding population when MD and CDM were used as parental populations (supplementary fig. S13, Supplementary Material online). The average weights of the "species topology" (T1) were 0.50, 0.57, and 0.45 when using populations AKS, CDM, and MD as the parent population of *P. vlangalii*, respectively, and those for the "geography topology" (T3) were 0.45, 0.39, and 0.52, respectively (supplementary fig. S13, Supplementary Material online). Topologies T1 and T3 had similar average weights and accounted for most of the tree weighting. From the perspective of the proportion of the windows with topology weighting of 1 (all samples conformed to a given tree shape), the geography topology (T3) was 0.33, 0.18, and 0.36, respectively, which was larger or slightly smaller than that of the species topology (T1; 0.19, 0.21, and 0.17, respectively; supplementary fig. S13, Supplementary Material online).

A heterogenetic and interlaced distribution pattern of different topology weightings across the genome showed that the species topology (T1) had the highest weighting in wide or narrow peaks on some chromosomes, whereas the geography topology (T3) had the higher weighting elsewhere in the genome (supplementary fig. S14, Supplementary Material online). For chromosomes LG17, LG10, LG19, and LG22, the average weighting of the geography topology was relatively higher across the entire chromosome. These results revealed large-scale phylogenetic discordance across the genome and many genomic regions supported the clustering of populations GHvla and GHput, further supporting the presence of genetic admixture upon contact between *P. vlangalii* and *P. putjatai*.

Admixture in the Contact Zone

D-statistics analyses of genetic introgression among populations found significant signals of admixture between GHput and GHvla from the contact zone (*Z*-score $> |4|$, supplementary table S13, Supplementary Material online), which was further supported by demographic analysis using G-PhoCS (supplementary fig. S15, Supplementary Material online). The introgressed proportions from GHput to GHvla measured using the F4 ratio were 0.134 $\pm$ 0.006, 0.187 $\pm$ 0.006, and 0.142 $\pm$ 0.005 when using populations AKS, CDM, and MD as the parent population of *P. vlangalii*, respectively (fig. 3*A*, supplementary table S14, Supplementary Material online). The iMAAPs analysis found a significant introgression signal around 1,000–2,000 generations ago (3,000–6,000 years ago at 3 years per generation; fig. 3*B*).

To quantify the proportions of admixture between populations GHput and GHvla across the genome, f_d, a window-based ABBA–BABA statistics, was applied for four populations as used in TWISST, including GHput, GHvla, one allopatric population of *P. vlangalii* (AKS, CDM, or MD), and *P. forsythii* as the OG. The QuIBL analysis indicated that increasing f_d-values were accompanied by increases in both internal branch lengths and non-ILS (non-incomplete lineage sorting) probabilities, and this correlation was significantly positive (supplementary fig. S16 and table S15, Supplementary Material online). A significant decrease in the absolute sequence difference (Dxy) was uncovered with increasing f_d-values (Tukey HSD test, *P*-value < 0.0001; supplementary fig. S17, Supplementary Material online). These results suggested that f_d reliably quantified introgression and differentiated it from shared ancestral variation.

Our analyses demonstrated a heterogeneous distribution of introgression, with a high proportion of introgression occurring across nearly the entire lengths of chromosomes LG17, LG10, LG19, and LG22 (fig. 3*C*–*E*). This was in accordance with the distribution pattern of the geography topology weighting. A relatively low level of introgression occurred at the middle of the chromosome and increased near the end, followed by a dramatic decrease at the very ends ($<$5% of the chromosome length) of the chromosome (supplementary fig. S18, Supplementary Material online). Considering that nearly all chromosomes of the two species of *Phrynocephalus* are telocentric (supplementary fig. S19, Supplementary Material online; Zeng et al. 1997; Wang et al. 2002), the low level of introgression at the very end may have been due to the centromere. Admixture proportions were significantly positively correlated with the population recombination rate of the admixed population GHvla (Spearman's $\rho = 0.356$, 0.358, and 0.367 with AKS, CDM, and MD as the parent population, respectively; *P*-value $= 0$; fig. 3*F*–*H*). A similar pattern occurred between population recombination rates and admixture proportions at different distances from the end of chromosomes (supplementary fig. S18, Supplementary Material online).

Fixed Differences Between *P. putjatai* and *P. vlangalii*

Considering the divergence time of these species, the density of fixed differences (d_f) between population GHput of *P. putjatai* and each population of *P. vlangalii* was estimated. The density of fixed differences corresponded to sites that were homozygous for one allele in GHput and homozygous for an alternative allele in one population of *P. vlangalii*, as measured by per site of available sequence data within each window. Because d_f is independent of within-species diversity, it reduced the amount of false-positive divergence detected between species when compared with F_{ST} measurements. Overall, the distribution of high-density or fixed differences between GHput and GHvla at the contact zone was sparser than that between GHput and other populations of *P. vlangalii* (fig. 4*A*).

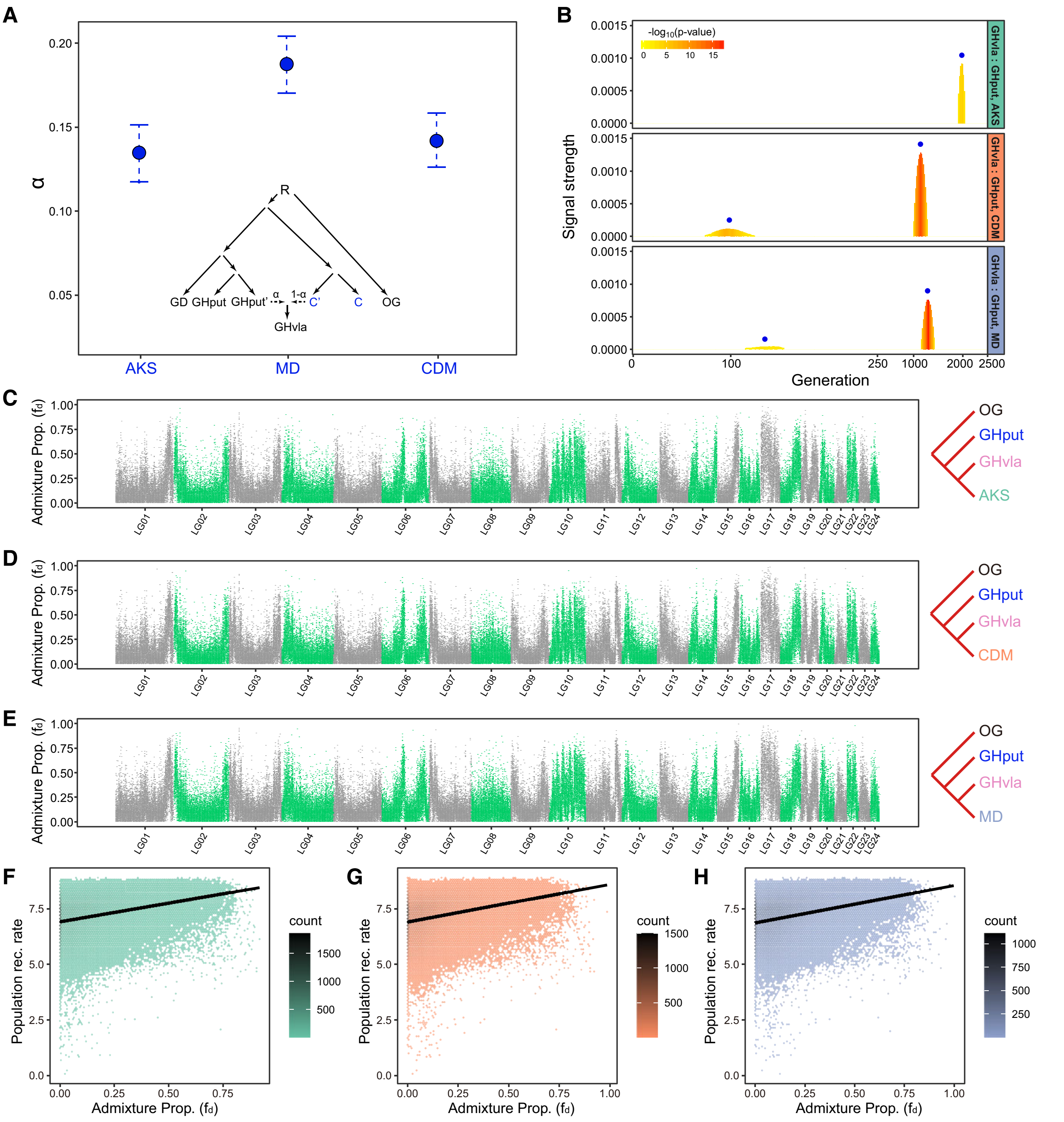

Figure 3. Admixture between populations GHvla and GHput of *Phrynocephalus*. (*A*) Introgression proportion between populations GHvla and GHput was inferred using the F4 ratio. α represents the proportion of introgression from GHput to GHvla. (*B*) Time of admixture was calculated using iMAAPs. Each generation of *Phrynocephalus* requires 3 years. (*C–E*) Distribution of admixture proportions along the entire genome calculated using f_d when the parent population of GHvla is population (*C*) AKS, (*D*) CDM, and (*E*) MD. Cladograms to the right of panels (*C–E*) shot the topology used to calculate admixture proportions. Sliding window size was 10 Kb. (*F–H*) Log2-transformed population recombination rates of population GHvla relative to admixture proportions calculated by f_d when the parent population of GHvla is population (*F*) AKS, (*G*) CDM, and (*H*) MD.

The mean density of fixed differences between GHput and GHvla (4.30e−05) was significantly smaller than that between GHput and the other three populations (AKS: 3.23e−04, CDM: 1.83e−04, and MD: 6.29e−04; Mann–Whitney *U* test, *P*-value = 0 for each of three comparisons; fig. 4*B*). Similar results were obtained using 10 randomly selected samples from each population of *P. vlangalii* to calculate d_f, which excluded effects of sample size variation among populations (supplementary fig. S20, Supplementary Material online).

We explored the potential effect that genetic admixture from GHput to GHvla had on the density of fixed differences between them. For this, 500 Mb of data without gene flow between GHput and GHvla was simulated by

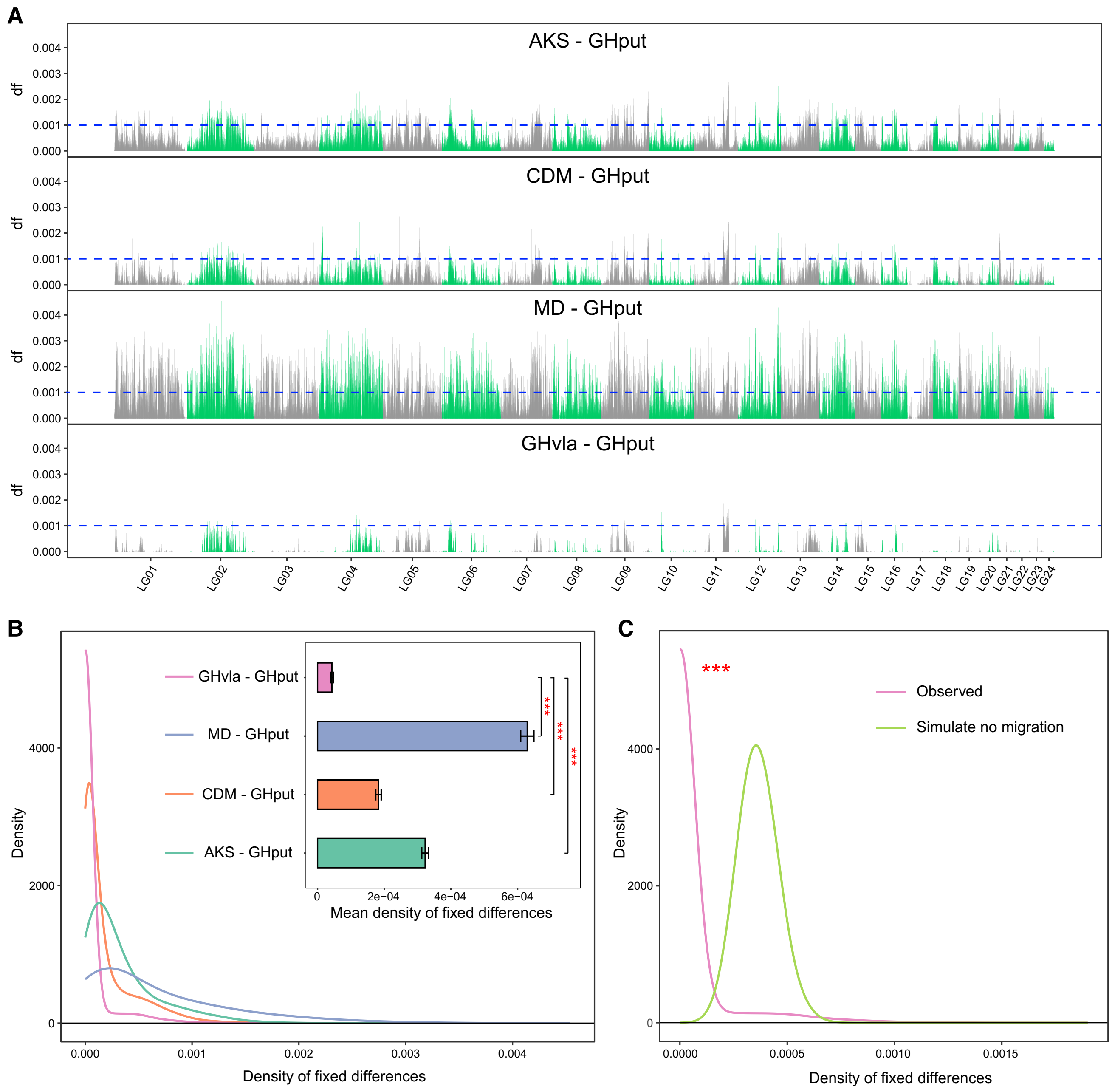

Figure 4. The density of fixed differences (d_f) between *Phrynocephalus vlangalii* populations and population GHput of *P. putjatai*. (*A*) Distribution of d_f between each population of *P. vlangalii* and population GHput along the whole genome. (*B*) Density distribution and comparison of d_f between each population of *P. vlangalii* and population GHput. Error bars present ± 3 CI. (*C*) Comparison of d_f between simulated and observed data. Here, d_f compares populations GHvla and GHput. Simulated data were obtained from msms assuming no gene flow between populations GHvla and GHput. Red asterisks indicate *P*-values tested by the Mann–Whitney *U* test < 0.001. Sliding window size is 50 Kb.

msms. The simulated data had a significantly larger mean density of fixed differences between GHput and GHvla than the empirical data (3.64e−04 vs. 4.30e−05, Mann–Whitney *U* test, *P*-value = 0; fig. 4C). This indicated that introgression after contact significantly reduced the divergence between GHvla and GHput.

