Title
Selection and environmental adaptation along a path to speciation in the Tibetan frog Nanorana parkeri.

Permalink
https://escholarship.org/uc/item/9w086955

Journal
Proceedings of the National Academy of Sciences of USA, 115(22)

Authors
Wang, Guo-Dong
Zhang, Bao-Lin
Zhou, Wei-Wei
et al.

Publication Date
2018-05-29

DOI
10.1073/pnas.1716257115

Peer reviewed
Selection and environmental adaptation along a path to speciation in the Tibetan frog *Nanorana parkeri*

Guo-Dong Wang, Bao-Lin Zhang, Wei-Wei Zhou, Yong-Xin Li, Jie-Qiong Jin, Yong Shao, He-chuan Yang, Yan-Hu Liu, Fang Yan, Hong-Man Chen, Li Jin, Feng Gao, Yaoguang Zhang, Haipeng Li, Bingyu Mao, Robert W. Murphy, David B. Wake, Ya-Ping Zhang, and Jing Che

*State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China; Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223, China; Kunming College of Life Science, University of Chinese Academy of Sciences, Kunming 650204, Yunnan, China; Southeast Asia Biodiversity Research Institute, Chinese Academy of Sciences, Yezin, 05282 Nay Pyi Taw, Myanmar; Human Genetics, Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore 138672, Singapore; Laboratory for Conservation and Utilization of Bio-Resources & Key Laboratory for Microbial Resources of the Ministry of Education, Yunnan University, Kunming 650091, China; Key Laboratory of Freshwater Fish Reproduction and Development of the Ministry of Education and Key Laboratory of Aquatic Science of Chongqing, Southwest University School of Life Sciences, Chongqing 400715, China; CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China; Centre for Biodiversity and Conservation Biology, Royal Ontario Museum, Toronto, ON, Canada M5S 2C6; and Department of Integrative Biology and Museum of Vertebrate Zoology, University of California, Berkeley, CA 94720-3160*

Contributed by David B. Wake, April 13, 2018 (sent for review September 17, 2017; reviewed by Scott V. Edwards and Craig Moritz)

Tibetan frogs, *Nanorana parkeri*, are differentiated genetically but not morphologically along geographical and elevational gradients in a challenging environment, presenting a unique opportunity to investigate processes leading to speciation. Analyses of whole genomes of 63 frogs reveal population structuring and historical demography, characterized by highly restricted gene flow in a narrow geographic zone lying between maternal lineages West (W) and East (E). A population found only along a single tributary of the Yalu Zangbu River has the mitogenome only of E, whereas nuclear genes of W comprise 89–95% of the nuclear genome. Selection accounts for 579 broadly scattered, highly divergent regions (HDRs) of the genome, which involve 365 genes. These genes fall into 51 gene ontology (GO) functional classes, 14 of which are likely to be important in driving reproductive isolation. GO enrichment analyses of E reveal many overrepresented functional categories associated with adaptation to high elevations, including blood circulation, response to hypoxia, and UV radiation. Four genes, including **DNAJC8** in the brain, **TMN1C** and **ADORA1** in the heart, and **LAMB3** in the lung, differ in levels of expression between low- and high-elevation populations. High-altitude adaptation plays an important role in maintaining and driving continuing divergence and reproductive isolation. Use of total genomes enabled recognition of selection and adaptation in and between populations, as well as documentation of evolution along a stepped cline toward speciation.

Significance

Central topics in evolutionary biology include uncovering the processes and genetic bases of speciation and documenting environmental adaptations and processes responsible for them. The challenging environment of the Qinghai-Tibetan Plateau (QTP) facilitates such investigations, and the Tibetan frog, *Nanorana parkeri*, offers a unique opportunity to investigate these processes. A cohort of whole-genome sequences of 63 individuals from across its entire range opens avenues for incorporating population genomics into studies of speciation. Natural selection plays an important role in maintaining and driving the continuing divergence and reproductive isolation of populations of the species. The QTP is a natural laboratory for studying how selection drives adaptation, how environments influence evolutionary history, and how these factors can interact to provide insight into speciation.

Author contributions: W.-W.Z., D.B.W., Y.-P.Z., and J.C. designed research; G.-D.W., Y.-P.Z., and J.C. managed this project; B.-L.Z., J.-Q.J., Y.S., and J.C. collected the samples and performed physiological measurement; B.-L.Z., J.-Q.J., Y.S., and J.C. collected the samples and performed physiological measurement; B.