Identification and Characteristics of HDRs

Windows with a mean density of fixed differences between GHput and each population of *P. vlangalii* that was >0.001 were treated as HDRs of the genome following Ellegren et al. (2012). HDRs between GHput and GHvla occupied 116 windows, accounting for 0.32% of the entire genome. In contrast, between GHput and AKS, CDM, MD, there were 2,093 (5.70% of the genome), 615 (1.67% of the genome), and 8,383 (22.82% of the genome) windows in HDRs, respectively. A total of 2,267 windows comprised HDRs between GHvla and the allopatric population of *P. putjatai* (GD), which accounted for 6.17% of the entire genome. Thus, the contact zone had fewer windows with HDRs and a smaller genome proportion of HDRs (fig. 5A and C). HDRs between GHput and GHvla exhibited significantly smaller continuous distribution lengths than HDRs

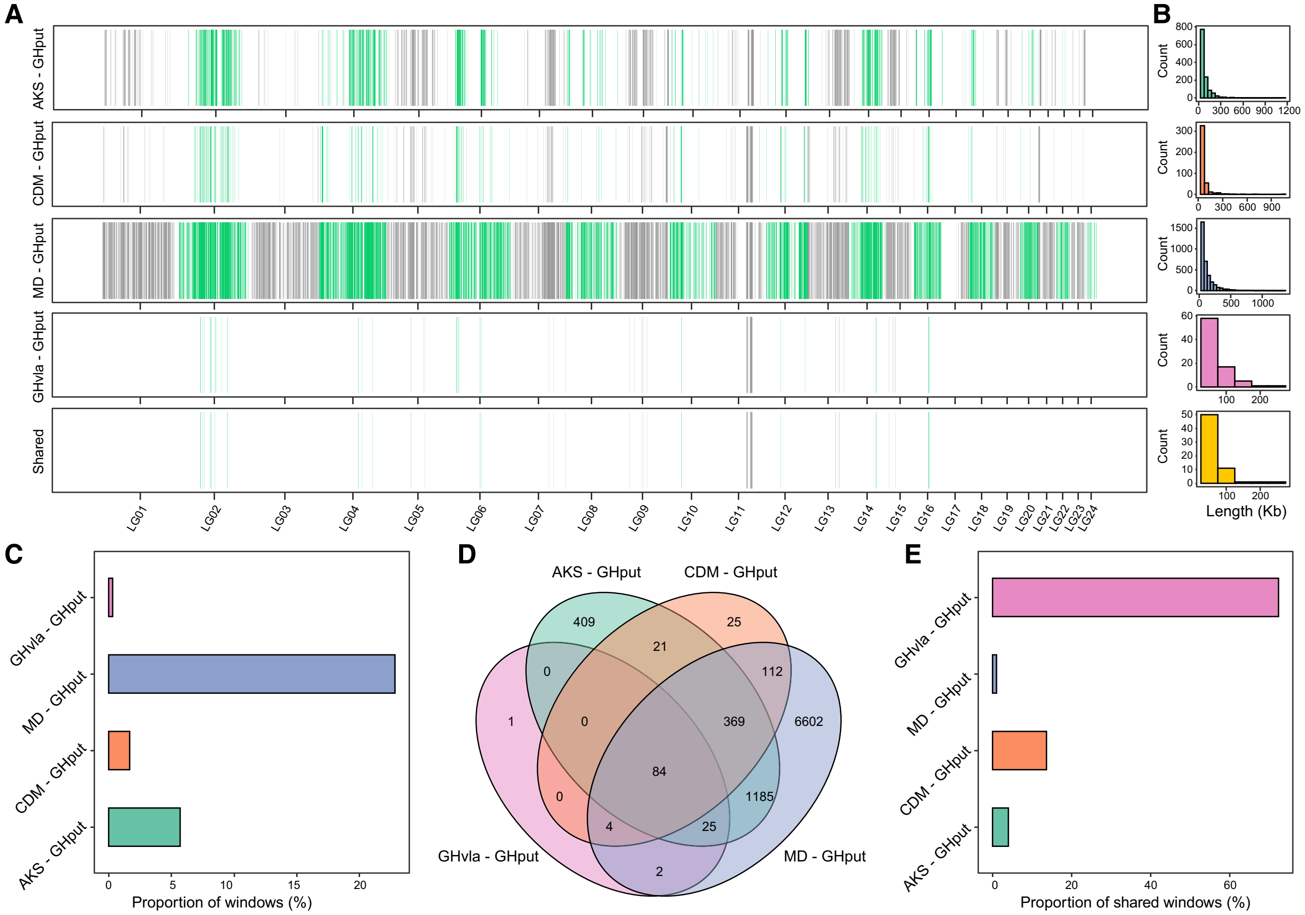

Figure 5. HDRs between populations of *Phrynocephalus vlangalii* and population GHput of *P. putjatai*. (*A*) Distribution of HDRs between each population of *P. vlangalii* and population GHput along the whole genome; (*B*) length distribution of HDRs between each population of *P. vlangalii* and population GHput; (*C*) window proportions of HDRs between each population of *P. vlangalii* and population GHput to the whole genome; (*D*) shared windows of HDRs between each population of *P. vlangalii* and population GHput; (*E*) proportion of shared HDRs windows between each population of *P. vlangalii* and population GHput to that of each population. The sliding window size is 50 Kb.

between GHput and allopatric populations of *P. vlangalii* (fig. 5*B*). Accordingly, HDRs of the two populations reduced and fragmented after contact. Comparisons of HDRs between GHput and four populations of *P. vlangalii* found a significantly higher proportion of shared HDRs windows for GHvla and the remaining three populations than expected by random chance (Fisher's exact test, GHvla vs. AKS: *P*-value = 4.94e−127; GHvla vs. CDM: *P*-value = 3.15e−133; GHvla vs. MD: *P*-value = 7.83e−73; fig. 5*D*). Thus, the HDRs may have formed in the ancestral population of *P. vlangalii*. The shared windows of HDRs in the four comparisons occupied 84 windows, accounting for 0.22% of the genome. Shared HDRs accounted for 72.41% of the total HDRs between the two populations from the contact zone, which was dramatically higher than compared with the other three allopatric populations (4.01%, 13.65%, and 1.00%; fig. 5*E*). Similar patterns were also observed when using the same sample size for all populations of *P. vlangalii* (supplementary fig. S21, Supplementary Material online).

To assess the genotype variation within *P. vlangalii* on the effect of the shared HDRs, we compared the position and genotype of the locus (SNP site) among the populations of *P. vlangalii* that showed fixed differences with GHput. In the shared HDRs, a total of 5,174 fixed difference loci between GHvla and GHput were identified. This number was slightly higher among allopatric populations, which contained 6,692 (AKS), 6,035 (CDM), and 7,755 (MD) (supplementary fig. S22, Supplementary Material online). Among them, 3,941 loci were shared among the four populations of *P. vlangalii* and 99.92% of them exhibited identical genotypes, which were contained in 76% of all fixed differences loci in GHvla (supplementary fig. S22, Supplementary Material online). In addition, comparisons of fixed differences loci of four populations of *P. vlangalii* found a significantly higher proportion of shared loci with identical genotypes for GHvla and the other three populations than expected at random (Fisher's exact test, GHvla vs. AKS: *P*-value = 8.99e−298; GHvla vs. CDM: *P*-value = 0; GHvla vs. MD: *P*-value = 1.11e−130; supplementary fig. S22, Supplementary Material online). Thus, despite low intraspecific variation, most of the fixed difference loci of the shared HDRs possessed identical genotypes among the four populations, suggesting that intraspecific genotypic variation hardly influences HDRs being shared within *P. vlangalii*.

The shared HDRs had a heterogeneous distribution pattern across the whole genome; all of them were located only on chromosomes LG02, LG04, LG05, LG06, LG07, LG09, LG10, LG11, LG12, LG13, LG14, LG15, and LG16, including the longest and eight shortest chromosomes (figs. 5A and 6A). Compared with the genomic background (GB), the HDRs exhibited significantly higher absolute divergence (Dxy; Mann–Whitney *U* test, *P*-value = 5.02e−31; figs. 6D and 7A, supplementary table S16, Supplementary Material online) and relative (F_{ST}; Mann–Whitney *U* test, *P*-value = 3.65e−51; figs. 6C and 7B, supplementary table S16, Supplementary Material online),

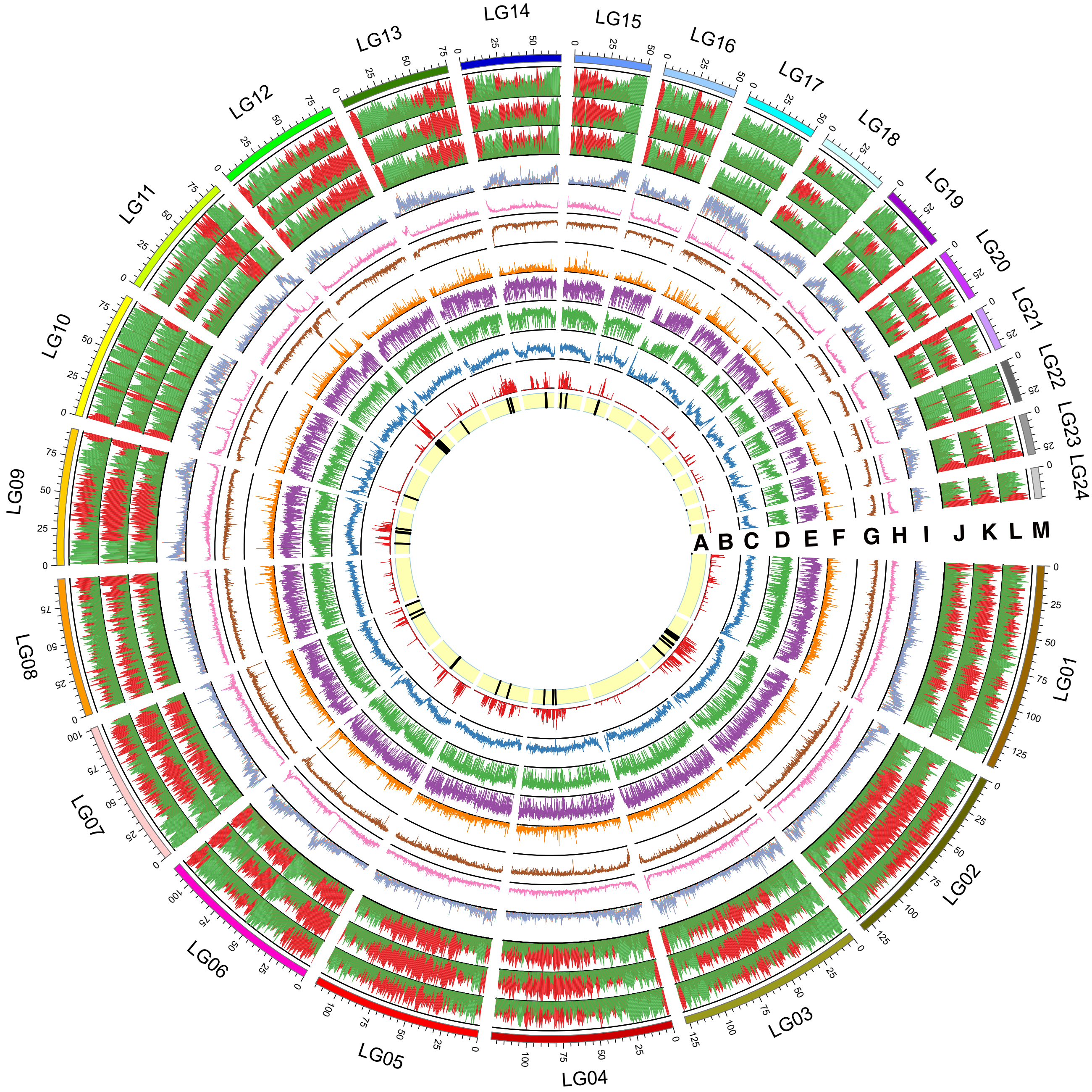

Figure 6. Genomic distribution characteristics of shared HDRs between *Phrynocephalus vlangalii* and *P. putjatai*. Genomic distribution of shared HDRs (*A*), the density of fixed differences between populations GHput and GHvla (*B*), F_{ST} between GHput and GHvla (*C*), Dxy between GHput and GHvla (*D*), nucleotide diversity of GHvla (*E*), XP-EHH between GHput and GHvla (*F*), the log2-transformed population recombination rate of GHvla (*G*), mean linkage coefficient of GHvla (*H*), admixture proportion between GHput and GHvla calculated using f_d (*I*), and topology weighting when AKS (*J*), CDM (*K*), MD (*L*) is used as the parent population of the GHvla, respectively. (*M*) Twenty-four chromosomes of the *P. vlangalii* genome. Green, orange, and blue lines in (*I*) represent the f_d-values calculated with AKS, CDM, and MD as the parent population of GHvla. Red area in (*J*–*L*) is consistent with T1 in supplementary fig. S13, Supplementary Material online, indicating the tree shape of the species tree; green area is consistent with T3 in supplementary fig. S13, Supplementary Material online, indicating the geographical tree structure. The sliding window size is 50 Kb.

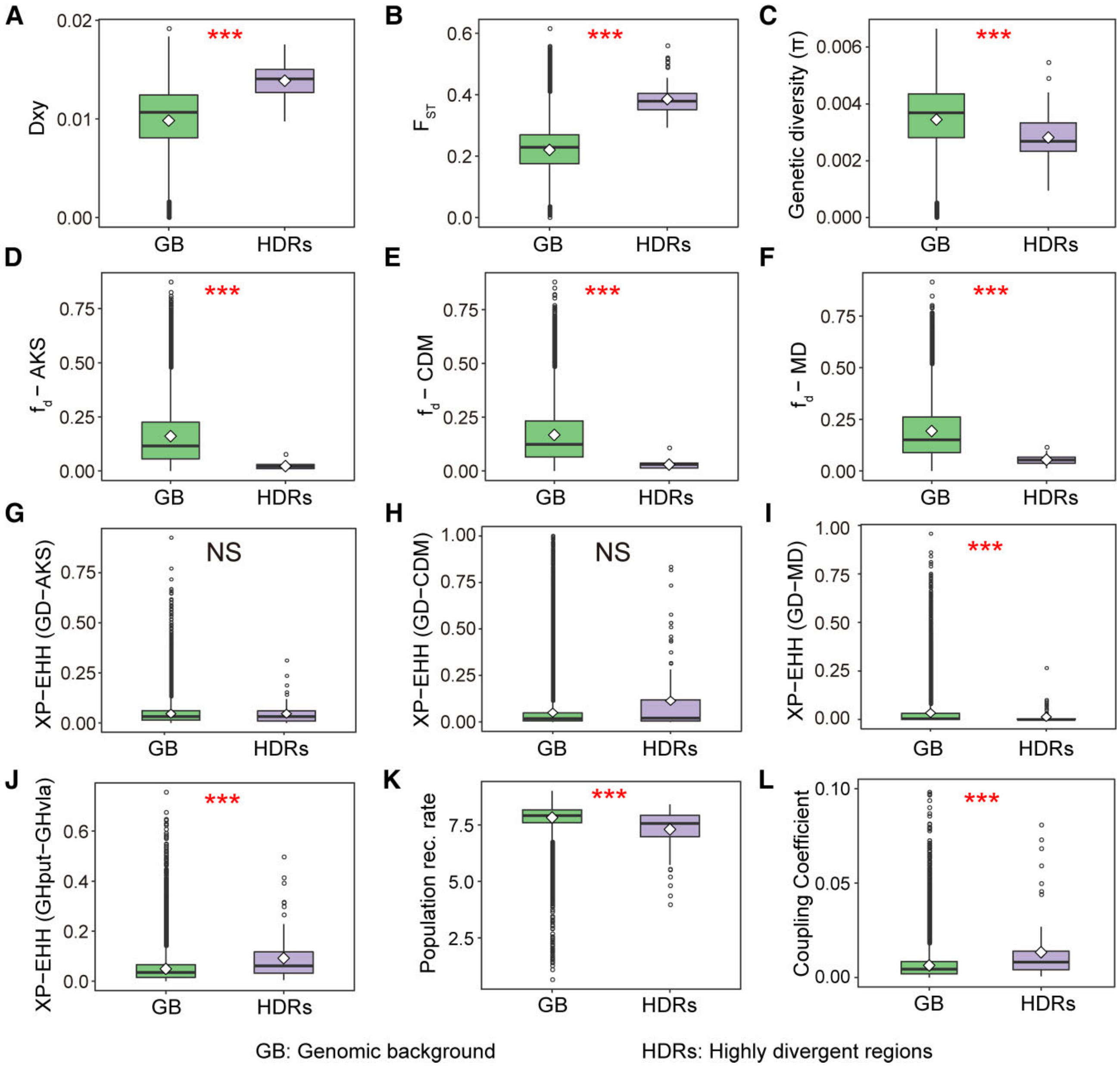

Figure 7. Comparison of genetic characteristics between shared HDRs and genomic background for *Phrynocephalus*. (*A*) Dxy between populations GHput and GHvla. (*B*) F_{ST} between GHput and GHvla. (*C*) Nucleotide diversity of GHvla. (*D–F*) Admixture proportion between GHput and GHvla calculated using f_d with AKS, CDM, and MD as the parent population of GHvla, respectively. The proportion of all extreme XP-EHH scores between GD and AKS (*G*), CDM (*H*), MD (*I*), respectively. (*J*) The proportion of all extreme XP-EHH scores between GHput and GHvla. (*K*) Log2-transformed population recombination rate of GHvla. (*L*) Coupling coefficient of divergent selection and recombination, the outliers larger than 0.1 are not shown. The difference of each comparison was tested by the Mann–Whitney *U* test. Red asterisks indicate the *P*-value was significantly <0.001 and NS indicate the *P*-value was larger than 0.05. The sliding window size is 50 Kb.

along with significantly lower nucleotide diversity (*π*; Mann–Whitney *U* test, *P*-value = 3.53e−10; figs. 6*E* and 7*C*, supplementary table S16, Supplementary Material online). This indicated that the divergence did not happen recently (Ravinet et al. 2017). Further, the lower proportions of introgression (f_d) of HDRs (Mann–Whitney *U* test, *P*-value = 1.93e−24, 1.52e−26, and 4.77e−27, respectively; figs. 6*I* and 7*D–F*, supplementary table S16, Supplementary Material online) were consistent with the higher weighting of the species topology (T1; Mann–Whitney *U* test, *P*-value = 1.62e−38, 6.03e−37, and 1.45e−37, respectively; fig. 6*J–L*, supplementary fig. S23 and table S16, Supplementary Material online), and lower weighting of the geography topology (T3; Mann–Whitney *U* test, *P*-value = 1.66e−37, 9.08e−36, and 4.13e−36, respectively; fig. 6*J–L*, supplementary fig. S23 and table S16, Supplementary Material online). This suggested that genetic admixture was less likely to occur in the HDRs.

To further explore how divergence was maintained in the face of gene flow, we first scanned for signals of selection across the whole genome using the cross-population extended haplotype homozygosity method (XP-EHH), which estimates the intensity of recent selection for selected pairs of populations (Sabeti et al. 2007). A significantly higher proportion of extreme XP-EHH scores was detected in HDRs compared with the GB between the two populations from the contact zone (GHput vs. Ghvla; Mann–Whitney *U* test, *P*-value = 4.21e−08; figs. 6*F* and 7*J*, supplementary table S16, Supplementary Material online). We did not observe this pattern in comparisons of allopatric populations of both species (GD vs. AKS, GD vs. CDM, GD vs. MD), indicating no selection signal of HDRs between allopatric populations (fig. 7*G*–*I*, supplementary table S16, Supplementary Material online). Thus, these HDRs appeared to have been subjected to strong recent divergent selection after contact, which constituted reinforcement. The HDRs also had a significantly decreased population recombination rate (ρ, Mann–Whitney *U* test, *P*-value = 1.06e−09; figs. 6*G* and 7*K*, supplementary table S16, Supplementary Material online) and increased mean linkage coefficient (Mann–Whitney *U* test, *P*-value = 3.45e−14; fig. 6*H*, supplementary fig. S24 and table S16, Supplementary Material online) in the admixed population GHvla. To account for the effects of local N_e on recombination, we divided ρ by genetic diversity (π) following Wang et al. (2016) and compared scaled ρ (ρ/π) between HDRs and the rest of the genome. We detected significantly suppressed scaled recombination of HDRs relative to the GB in GHvla (supplementary fig. S25 and table S16, Supplementary Material online). Targeted LD analysis showed significantly increased LD in the contact zone relative to allopatric populations (mean linkage coefficient: 0.18 vs. 0.14, Mann–Whitney *U* test, *P*-value = 3.57e−15). Taken together, the HDRs exhibited a significantly higher coupling coefficient (Mann–Whitney *U* test, *P*-value = 9.59e−12; fig. 7*L*, supplementary fig. S25 and table S16, Supplementary Material online), which indicated the maintenance of divergence upon contact. To test whether the low recombination and higher coupling in the HDR region were due to the centromere, we analyzed the relative position of HDRs to the centromere on chromosomes. Because most chromosomes are telocentric (Zeng et al. 1997; Wang et al. 2002), we used the relative distance from the end of the chromosome as a measure of the distance to the centromere. Most of the HDRs had a relative distance to the chromosome ends between 0.4 and 0.7, with very few HDRs having distances between 0 and 0.2 (supplementary fig. S26, Supplementary Material online). Thus, the vast majority of HDRs occurred at higher distances from the centromere, and thus the centromere did not affect their low recombination and higher coupling rates.

To test for the potential role HDRs play in asymmetric introgression, we examined the allele frequencies of HDRs in GHput and GHvla using XP-EHH. A significant difference was observed for the proportion of negative extreme scores (Mann–Whitney *U* test, *P*-value = 1.31e−05; supplementary fig. S27 and table S16, Supplementary Material online), but not for the positive ones (Mann–Whitney *U* test, *P*-value = 0.49; supplementary fig. S27 and table S16, Supplementary Material online). This suggested a significantly higher proportion of high frequency or fixed alleles in GHput remained polymorphic in GHvla than in the GB. This observed pattern could have driven the asymmetric gene flow from GHput to GHvla.

Genomic and Geographical Cline in the Contact Zone

We used genomic and geographical cline analyses to further assess admixture and selection on admixed genomes. Datasets for these analyses were the GB, shared HDRs, and highly introgressed regions (HIRs). GHvla had a relatively high hybrid index at the genome-wide level (fig. 8*A*), which was consistent with our STRUCTURE analysis. Loci subject to divergent selection were associated with lower fitness in heterospecific GB and were less likely to be introgressed than non-selected loci. The genomic-cline analysis was used to assess recent selection against hybridization, wherein parameter v described the departure of introgression from the genome average, and elevated values were consistent with incompatibilities (Knief et al. 2019). HDRs had a significantly steeper cline (larger v) than the GB (fig. 8*B*; Mann–Whitney *U* test, *P*-value = 4.36e−04), which supported divergent selection. In contrast, HIRs had a significantly smoother cline (smaller v) than the GB (fig. 8*B*; Mann–Whitney *U* test, *P*-value = 1.16e−53), demonstrating that these regions were more susceptible to genetic exchange.

For the geographical cline analysis, we defined locality 10 (the locality nearest to GHvla in GHput) as the tentative center and estimated geographical distances (in km) of each locality from locality 10. For *P. vlangalii* localities, distances were expressed as negative values, and positive for those containing *P. putjatai* (supplementary fig. S28 and table S17, Supplementary Material online). Data sets GB and HIRs best-fit model I, and HDRs model II (supplementary table S18, Supplementary Material online). The GB generated a modest cline (width of 25.9 km), which was centered between the two species (at −7.4 km; approximately midway between localities 10 and 8; fig. 8*C* and *D*). A smooth and broad cline (width of 92.2 km) was identified in HIRs and its center shifted toward GHvla relative to the GB cline (at −21.7 km, near locality 8; fig. 8*C* and *D*). Thus, the HIRs were more likely introgressed to GHvla in the contact zone. By contrast, the HDRs produced a sharply narrow, step-like cline (width of 1.0 km) and its center was estimated at −0.8 km (near locality 10; fig. 8*C* and *D*). Thus, barrier loci under divergent selection confer a cost to hybrids that prevents unabated dispersion across the contact zone; as expected, they displayed steeper clines and reduced width (Barton and Hewitt 1985; McKenzie et al., 2015). These results also correspond with evidence from the XP-EHH and genomic-cline analyses, supporting divergent selection against hybridization in the contact zone.

Functional Annotation of HDRs

Functional annotation of the shared HDRs identified 112 genes, 89 of which were annotated with known functions

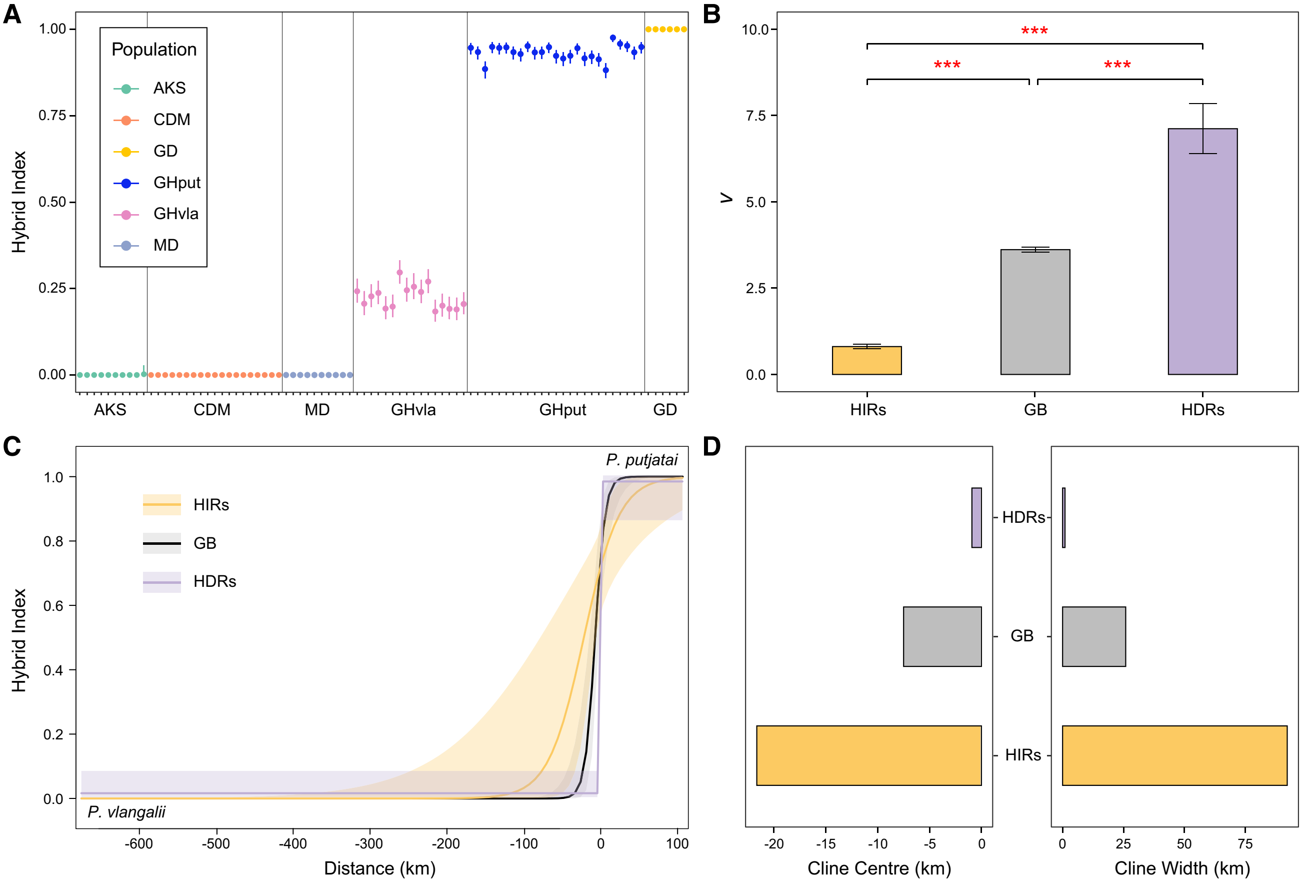

Figure 8. Genomic and geographic cline analyses. (*A*) Hybrid index of each sample on the genome-wide level. Dots indicate the average hybrid indices and vertical lines indicate 95% confidence intervals. (*B*) Statistics of the genetic cline rate *v*. Higher values indicate steeper clines. Error bars present ± SE. Red asterisks indicate the *P*-value was significantly <0.001. (*C*) Comparisons of three geographic clines for hybrid indices. ML clines and observed average hybrid indices per sample location shown. Shading indicates 95% confidence intervals. *X*-axis indicates geographical distances (in km) of each locality from locality 10 (in fig. 2*A* and *B*). For localities of *Phrynocephalus vlangalii*, distances were expressed as negative values, and for those of *P. putjatai* as positive values. (*D*) Corresponding mean cline centers and cline widths (in km). GB, genomic background; HDRs, highly divergent regions; HIRs, highly introgressed regions.

in the reference genome of *P. vlangalii*. The Metascape and KOBAS enrichment analyses of these genes identified those involved in the estrogen receptor signaling pathway (supplementary tables S19 and S20, Supplementary Material online). This pathway involves many functions, including the intracellular estrogen receptor signaling pathway (GO: 0030520), regulation of intracellular estrogen receptor signaling pathway (GO: 0033146), with genes *ESR1*, *CARM1*, and *KANK2* involved. Both methods detected genes, such as *NCAPD2*, *TEX11*, *SYCE3*, and *ZNF318*, enriched in functions related to meiosis, though without an identical GO term (supplementary tables S19 and S20, Supplementary Material online). Pathways related to estrogen receptors and meiosis play important roles in sexual reproduction. In addition to the functional enrichment, several other genes related to reproduction were identified, such as sperm formation, the fertilization process, infertility, and meiosis (details in table 1). Because many genes involved in reproduction were located in the HDRs between the two species, we suspect that these regions were involved in the evolution of RI between *P. putjatai* and *P. vlangalii*.

Discussion

During allopatric speciation, genetic differentiation accumulates during spatial isolation. If contact reoccurs, populations that have not developed a mechanism for complete RI often have the opportunity to experience gene flow (Abbott et al. 2013). Lizards of the genus *Phrynocephalus* can provide insight into the maintenance of genetic barriers to gene flow, promoting species persistence, particularly during the late stage of speciation.

Highly divergent genes that are associated with reproduction can play an important role in the maintenance of RI. Barriers to genetic exchange, such as the evolutionary of RI, are most likely to develop within species that are distributed allopatrically, although most barriers may have developed through background selection, selective sweeps, or genetic drift (Feder et al. 2013). The maintenance of species boundaries in *P. putjatai* and *P. vlangalii* when they are in contact involves many genes in the HDRs that have functions involved in reproduction, such as sperm formation, the fertilization process, infertility, and meiosis (table 1, supplementary tables S19 and S20,

Table 1. List and Detail Functions of Genes Related to Reproduction in the Shared HDRs of the Genome.

Gene symbol	Chromosome	Function	Reference
TCP11	LG02	Encodes the receptor of fertilization promoting peptide (FPP), acts on spermatogenesis and sperm function, which may be responsible for the sperm tail morphology and motility. May be important in fertilization.	Ma et al. (2002); Liu et al. (2011)
TGIF2	LG02	May participate in spermiogenesis and folliculogenesis.	Hu et al. (2011)
ESR1	LG04	Mediate the physiological responses to estrogens during sperm production. Plays a prominent role in successful spermatogenesis and fertility. Polymorphisms in the *ESR1* may have differential roles in the predisposition to male infertility.	Corbo et al. (2007); Giwercman (2011); Gunawan et al. (2011); Ge et al. (2014)
ZNF318	LG04	May act as a transcriptional regulator during spermatogenesis and, in particular, during meiotic division.	GeneCards (https://www.genecards.org/)
PRSS21	LG07	Also known as testisin, a proteolytic factor that directs epididymal sperm cell maturation and sperm-fertilizing ability. Lacking *PRSS21* can reduce sperm capacity to penetrate the cumulus to reach the oocyte and fertilize.	Yamashita et al. (2008); Netzel-Arnett et al. (2009); Zhou et al. (2012)
LDLR	LG11	Involved in steroidogenesis.	Ye et al. (2014)
MRC1	LG11	Affects minisatellite stability during meiosis.	LeClere et al. (2013)
NCAPD2	LG11	Play pivotal roles in chromosome assembly and segregation during both mitosis and meiosis.	Yuan et al. (2019)
RAD23A	LG11	Associated with male infertility.	Rockett et al. (2004)
SYCE3	LG11	Involved in physically linking homolog pairs to complete proper segregation of homologous chromosomes during meiosis. Loss of *SYCE3* leads to infertility.	Schramm et al. (2011)
CETN1	LG16	Plays an essential role in the late steps of spermiogenesis and spermatid maturation. Germline deletion of *CETN1* causes infertility.	Avasthi et al. (2013)
OGT	LG16	OGT and O-GlcNAcylation in general are needed for germ cells meiosis.	Olivier-Van Stichelen et al. (2012)
TAF1	LG16	*TAF1*, together with testis-specific TBP-associated factors, regulate the transcription of genes necessary for spermatocyte entry into meiosis.	Metcalf and Wassarman (2007)
TEX11	LG16	Modulates germ cell proliferation and is essential for meiosis and male fertility.	Wang et al. (2001); Tang et al. (2011); Yu et al. (2012)

Supplementary Material online). Genes associated with reproduction have also diverged during the speciation process in other groups, such as *Anopheles gambiae* (Lawniczak et al. 2010) and *Nanorana parkeri* (Wang et al. 2018). They will result in abnormal reproduction, and thus genetic isolation. In other taxa, the genes that play a role in speciation may associate with assortative mating characteristics, such as skin coloration in neotropical cichlid fishes (Kautt et al. 2020), feather coloration in seedeaters (Turbek et al. 2021), wing patterning in *Heliconius* butterflies (Martin et al. 2013), and also with adaptive phenotypes, such as beak shapes in Darwin's finches (Lamichhaney et al. 2015).

The ability of species that have diverged over long periods of time to exchange genes is typically restricted upon contact. Asymmetric introgression between *P. putjatai* and *P. vlangalii* appears to have occurred 3,000–6,000 years ago following a long period of geographical isolation (Guo and Wang 2007; Jin et al. 2008; Jin and Brown 2013), rather than consisting of ongoing, and random gene flow (figs. 2*F* and 3*B*). The asymmetric pattern of this introgression may be due to biased backcrossing (Takahashi et al. 2017), wherein hybrid offspring tend to backcross with the parental species *P. vlangalii*, and/or prezygotic or postzygotic isolation between hybrid offspring and *P. putjatai*. Prezygotic isolation can result in hybrid offspring producing gametes that are not compatible with those of *P. putjatai*, and postzygotic isolation or postzygotic isolation may involve backcross-sterility or backcross-lethality with *P. putjatai* via Dobzhansky–Muller (DM) incompatibilities (Muller 1942; Dobzhansky 1982). If so, more genes related to reproduction should remain conserved in GHput, but exhibit some diversity in GHvla. We found a significantly higher proportion of alleles in HDRs have risen to high frequencies or fixation in GHput than the genomic background, but remain polymorphic in GHvla (fig. S27 and table S16, Supplementary Material online). Our hypothesis predicts this result to some extent. However, only future behavioral and hybridization experiments can ultimately test this hypothesis.

Gene flow plays an important role in shaping the heterogeneous distribution of genetic differentiation. HDRs of the genome involved in RI appear to be shielded from gene flow, whereas other regions exhibit some homogenization by gene flow and correspondingly show lower levels of differentiation. However, several genome-level investigations have suggested that gene flow was not a major factor in shaping the genomic landscape of differentiation or the formation of genomic islands (Renaut et al. 2013; Cruickshank and Hahn 2014; Burri et al. 2015; Han et al. 2017). In our lizards, the mean divergence, window number, and continuous length of HDRs between the two species at the contact zone are significantly lower than that from comparisons of allopatric *P. putjatai* and *P. vlangalii* (figs. 4*A*,*B* and 5*A*–*C*). Simulated data without introgression also obtain significantly higher divergence between sympatric populations (fig. 4C). These findings indicate that gene flow is likely responsible for the reduced genetic differentiation of *P. putjatai* and *P. vlangalii* in contact and that lower proportions of introgression in HDRs may promote the maintenance of genomic divergence related to RI (Poelstra et al. 2014).

Divergent selection and variation in recombination rates may maintain RI of species upon contact. The strength of the antagonism between selection and recombination will determine the extent to which the species and RI persist in the face of gene flow (Felsenstein 1981; Abbott et al. 2013). Speciation rarely involves a single-locus barrier to gene flow. Thus, associations between more barrier loci and their closely linked loci produce much more stable genomic divergence. These associations could be maintained and strengthened by the selection, but shattered by recombination (Nosil et al. 2009; Abbott et al. 2013; Ortiz-Barrientos et al. 2016). In admixed population GHvla, HDRs have significantly lower population recombination rates compared with the remaining portions of the genome (figs. 6G and 7K, supplementary table S16, Supplementary Material online). The positive correlation between the proportion of introgression and relative population recombination rate suggests that variation in recombination mediates the heterogeneous distribution of introgression (fig. 3F–H; Edelman et al. 2019; Martin et al. 2019). In our case, the role that selection plays in this process is apparent in the significantly higher proportion of extreme XP-EHH values of HDRs occurring at the contact zone (figs. 6F and 7J, supplementary table S16, Supplementary Material online), but not between comparisons of allopatric populations between *P. putjatai* and *P. vlangalii* (fig. 7G–I, supplementary table S16, Supplementary Material online). Thus, strong, recent divergent selection pressure, and reinforcement in the contact zone may counter the effects of hybridization. Both GHvla and GHput in the contact zone have negative mean Tajima's *D*-values (supplementary fig. S11, Supplementary Material online), they do not appear to have experienced significant population size expansion (supplementary fig. S12, Supplementary Material online), and both have much steeper genomic and geographic clines of HDRs than the GB (fig. 8). These findings indicate the effects of selection. The balance between selection and recombination can determine the strength of the association among diverging loci and the barrier they pose to genetic homogenization. Strong coupling of selection and recombination maintains associations, resulting in a strong barrier to homogenization, whereas weak coupling causes barrier loci to act independently, making population isolation less likely to persist (Felsenstein 1981; Barton 1983; Butlin 2005; Abbott et al. 2013). Thus, strong and recent divergent selection paired with low recombination at loci related to RI appears to maintain both species of *Phrynocephalus* upon contact. This mechanism can drive the persistence of newly generated species.

It is important to distinguish between primary and secondary contact. In the former case, HDRs reflect accumulated RI, yet in the latter case, they reflect either maintenance or heterogeneity in disrupting RI (Feder et al. 2013). The contact zone in the Gonghe Basin of *P. putjatai* and *P. vlangalii* was reported to be secondary contact, but with little evidence (Jin and Liu 2008). If this is a primary contact zone, then gene flow should have been continuous since the species' initial divergence. However, introgression in the contact zone is a recent event (fig. 3B), which rejects the hypothesis of the primary contact. A more precise distinction between the two scenarios requires new analytical methods, but this remains a major challenge (Strasburg and Rieseberg 2013).

Materials and Methods

Sequencing, Assembly, and Annotation of the Genome of *P. vlangalii*

To obtain a high-connectivity genome, which is necessary to explore the genomic landscape and driving forces of species persistence, we first performed sequencing, chromosome-level assembly, and annotation of the genome of *P. vlangalii*.

(1) Sampling, library construction, and sequencing of the *P. vlangalii* genome

A single male *P. vlangalii* was collected in July 2018 from Madoi, Qinghai, China. After euthanasia, seven tissues (liver, muscle, heart, lung, brain, skin, and testis) were excised and stored in liquid nitrogen.

High-quality genomic DNA was extracted from muscle tissue using a Qiagen Genomic DNA extraction kit (Qiagen, Valencia, CA, USA). DNA concentration was measured using a Qubit Fluorometer. Sample integrity and purity were assessed using agarose gel electrophoresis. For long-read sequencing, portions of the DNA were used to construct circular consensus sequencing libraries with a fragment size of 20 kb using the SMRTbell template preparation kit (Pacific Biosciences, Menlo Park, CA, USA). We then sequenced this library using the PacBio Sequel system (Pacific Biosciences) with 26 cells. The raw reads were filtered, and only high-quality data were used for further genome assembly. Other portions of genomic DNA were prepared for short-read sequencing which was then used to estimate genome size and perform error correction of the assembled genome. A paired-end library with short insert sizes of about 270 bp was constructed and sequenced using Illumina HiSeq X Ten System with paired read lengths of 150 bp. SOAPnuke (v1.5.3; Chen et al. 2018) software was used to filter out adapters and low-quality data.

We then applied the Hi-C (high-throughput chromosome conformation capture) technique to assist the chromosome-level assembly. Chromatin in the muscle samples was cross-linked to DNA and fixed with formaldehyde in a concentration of 1%. Fixed samples were used for constructing Hi-C libraries following the standard Hi-C library protocol. We assessed the insert size and concentration of the libraries using an Agilent 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA) and the ABI StepOnePlus Real-Time PCR System, respectively. The libraries were then sequenced on an Illumina HiSeq X Ten platform to generate paired-end reads that were 100 bp in length. SOAPnuke (v1.5.3) was used to trim adapters and filter out low-quality data. We performed additional quality control on the Hi-C data using HiC-Pro (Servant et al. 2015).

To assist in gene annotation of the genome for *P. vlangalii*, PacBio ISO-Seq was used on all dissected tissue types. Total RNA was extracted and purified from each of the seven tissues using TRIzol (Invitrogen Corp., Carlsbad, CA, USA) and an RNeasy Mini Kit (Qiagen, Chatsworth, CA, USA). A NanoDrop and an Agilent 2100 Bioanalyzer were used for qualifying and quantifying these total RNA. We then mixed equal volumes of the RNA extracted from each tissue into a single pooled RNA sample. Total RNA was synthesized to first-strand cDNA using Clontech SMARTer PCR cDNA Synthesis Kit. A large-scale PCR was performed to synthesize second-strand cDNA after PCR optimization. The obtained cDNA was used for SMRTbell library construction, which was sequenced using the PacBio Sequel platform. To obtain consensus full-length isoforms, we performed SMRT analysis on high-quality sequencing data.

(2) Chromosome-level assembly and annotation of the genome of *P. vlangalii*

A *k*-mer depth frequency distribution analysis of the short-read sequencing data was performed to estimate the genome size and heterozygosity of the genome. The *k*-mer distribution was measured by KMERFREQ_AR (Luo et al. 2012) with $k = 17$ and 21. The genome size (G) was estimated with the formula $G = Nk\text{-mer}/Dk\text{-mer}$, where Nk-mer is the total number of k-mers and Dk-mer is k-mer depth of the homozygous peak.

The primary *de novo* assembly of the genome was carried out using Falcon (v0.7; Chin et al. 2016) using the long reads produced by PacBio Sequel system and then polished by SMRT Link (v5.0; https://www.pacb.com/support/software-downloads/). After this, Pilon (v1.2.4; Walker et al. 2014) was further applied to polish and improve the genome assembly again using all filtered short reads. To perform chromosome-level assembly of the genome, Juicer (v1.9.9) and 3D de novo assembly (3D-DNA; v1.9) pipelines (Durand et al. 2016; Dudchenko et al. 2017) were used to cluster, order, and orient the contigs to the chromosome-level scaffolds based on valid reads of Hi-C.

TEs and tandem repeats of the *P. vlangalii* chromosome-level genome were annotated using two strategies. Firstly, known repeats were predicted with the homology prediction method. These TEs were annotated using RepeatMasker (v4.08) and RepeatProteinMask (v4.08; Tarailo-Graovac and Chen 2009) based on the RepBase TE library (https://www.girinst.org/repbase/). Secondly, a *de novo* prediction method was used to discover novel repeats. We used RepeatModeler (http://www.repeatmasker.org/RepeatModeler/; v1.08) to build a *de novo* repeat database and then annotated TEs with this database using RepeatMasker. We used Tandem Repeat Finder (v4.07; Benson 1999) to annotate the tandem repeats portions of the genome. Finally, we estimated the divergence of a repeat copy from its consensus sequence with RepeatMasker.

Integrated methods were used to annotate the position and structure of protein-coding genes. We aligned full-length transcripts produced by PacBio ISO-Seq to the genome using Gmap (v2017-09-11; Wu and Watanabe 2005). SMRT Link was then applied to filter the alignments with the criteria that the coverage must be ≥0.99 and similarity ≥0.85. Protein sequence data of *Homo sapiens*, *Gallus gallus*, and *Anolis carolinensis* were downloaded and aligned to our genome using Exonerate (Slater and Birney 2005). We then randomly selected 2,000 genes with complete structure, which were obtained by homology-based prediction with the above three species, to train Augustus (v3.3; Stanke et al. 2006) and SNAP (v2006-07-28; Korf 2004) for the first time. The protein sequence data of *H. sapiens*, *G. gallus*, and *A. carolinensis*, well-aligned full-length transcripts by PacBio ISO-Seq, and annotation file of repeats were imported into Maker-P pipeline (Cantarel et al. 2008) for preliminary annotation. We then randomly selected 2,000 genes with an AED value <0.1 from the output of Maker-P to train Augustus and SNAP for the second time. Next, we repeated the Maker-P pipeline as above to obtain the final annotated gene sets.

We used five different public protein databases to perform gene functional annotation using Blast+ (v2.29; McGinnis and Madden 2004), including Kyoto Encyclopedia of Genes and Genomes (KEGG; Ogata et al. 1999), GO (Ashburner et al. 2000), TrEMBL (Bairoch and Apweiler 2000), SwissProt, and the NCBI NR protein sequence database (Deng et al. 2006). InterProScan (v5.30-69.0; Zdobnov and Apweiler 2001) was used to identify motifs and domains by searching against the public databases, which were SMART (Schultz et al. 2000), PRINTS (Attwood et al. 2000), ProSiteProfiles (Hulo et al. 2007), PANTHER (Mi et al. 2005), ProDom (Corpet et al. 1999), Pfam (Bateman et al. 2000), and ProSitePatterns (Sigrist et al. 2012). Finally, we integrated all the above results to obtain the final functional annotation of the *P. vlangalii* gene set.

Three types of noncoding RNAs (ncRNAs), including rRNAs, miRNAs, and snRNAs, were identified by aligning the genome of *P. vlangalii* to the Rfam database (Griffiths-Jones et al. 2005). Final ncRNAs were obtained by filtering with Infernal (v1.1; Nawrocki and Eddy 2013). The tRNAs were predicted by tRNAscan-SE (v2.0; Lowe and Eddy 1997) based on sequence structure.

Whole-Genome Resequencing and SNP Calling of Population Genomes

To obtain the population genomic data for exploring the potential genomic mechanisms of species or RI persistence, we firstly performed sampling, whole-genome resequencing, and SNP calling.

(1) Sampling and whole-genome resequencing of population genomes

All samples were originated from the northeastern Qinghai-Tibet Plateau (fig. 2A). We sampled a total of 55 individuals of *P. vlangalii* and 31 individuals of *P. putjatai* (fig. 2A and B, supplementary table S9, Supplementary Material online). These included a fine-scale sampling of 16 and 25 individuals from the contact zone for *P. vlangalii* (three sites) and *P. putjatai* (four sites), respectively (fig. 2A and B, supplementary table S9, Supplementary Material online). One individual of *P. forsythii* was used as the OG taxon. Detailed information on all samples, including sampling locations, is provided in supplementary table S9, Supplementary Material online. All collections were made according to animal use protocols approved by the Kunming Institute of Zoology Animal Care and Ethics Committee with the number SMKX-20160301-03.

Following euthanasia, liver tissues from the above lizards were subsampled and stored in 95% ethyl alcohol. Total genomic DNA was extracted using the phenol/chloroform method. Subsequently, we used 1–3 µg of DNA to prepare one paired-end library per individual with an insert size of 300–800 bp. Libraries were sequenced using the Illumina HiSeq X Ten platform with a read length of 150 bp. Most individuals were sequenced to a target depth of 10×, with one individual from each population and OG taxa sequenced to a target depth of 15×. Detailed information on sequencing coverage can be found in supplementary table S9, Supplementary Material online.

(2) Genome mapping, SNP calling, and filtering

We used BWA-MEM (0.7.12) with default settings (Li and Durbin 2009) to map the raw sequence reads of each individual to the *P. vlangalii* chromosome-level reference genome. SAMtools (v1.3.1; Li et al. 2009) was used to convert SAM files to BAM files. The "sort" and "rmdup" options within SAMtools were used to detect and remove PCR duplicates. To enhance the alignments in regions of insertion-deletion polymorphisms, local realignment around indels was performed using the GATK (Genome Analysis Tool Kit v3.5; DePristo et al. 2011). The variant discovery was performed according to all the realigned BAM files using the "mpileup" option in SAMtools with the parameters "-C 50 -q 20 -Q 20 -uDf".

To obtain high-quality SNPs for downstream analysis, variants that met the following criteria were removed: sites shorter than 5 bp away from the indels; sites with a quality score below 40; sites with triallelic alleles and indels; sites with an overall sequencing depth <2.5% and >97.5%; sites occurring in fewer than 85% of individuals; and sites with a minimum allele frequency (MAF) lower than 0.01.

Population Structure, Characteristics, and Demographic History

(1) Population structure

Because a clear population structure was required for subsequent analyses, we used the following approaches:

Phylogenetic relationships were reconstructed by both concatenation and coalescent-based methods. SNPs located in the repeat region were filtered out to reduce potential alignment errors before phylogenetic reconstruction. To avoid capturing regions with strong natural selection, putatively neutral genomic regions that were >10 Kb from exons were obtained. Of these regions, 1,000 loci with a window size of 100 Kb were randomly selected and used for subsequent analyses. We first concatenated the loci and reconstructed phylogenies using the maximum-likelihood (ML) framework implemented in RAxML (v8.2.12; Stamatakis 2014) assuming the GTR + GAMMA model and with 100 rapid bootstrap replicates to infer support values for each node. Secondly, for the coalescent-based method, individual gene trees were reconstructed using RAxML with the GTR + GAMMA model. Following this, MP-EST (v1.6; Liu et al. 2010) was used to generate a species tree with each population as the tips of the tree.

We investigated population genetic structure using PCA and a model-based clustering approach. The PCA was performed on all SNPs using GCTA (v1.24.4; Yang et al. 2011). ADMIXTURE (v1.23; Alexander et al. 2009) and FRAPPE (v1.0; Tang et al. 2005) were used for ancestry estimation of each individual. Both methods were used also to detect admixture between GHvla and GHput. For these analyses, we first obtained genotype data (in PLINK PED format) from VCF files using VCFtools (v0.1.13; Danecek et al. 2011). Then PLINK (v1.90b6.17; Purcell et al. 2007) was used to generate BED files with PED files using the parameter flags "–make-bed" (for PCA and ADMIXTURE) and "–recode12" (for FRAPPE), respectively. To account for the effects of LD, SNPs with an interval of 50 Kb were selected for all population structure analyses. Two strategies were used depending on the purpose. Firstly, we set the possible ancestral clusters number (K) to 2 for all samples to explore the admixture between *P. vlangalii* and *P. putjatai*. We additionally estimated the population structure for each species with K set from 2 to 5 and 7, respectively. For both strategies, the maximum number of expectation-maximization iterations for each ancestral cluster was set to 10,000.

(2) Intrapopulation and interpopulation summary statistics

We calculated several intrapopulation and interpopulation summary statistics to compare genomic differences 1) between populations in the contact zone and allopatric populations, and 2) between HDRs and genomic background. This clarified the genetic exchange and differentiation of populations in the contact zone.

Nucleotide diversity and Tajima's *D* for each population were calculated using VCFtools (v0.1.13) with the parameter flags "–window-pi" and "–TajimaD", respectively. The sliding window size was set to 50 Kb. We estimated genome-wide LD decay using PopLDDecay (v3.4.0; Zhang et al. 2019), setting the maximum distance between two SNPs to 1,000 Kb. Window-based LD was also measured by PLINK (v1.90b6.17) with the parameters "–r2 –ld-window-r2 0 –ld-window 1000 –ld-window-kb 50" to obtain the linkage coefficient of individual SNPs. The mean linkage coefficient for each 50 Kb window was calculated through an in-house Perl script. We calculated windows-based hierarchical *F*-statistics (F_{ST}) as implemented in VCFtools (v0.1.13) for all possible pairwise comparisons among populations with a window size of 50 Kb (with parameter "–F_{ST}-window-size"). Negative values of F_{ST} were treated as 0. Custom Perl scripts were used to calculate the absolute sequence divergence (Dxy; Ai et al. 2015) based on SNPs averaged to obtain window-based estimates. The window size was set to 50 Kb and raw results were standardized by the total number of available sites per window.

(3) Population recombination rate

To explore the potential role recombination played in the distribution of introgression across the genome and the maintenance of RI in the contact zone, we estimated the population recombination rate (ρ). Beagle (v5.0; Browning and Browning 2007) was used to phase the filtered SNPs, and phased data were then as input into the *FastEPRR_VCF_step1* function in FastEPRR (Gao et al. 2016) to scan the input and store required information into files for each 10 and 50 Kb window (with parameters inSNPThreshold = 30 and qualThreshold = 20). Next, *FastEPRR_VCF_step2* was used to estimate the recombination rate for each window. Finally, we applied *FastEPRR_VCF_step3* to merge the files generated by step 2 for each chromosome. Raw population recombination rates were normalized using the log2-transformed method. The distribution pattern of population recombination rates at different distances from the end of chromosomes was explored further.

(4) Demographic history

To exclude the influence of population history on the detection of natural selection, we estimated the trajectories of demographic histories for all the populations using the PSMC (Pairwise Sequentially Markovian Coalescent) approach (Li and Durbin 2011). To avoid high false-negative rates at low sequence coverage of PSMC, we restricted this analysis to the individual with the highest coverage (≥15) in each population. The PSMC analysis was set as the following parameters: -N25 -t15 -r5 -b -p "4 + 25 * 2 + 4+6", together with 100 bootstrapping replicates. A generation time of 3 years and a neutral mutation rate of 1.4e−09 per site per year were used to scale the raw results. Demographic history was also inferred using the G-PhoCS (Generalized Phylogenetic Coalescent Sampler, v1.2.3; Gronau et al. 2011) to provide the population genetic parameters for subsequent simulations. Accordingly, 1,000 neutral loci with a length of 100 Kb were selected to estimate the above parameters of all four populations of *P. vlangalii* and population GHput of *P. putjatai*. Migration scenarios were added by combining the results of the *D*-statistic tests, ADMIXTURE, and FRAPPE. We ran each Markov chain for 2,000,000 generations and sampled parameter values every 20th iteration. TRACER (v1.7.1; Rambaut et al. 2018) was used to determine burn-in and convergence of each run. More information about the control file of G-PhoCS is provided in the supplementary text, Supplementary Material online.

Admixture in the Contact Zone Between *P. vlangalii* and *P. putjatai*

It was important to determine the extent of introgression in the contact zone to understand the maintenance of genomic integrity. Our ADMIXTURE and FRAPPE analyses indicated asymmetric introgression at the contact zone from population GHput of *P. putjatai* to population GHvla of *P. vlangalii*. To obtain more detailed admixture information on the contact between these two species, we investigated introgression in terms of time, proportion, and distribution across the genome.

(1) Time and proportion of the introgression in the contact zone

D-statistic tests based on the four-taxon test were performed to confirm the admixture among populations with the qpDstat procedure in the ADMIXTOOLS package (v5.0; Patterson et al. 2012). An absolute *Z*-score value of ≥4 was considered significant. The software iMAAPs (v1.0.0; Zhou et al. 2017) was used to calculate the admixture time using default parameters. We set GHvla to be the admixture population, GHput as one of the reference populations, and populations AKS, CDM, and MD as the others, respectively. The analysis was performed chromosome-by-chromosome with a time horizon from 1 to 1,000,000 generations. Admixture proportions from populations GHput to GHvla were estimated using the qpF4ratio procedure in ADMIXTOOLS. We selected four populations and one OG with a phylogenetic relationship shown in fig. 3A, and set populations AKS, CDM, and MD as the parent population (labeled in C) of *P. vlangalii*, respectively. Then we computed the admixture proportion α (proportion from GHput population) in the formula: $\alpha = $ f4 (GD, OG; C, GHvla)/f4 (GD, OG; C, GHput) as described in Patterson et al. (2012).

(2) Distribution of introgression across the genome in the contact zone

To examine the genomic landscape of the introgression between *P. vlangalii* and *P. putjatai* in the contact zone, we estimated admixture in windows through tree- and *D*-statistic-based methods. Four populations were selected for each of these analyses, including OG, GHvla, GHput, and one allopatric population of *P. vlangalii*.

Tree-based TWISST (v0.2; Martin and Van Belleghem 2017) was used to quantify genealogical relationships among populations in sliding windows. Firstly, we constructed an ML tree for each 50 Kb non-overlapping window using IQ-TREE (v1.6.12; Nguyen et al. 2015) with the GTR + G model and 1,000 ultrafast bootstraps. Based on these trees, we used TWISST to calculate the topology weighting by iterative sampling of subtrees. The "species topology" referred to the topology where population GHvla clustered with the allopatric populations of *P. vlangalii*, which reflected its phylogenetic history. The tree topology that clustered GHvla and GHput together was considered the "geography topology", which implied the influence of admixture. We then calculated the mean topology weighting and the proportion of the windows with topology weighting 1 for each topology.

ABBA–BABA statistics in sliding windows, which is a *D*-statistic-based method, was used to estimate admixture proportions based on f_d (Martin et al. 2015). Unlike TWISST, this SNP-based method combines *D*-statistics and *f* estimators to calculate the genome-wide fraction of admixture. The script *parseVCF.py* (https://github.com/simonhmartin/genomics_general; v0.3) was used first to phase the VCF file, and then admixture proportions were estimated using the *ABBABABAwindows.py* script. We used window sizes of 10 and 50 Kb and minimum good sites per window of 10. The windows with an f_d-value smaller than 0 or larger than 1 were removed. Further analysis using QuIBL (Edelman et al. 2019) was applied to assess the reliability of detecting introgression versus ILS. As required, we randomly selected one sample for each population and constructed an ML tree in sliding windows using IQ-TREE (v1.6.12) with the GTR + G model and 1,000 ultrafast bootstraps. All trees were used as input files of QuIBL to obtain the relative ratio of introgression and ILS across the whole genome with the parameter "likelihoodthresh" set to 0.01. Finally, we used the *perlocus_formatter.py* script to calculate internal branch lengths and the non-ILS probabilities for each window and tested relationships with the admixture proportions calculated using f_d.

To investigate the potential role recombination plays in the distribution of introgression across the whole genome, we performed a correlation analysis between the population recombination rate and admixture proportions in sliding windows using Spearman's correlation coefficient method.

Analysis of Highly Divergent Genomic Regions between *P. vlangalii* and *P. putjatai*

To explore the genomic mechanisms of the maintenance of RI in the contact zone, we analyzed the characteristics, selection pressure, and functional annotation of HDRs.

(1) Identification of HDRs

To identify the HDRs between the two species, we first calculated the density of fixed differences (d_f; Ellegren et al. 2012) in 50 Kb sliding windows between GHput and each population of *P. vlangalii*. To determine if admixture would reduce divergence between *P. putjatai* and *P. vlangalii*, we compared the differences of mean d_f between sympatric and allopatric populations of *P. vlangalii* with GHput. In addition, we simulated 500 Mb data using msms (v3.2rc Build:162; Ewing and Hermisson 2010) with the same parameters that were estimated by G-PhoCS with the exception that we removed gene flow between GHput and GHvla in the contact zone. The mean d_f between GHput and GHvla for simulated and empirical data was compared and significance was assessed using the Mann–Whitney *U* test.

Windows with a d_f larger than 0.001 were considered as HDRs following Ellegren et al. (2012). We compared the number, distribution, and continuous length of the HDRs of the four populations of *P. vlangalii*. We assessed the proportion of shared HDRs between two of the four populations to determine if it was significantly higher (Fisher's exact test) than expected. The HDRs shared by all four populations were considered as the most representative of highly divergent genomic regions between *P. putjatai* and *P. vlangalii* and these were used for further analysis. To assess the genotype variation within *P. vlangalii* of the shared HDRs, we analyzed the position and genotype of the locus (SNP site) among the populations of *P. vlangalii* that showed fixed differences with GHput. The fixed different loci with identical genotypes among the four populations were identified and tested to determine if they were higher than expected using Fisher's exact test.

(2) Genomic characteristics and natural selection of HDRs

We explored differences in the distribution patterns of genetic characteristics and statistics between the shared HDRs and genomic background including F_{ST}, Dxy, XP-EHH, admixture proportions (f_d) between populations GHput and GHvla, nucleotide diversity (π), population recombination rate (ρ), local LD of GHvla, and the topology weighting of both the "geography topology" and "species topology". To further explore the potential role of recombination in the maintenance of RI in the face of gene flow, we also performed targeted LD analysis by comparing the linkage coefficient of HDRs between contact zone populations and allopatric populations. Significance tests were performed using the Mann–Whitney *U* test.

We used the XP-EHH (Sabeti et al. 2007) to detect recent selective sweeps as implemented in selscan (v1.2.0a; Szpiech and Hernandez 2014). For this, we estimated divergent selection between GHvla and GHput in the contact zone by setting GHvla as the first (nonref) population and GHput as the second (ref) population. To test for the role reinforcement played in the contact zone, we compared the strength of divergent selection in sympatric and allopatric populations. We also estimated divergent selection between allopatric GD versus AKS, GD versus CDM, and GD versus MD by always setting GD as the second (ref) population. We defined all SNPs' distances according to approximate recombination maps then calculated and compared the XP-EHH score between population comparisons within each chromosome using the default parameters (selscan –xpehh –hap < pop1 haps > –ref < pop2 haps > –map –out). Raw extreme scores (≥2 by default) were normalized using windows of the constant size of 50 Kb with varying numbers of SNPs. Windows <50 Kb were removed from the above analyses.

We quantified the antagonism between selection and recombination by calculating a relative coupling coefficient as the proportion of extreme XP-EHH scores between GHput and GHvla divided by the population recombination rate of GHvla, and then compared the difference between the shared HDRs and genomic background. Mann–Whitney *U* test was used to perform significance tests for the comparisons. To test if the centromere influenced recombination and coupling coefficients in the HDR region, we analyzed the relative position of HDRs to the centromeres on the chromosomes. Most chromosomes of both species are telocentric, and thus the centromere was at the end of the genome (supplementary fig. S19, Supplementary Material online; Zeng et al. 1997; Wang et al. 2002). Therefore, we calculated the relative distance from HDRs to chromosome ends using a scale from 0 to 1, where 0 indicated the end of the chromosome and 1 indicated the farthest from the end (middle of the chromosome).

To test the potential role HDRs played in asymmetric introgression, we explored allele frequencies of HDRs in GHput and GHvla. XP-EHH was used because it could identify and statistically assess alleles that were nearly fixed or fixed in one population but remained polymorphic in another. GHvla was set as the first population and GHput as the second; a positive score represented a nearly or fixed allele in GHvla, and the negative score represented the same in GHput. We counted the proportion of positive and negative extreme values (≥2) separately with a sliding window of 50 Kb and compared them in HDRs with the genomic background. Mann–Whitney *U* test was used for significance tests.

(3) Functional annotation of HDRs

Based on the annotated genome of *P. vlangalii*, we extracted genes partially or completely located in HDRs using the “intersect” option within BEDtools (v2.29.0; Quinlan and Hall 2010). Functional enrichment was determined using Metascape (v3.5; Zhou et al. 2019) and KOBAS (v3.0; Bu et al. 2021). Hypergeometric test and Benjamini–Hochberg *P*-value correction algorithm were conducted using Metascape to identify ontology terms that contained a statistically greater number of genes than expected by chance, and where Fisher’s exact test, χ^2 test, and the Binomial test were used in KOBAS. To assure reliability, we chose a strict threshold for *P*-value (<0.01) and only the GO terms that passed the threshold in both software and all tests were considered to be significantly enriched. We also performed gene annotation by searching for related biological functions on public gene functional databases and published studies.

Genomic and Geographical Cline Analyses

Genomic and geographical cline analysis further assessed introgression and selection in the contact zone. Genomic clines used the hybrid index to detect loci that may be subject to selection or introgression. Representative SNPs sets included the following: (1) genome-wide level or GB, which consisted of randomly selected one SNP per 500 Kb sliding window; (2) HDRs, which randomly selected one SNP in each window of shared HDRs that was highly differentiated (top 1% sites of F_{ST}) between the allopatric populations of *P. putjatai* and *P. vlangalii*; and (3) HIRs with a randomly selected SNP in each window of HIRs (top 1% windows of f_d) that was highly differentiated (top 1% sites of F_{ST}) between the allopatric populations of *P. putjatai* and *P. vlangalii*. Genomic-cline models were fitted for each SNP set using the gghybrid R package (v1.0.0; Bailey 2020) including all individuals, discarding the first 5,000 iterations as burn-in, and estimating parameters and posterior probabilities from subsequent 5,000 MCMC iterations. Bayesian hybrid index and parameter v of Fitzpatrick’s logit-logistic genomic-cline were estimated, where v was always positive and higher values indicated steeper clines.

A geographic cline analysis assessed asymmetric gene flow and assessed whether the shared HDRs had a signal indicating divergent selection. The Metropolis-Hastings MCMC algorithm implemented in the R package HZAR (Derryberry et al. 2014) was used to fit a series of geographical cline models to the hybrid index of the above three datasets. To collapse sample localities on a one-dimensional axis, we defined locality 10 (the locality nearest GHvla in GHput) as the tentative center and estimated geographical distances (in km) of each locality from locality 10. For *P. vlangalii* localities, distances were expressed as negative values and those of *P. putjatai* as positive values (supplementary fig. S28 and table S17, Supplementary Material online). Three commonly used models (I–III; Brumfield et al. 2001; McKenzie et al. 2015) were run on each dataset with 50 replicates, all of which estimated cline center and width. Each model was run for three independent MCMC chains (100,000 iterations per chain) and compared with each other and a null model of no clinal transition using the corrected Akaike Information Criterion. The best-fit model was determined through the lowest AICc score.

Supplementary Material

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

Acknowledgments

We thank Guo-Dong Wang and Wei-Wei Zhai for helpful comments and discussions, and Yun-He Wu for providing the map. This work was supported by the Strategic Priority Research Program, CAS (XDB31000000 and XDPB17), the National Natural Science Foundation of China (32100341), the Second Tibetan Plateau Scientific Expedition and Research Program (STEP, Grant No. 2019QZKK0501), and the Animal Branch of the Germplasm Bank of Wild Species, Chinese Academy of Sciences (the Large Research Infrastructure Funding).

Author Contributions

J.C., Y.-P.Z., and W.-W.Z. designed research; W.G. and W.-W.Z. did the fieldwork and samples collection; W.G., C.-X.Y., and J.-Q.J. performed the extraction of the genomic DNA and coordinated the genome sequencing; W.G., C.-X.Y., and B.-L.Z. performed genome assembly, annotation, and population genetic analysis; W.G., E.A.C., H.A.D., R.W.M., Y.-P.Z., and J.C. discussed and wrote the manuscript.

Data Availability

The raw sequencing data for our genome assembly, annotation, and population genomics have been deposited in the Genome Sequence Archive (GSA) of the National Genomics Data Center (NGDC) at https://ngdc.cncb.ac.cn/gsa/with accession numbers CRA004364 and CRA004746. Whole-genome assembly data were deposited in the Genome Warehouse (GWH) of NGDC at https://bigd.big.ac.cn/gwh under accession number GWHBCKV00000000. The annotation files can be found in Figshare (https://figshare.com/projects/Genomic_data_of_Phrynocephalus_vlangalii/116070).

References

Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE, Bierne N, Boughman J, Brelsford A, Buerkle CA, Buggs R, *et al.* 2013. Hybridization and speciation. *J Evol Biol.* **26**:229–246.

Ai H, Fang X, Yang B, Huang Z, Chen H, Mao L, Zhang F, Zhang L, Cui L, He W, *et al.* 2015. Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nat Genet.* **47**:217–225.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**: 1655–1664.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al.* 2000. Gene ontology: tool for the unification of biology. *Nat Genet.* **25**:25–29.

Attwood TK, Croning MDR, Flower DR, Lewis AP, Mabey JE, Scordis P, Selley JN, Wright W. 2000. PRINTS-S: the database formerly known as PRINTS. *Nucleic Acids Res.* **28**:225–227.

Avasthi P, Scheel JF, Ying G, Frederick JM, Baehr W, Wolfrum U. 2013. Germline deletion of *Cetn1* causes infertility in male mice. *J Cell Sci.* **126**:3204–3213.

Bailey R. 2020. gghybrid: R package for evolutionary analysis of hybrids and hybrid zones (v1.0.0). Zenodo.

Bairoch A, Apweiler R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**:45–48.

Barton NH. 1983. Multilocus clines. *Evolution* **37**:454–471.

Barton NH, Hewitt GM. 1985. Analysis of hybrid zones. *Annu Rev Ecol Evol Syst.* **16**:113–148.

Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, Sonnhammer ELL. 2000. The Pfam protein families database. *Nucleic Acids Res.* **28**: 263–266.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**:573–580.

Browning SR, Browning BL. 2007. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet.* **81**:1084–1097.

Brumfield RT, Jernigan RW, McDonald DB, Braun MJ. 2001. Evolutionary implications of divergent clines in an avian (*Manacus*: Aves) hybrid zone. *Evolution* **55**:2070–2087.

Bu D, Luo H, Huo P, Wang Z, Zhang S, He Z, Wu Y, Zhao L, Liu J, Guo J, *et al.* 2021. KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Res.* **49**:W317–W325.

Burri R, Nater A, Kawakami T, Mugal CF, Olason PI, Smeds L, Suh A, Dutoit L, Bureš S, Garamszegi LZ, *et al.* 2015. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Res.* **25**:1656–1665.

Butlin RK. 2005. Recombination and speciation. *Mol Ecol Resour.* **14**: 2621–2635.

Cantarel BL, Korf I, Robb SM, Parra G, Ross E, Moore B, Holt C, Sánchez Alvarado A, Yandell M. 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**:188–196.

Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, Li Y, Ye J, Yu C, Li Z, *et al.* 2018. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *GigaScience* **7**:1–6.

Chin C-S, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, *et al.* 2016. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods* **13**:1050–1054.

Corbo RM, Ulizzi L, Piombo L, Martinez-Labarga C, De Stefano GF, Scacchi R. 2007. Estrogen receptor alpha polymorphisms and fertility in populations with different reproductive patterns. *Mol Hum Reprod.* **13**:537–540.

Corpet F, Gouzy J, Kahn D. 1999. Recent improvements of the ProDom database of protein domain families. *Nucleic Acids Res.* **27**:263–267.

Coughlan JM, Matute DR. 2020. The importance of intrinsic postzygotic barriers throughout the speciation process. *Phil Trans R Soc B.* **375**:20190533.

Coyne JA, Orr HA. 2004. *Speciation*. Sunderland (MA): Sinauer Associates.

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.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, *et al.* 2011. The variant call format and VCFtools. *Bioinformatics* **27**: 2156–2158.

Darwin C. 1859. *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life.* London: John Murray.

Deng YY, Li JQ, Wu SF, Zhu YP, Chen YW, He FC. 2006. Integrated nr database in protein annotation system and its localization. *Comput Eng.* **32**:71–74.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, *et al.* 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* **43**:491–498.

Derryberry EP, Derryberry GE, Maley JM, Brumfield RT. 2014. HZAR: hybrid zone analysis using an R software package. *Mol Ecol Resour.* **14**:652–663.

Dobzhansky T. 1982. *Genetics and the origin of species*. New York: Columbia University Press.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, *et al.* 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**:92–95.

Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. 2016. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**:95–98.

Dynesius M, Jansson R. 2014. Persistence of within-species lineages: a neglected control of speciation rates. *Evolution* **68**:923–934.

Edelman NB, Frandsen P, Miyagi M, Clavijo BJ, Davey J, Dikow R, Accinelli GG, Van Belleghem S, Patterson NJ, Neafsey DE, *et al.* 2019. Genomic architecture and introgression shape a butterfly radiation. *Science* **366**:594–599.

Ellegren H, Smeds L, Burri R, Olason PI, Backström N, Kawakami T, Künstner A, Mäkinen H, Nadachowska-Brzyska K, Qvarnström A, *et al.* 2012. The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* **491**:756–760.

Etienne RS, Morlon H, Lambert A. 2014. Estimating the duration of speciation from phylogenies. *Evolution* **68**:2430–2440.

Ewing G, Hermisson J. 2010. MSMS: a coalescent simulation program including recombination, demographic structure and selection at a single locus. *Bioinformatics* **26**:2064–2065.

Feder JL, Flaxman SM, Egan SP, Comeault AA, Nosil P. 2013. Geographic mode of speciation and genomic divergence. *Annu Rev Ecol Evol Syst.* **44**:73–97.

Felsenstein J. 1981. Skepticism towards Santa Rosalia, or why are there so few kinds of animals? *Evolution* **35**:124–138.

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 Genom Genet.* **6**:1563–1571.

Gao W, Sun Y-B, Zhou W-W, Xiong Z-J, Chen L, Li H, Fu T-T, Xu K, Xu W, Ma L, *et al.* 2019. Genomic and transcriptomic investigations of the evolutionary transition from oviparity to viviparity. *Proc Natl Acad Sci U S A.* **116**:3646–3655.

Ge Y-Z, Xu L-W, Jia R-P, Xu Z, Li W-C, Wu R, Liao S, Gao F, Tan S-J, Song Q, *et al.* 2014. Association of polymorphisms in estrogen receptors (*ESR1* and *ESR2*) with male infertility: a meta-analysis and systematic review. *J Assist Reprod Genet.* **31**:601–611.

Giwercman A. 2011. Estrogens and phytoestrogens in male infertility. *Curr Opin Urol.* **21**:519–526.

Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, Bateman A. 2005. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33**:D121–D124.

Gronau I, Hubisz MJ, Gulko B, Danko CG, Siepel A. 2011. Bayesian inference of ancient human demography from individual genome sequences. *Nat Genet.* **43**:1031–1034.

Gunawan A, Kaewmala K, Uddin MJ, Cinar MU, Tesfaye D, Phatsara C, Tholen E, Looft C, Schellander K. 2011. Association study and expression analysis of porcine *ESR1* as a candidate gene for boar fertility and sperm quality. *Anim Reprod Sci.* **128**:11–21.

Guo XG, Wang YZ. 2007. Partitioned Bayesian analyses, dispersal-vicariance analysis, and the biogeography of Chinese toad-headed lizards (Agamidae: *Phrynocephalus*): a re-evaluation. *Mol Phylogenet Evol.* **45**:643–662.

Han F, Lamichhaney S, Grant BR, Grant PR, Andersson L, Webster MT. 2017. Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. *Genome Res.* **27**:1004–1015.

Harrison RG, Larson EL. 2014. Hybridization, introgression, and the nature of species boundaries. *J Hered.* **105**:795–809.

Hirase S, Yamasaki YY, Sekino M, Nishisako M, Ikeda M, Hara M, Merilä J, Kikuchi K. 2021. Genomic evidence for speciation with gene flow in broadcast spawning marine invertebrates. *Mol Biol Evol.* **38**, 4683–4699.

Howard DJ. 1993. Reinforcement: origin, dynamics, and fate of an evolutionary hypothesis. In *Hybrid zones and the evolutionary process*. Oxford: Oxford University Press. p. 46–69.

Hu Y, Yu H, Shaw G, Renfree MB, Pask AJ. 2011. Differential roles of TGIF family genes in mammalian reproduction. *BMC Dev Biol.* **11**:58.

Hulo N, Bairoch A, Bulliard V, Cerutti L, Cuche BA, de Castro E, Lachaize C, Langendijk-Genevaux PS, Sigrist CJA. 2007. The 20 years of PROSITE. *Nucleic Acids Res.* **36**:D245–D249.

Jin YT, Brown RP. 2013. Species history and divergence times of viviparous and oviparous Chinese toad-headed sand lizards (*Phrynocephalus*) on the Qinghai-Tibetan Plateau. *Mol Phylogenet Evol.* **68**:259–268.

Jin Y-T, Brown RP, Liu N-F. 2008. Cladogenesis and phylogeography of the lizard *Phrynocephalus vlangalii* (Agamidae) on the Tibetan plateau. *Mol Ecol.* **17**:1971–1982.

Jin Y-T, Liu N-F. 2008. Introgression of mtDNA between two *Phrynocephalus* lizards in Qinghai Plateau (in Chinese). *Acta Zool Sin.* **54**:111–121.

Jin Y, Yang Z, Brown RP, Liao P, Liu N. 2014. Intraspecific lineages of the lizard *Phrynocephalus putjatia* from the Qinghai-Tibetan Plateau: impact of physical events on divergence and discordance between morphology and molecular markers. *Mol Phylogenet Evol.* **71**:288–297.

Kautt AF, Kratochwil CF, Nater A, Machado-Schiaffino G, Olave M, Henning F, Torres-Dowdall J, Härer A, Hulsey CD, Franchini P, *et al.* 2020. Contrasting signatures of genomic divergence during sympatric speciation. *Nature* **588**:106–111.

Knief U, Bossu CM, Saino N, Hansson B, Poelstra J, Vijay N, Weissensteiner M, Wolf J. 2019. Epistatic mutations under divergent selection govern phenotypic variation in the crow hybrid zone. *Nat Ecol Evol.* **3**:570–576.

Korf I. 2004. Gene finding in novel genomes. *BMC Bioinform.* **5**:59.

Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, Martinez-Barrio A, Promerová M, Rubin C-J, Wang C, Zamani N, *et al.* 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature* **518**:371–375.

Lawniczak MKN, Emrich SJ, Holloway AK, Regier AP, Olson M, White B, Redmond S, Fulton L, Appelbaum E, Godfrey J, *et al.* 2010. Widespread divergence between incipient *Anopheles gambiae* species revealed by whole genome sequences. *Science* **330**: 512–514.

LeClere AR, Yang JK, Kirkpatrick DT. 2013. The role of *CSM3*, *MRC1*, and *TOF1* in minisatellite stability and large loop DNA repair during meiosis in yeast. *Fungal Genet Biol.* **50**:33–43.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**:1754–1760.

Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* **475**:493–496.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**:2078–2079.

Liu Y, Jiang M, Li C, Yang P, Sun H, Tao D, Zhang S, Ma Y. 2011. Human t-complex protein 11 (TCP11), a testis-specific gene product, is a potential determinant of the sperm morphology. *Tohoku J Exp Med.* **224**:111–117.

Liu L, Yu L, Edwards SV. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evol Biol.* **10**:302.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**:955–964.

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, *et al.* 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience* **1**:18.

Ma Y, Zhang S, Xia Q, Zhang G, Huang X, Huang M, Xiao C, Pan A, Sun Y, Lebo R, *et al.* 2002. Molecular characterization of the *TCP11* gene which is the human homologue of the mouse gene encoding the receptor of fertilization promoting peptide. *Mol Hum Reprod.* **8**:24–31.

Mallet J, Meyer A, Nosil P, Feder JL. 2009. Space, sympatry and speciation. *J Evol Biol.* **22**:2332–2341.

Martin SH, Dasmahapatra KK, Nadeau NJ, Salazar C, Walters JR, Simpson F, Blaxter M, Manica A, Mallet J, Jiggins CD. 2013. Genome-wide evidence for speciation with gene flow in Heliconius butterflies. *Genome Res.* **23**:1817–1828.

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.

Matute DR. 2010. Reinforcement of gametic isolation in *Drosophila*. *PLoS Biol.* **8**:e1000341.

Mayr E. 1963. *Animal species and evolution*. Cambridge: Harvard University Press.

McGinnis S, Madden TL. 2004. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* **32**: W20–W25.

McKenzie JL, Dhillon RS, Schulte PM. 2015. Evidence for a bimodal distribution of hybrid indices in a hybrid zone with high admixture. *R Soc Open Sci.* **2**:150285.

Metcalf CE, Wassarman DA. 2007. Nucleolar colocalization of TAF1 and testis-specific TAFs during *Drosophila* spermatogenesis. *Dev Dyn.* **236**:2836–2843.

Mi H, Lazareva-Ulitsky B, Loo R, Kejariwal A, Vandergriff J, Rabkin S, Guo N, Muruganujan A, Doremieux O, Campbell MJ, *et al.* 2005. The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res.* **33**:D284–D288.

Muller H. 1942. Isolating mechanisms, evolution, and temperature. *Biol Symp.* **6**:71–125.

Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**:2933–2935.

Netzel-Arnett S, Bugge TH, Hess RA, Carnes K, Stringer BW, Scarman AL, Hooper JD, Tonks ID, Kay GF, Antalis TM. 2009. The glycosylphosphatidylinositol-anchored serine protease PRSS21 (testisin) imparts murine epididymal sperm cell maturation and fertilizing ability. *Biol Reprod.* **81**:921–932.

Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* **32**:268–274.

Noble DWA, Qi Y, Fu J. 2010. Species delineation using Bayesian model-based assignment tests: a case study using Chinese toad-headed agamas (genus *Phrynocephalus*). *BMC Evol Biol.* **10**:197.

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.

Nosil P, Egan SP, Funk DJ. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes: "isolation by adaptation" and multiple roles for divergent selection. *Evolution* **62**:316–336.

Nosil P, Feder JL. 2012. Genomic divergence during speciation: causes and consequences. *Phil Trans R Soc B.* **367**:332–342.

Nosil P, Funk DJ, Ortiz-Barrientos D. 2009. Divergent selection and heterogeneous genomic divergence. *Mol Ecol.* **18**:375–402.

Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. 1999. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **27**:29–34.

Olivier-Van Stichelen S, Drougat L, Dehennaut V, El Yazidi-Belkoura I, Guinez C, Mir AM, Michalski JC, Vercoutter-Edouart AS, Lefebvre T. 2012. Serum-stimulated cell cycle entry promotes ncOGT synthesis required for cyclin D expression. *Oncogenesis* **1**:e36.

Ortiz-Barrientos D, Engelstädter J, Rieseberg LH. 2016. Recombination rate evolution and the origin of species. *Trends Ecol Evol.* **31**:226–236.

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.

Poelstra JW, Vijay N, Bossu CM, Lantz H, Ryll B, Müller I, Baglione V, Unneberg P, Wikelski M, Grabherr MG, *et al.* 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* **344**:1410.

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.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**:841–842.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol.* **67**:901–904.

Ravinet M, Faria R, Butlin RK, Galindo J, Bierne N, Rafajlović M, Noor MAF, Mehlig B, Westram AM. 2017. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *J Evol Biol.* **30**:1450–1477.

Renaut S, Grassa CJ, Yeaman S, Moyers BT, Lai Z, Kane NC, Bowers JE, Burke JM, Rieseberg LH. 2013. Genomic islands of divergence are not affected by geography of speciation in sunflowers. *Nat Commun.* **4**:1827.

Rockett JC, Patrizio P, Schmid JE, Hecht NB, Dix DJ. 2004. Gene expression patterns associated with infertility in humans and rodent models. *Mutat Res.* **549**:225–240.

Rosenblum EB, Sarver BAJ, Brown JW, Des Roches S, Hardwick KM, Hether TD, Eastman JM, Pennell MW, Harmon LJ. 2012. Goldilocks meets Santa Rosalia: an ephemeral speciation model explains patterns of diversification across time scales. *Evol Biol.* **39**:255–261.

Rundle HD, Nagel L, Boughman JW, Schluter D. 2000. Natural selection and parallel speciation in sympatric sticklebacks. *Science* **287**:306–308.

Rundle HD, Nosil P. 2005. Ecological speciation. *Ecol Lett.* **8**:336–352.

Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, Xie X, Byrne EH, McCarroll SA, Gaudet R. 2007. Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**:913–918.

Schluter D. 2000. *The ecology of adaptive radiation*. Oxford: Oxford University Press.

Schramm S, Fraune J, Naumann R, Hernandez-Hernandez A, Höög C, Cooke HJ, Alsheimer M, Benavente R. 2011. A novel mouse synaptonemal complex protein is essential for loading of central element proteins, recombination, and fertility. *PLoS Genet.* **7**: e1002088.

Schultz J, Copley RR, Doerks T, Ponting CP, Bork P. 2000. SMART: a web-based tool for the study of genetically mobile domains. *Nucleic Acids Res.* **28**:231–234.

Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, Peichel CL, Saetre G-P, Bank C, Brännström Å, *et al.* 2014. Genomics and the origin of species. *Nat Rev Genet.* **15**:176–192.

Servant N, Varoquaux N, Lajoie BR, Viara E, Chen C-J, Vert J-P, Heard E, Dekker J, Barillot E. 2015. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**:259.

Sigrist CJA, de Castro E, Cerutti L, Cuche BA, Hulo N, Bridge A, Bougueleret L, Xenarios I. 2012. New and continuing developments at PROSITE. *Nucleic Acids Res.* **41**:D344–D347.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**: 3210–3212.

Slater GSC, Birney E. 2005. Automated generation of heuristics for biological sequence comparison. *BMC Bioinform.* **6**:31.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.

Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**:W435–W439.

Strasburg JL, Rieseberg LH. 2013. Methodological challenges to realizing the potential of hybridization research. *J Evol Biol.* **26**: 259–260.

Szpiech ZA, Hernandez RD. 2014. Selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Mol Biol Evol.* **31**:2824–2827.

Takahashi H, Toyoda A, Yamazaki T, Narita S, Mashiko T, Yamazaki Y. 2017. Asymmetric hybridization and introgression between sibling species of the pufferfish *Takifugu* that have undergone explosive speciation. *Mar Biol.* **164**:90.

Tang H, Peng J, Wang P, Risch NJ. 2005. Estimation of individual admixture: analytical and study design considerations. *Genet Epidemiol.* **28**:289–301.

Tang L, Zeng W, Clark RK, Dobrinski I. 2011. Characterization of the porcine testis-expressed gene 11 (Tex11). *Spermatogenesis* **1**: 147–151.

Tarailo-Graovac M, Chen N. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinform.* **25**:4.10.1–4.10.14.

Turbek SP, Browne M, Di Giacomo AS, Kopuchian C, Hochachka WM, Estalles C, Lijtmaer DA, Tubaro PL, Silveira LF, Lovette IJ, *et al.* 2021. Rapid speciation via the evolution of pre-mating isolation in the Iberá Seedeater. *Science* **371**:eabc0256.

Vijay N, Bossu CM, Poelstra JW, Weissensteiner MH, Suh A, Kryukov AP, Wolf JBW. 2016. Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nat Commun.* **7**:13195.

Wagner CE, Mandeville EG. 2017. Speciation, species persistence and the goals of studying genomic barriers to gene flow. *J Evol Biol.* **30**: 1512–1515.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, *et al.* 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**:e112963.

Wang PJ, McCarrey JR, Yang F, Page DC. 2001. An abundance of X-linked genes expressed in spermatogonia. *Nat Genet.* **27**: 422–426.

Wang J, Street NR, Scofield DG, Ingvarsson PK. 2016. Variation in linked selection and recombination drive genomic divergence during allopatric speciation of European and American aspens. *Mol Biol Evol.* **33**:1754–1767.

Wang YZ, Zeng XM, Fang ZL, Wu GF, Liu ZJ, Papenfuss TJ, Macey JR. 2002. A valid species of the genus *Phrynocephalus*: *P. putjatia* and a discussion on taxonomy of *Phrynocephalus hongyuanensis* (Sauria: Agamidae) (in Chinese). *Acta Zootaxon Sin.* **27**: 372–383.

Wang Y, Zhan A, Fu J. 2009. Testing historical phylogeographic inferences with contemporary gene flow data: population genetic structure of the Qinghai toad-headed lizard. *J Zool.* **278**:149–156.

Wang G-D, Zhang B-L, Zhou W-W, Li Y-X, Jin J-Q, Shao Y, Yang H-C, Liu Y-H, Yan F, Chen H-M, *et al.* 2018. Selection and environmental adaptation along a path to speciation in the Tibetan frog *Nanorana parkeri*. *Proc Natl Acad Sci U S A.* **115**: E5056–E5065.

Wolf JB, Ellegren H. 2017. Making sense of genomic islands of differentiation in light of speciation. *Nat Rev Genet.* **18**:87–100.

Wu C-I. 2001. The genic view of the process of speciation. *J Evol Biol.* **14**:851–865.

Wu TD, Watanabe CK. 2005. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**: 1859–1875.

Yamashita M, Honda A, Ogura A, Kashiwabara S-I, Fukami K, Baba T. 2008. Reduced fertility of mouse epididymal sperm lacking Prss21/Tesp5 is rescued by sperm exposure to uterine microenvironment. *Genes Cells* **13**:1001–1013.

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.

Ye T, Kang M, Huang Q, Fang C, Chen Y, Shen H, Dong S. 2014. Exposure to DEHP and MEHP from hatching to adulthood causes reproductive dysfunction and endocrine disruption in marine medaka (*Oryzias melastigma*). *Aquat Toxicol.* **146**:115–126.

Yu Y-H, Siao F-P, Hsu LC-L, Yen PH. 2012. TEX11 modulates germ cell proliferation by competing with estrogen receptor β for the binding to HPIP. *Mol Endocrinol.* **26**:630–642.

Yuan C-W, Sun X-L, Qiao L-C, Xu H-X, Zhu P, Chen H-J, Yang B-L. 2019. Non-SMC condensin I complex subunit D2 and non-SMC condensin II complex subunit D3 induces inflammation *via* the IKK/NF-κB pathway in ulcerative colitis. *World J Gastroenterol.* **25**:6813–6822.

Zdobnov EM, Apweiler R. 2001. InterProScan–an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17**:847–848.

Zeng X-M, Wang Y-Z, Liu Z-J, Fang Z-L, Wu G-F, Papenfuss TJ, Macey RJ. 1997. Karyotypes of nine species in the genus *Phrynocephalus*, with discussion of karyotypic evolution of Chinese *Phrynocephalus* (in Chinese). *Acta Zool Sin.* **43**:399–410.

Zhang C, Dong S-S, Xu J-Y, He W-M, Yang T-L. 2019. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **35**:1786–1788.

Zhou C, Kang W, Baba T. 2012. Functional characterization of double-knockout mouse sperm lacking SPAM1 and ACR or SPAM1 and PRSS21 in fertilization. *J Reprod Dev.* **58**:330–337.

Zhou Y, Yuan K, Yu Y, Ni X, Xie P, Xing EP, Xu S. 2017. Inference of multiple-wave population admixture by modeling decay of linkage disequilibrium with polynomial functions. *Heredity* **118**: 503–510.

Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. 2019. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun.* **10**:1523.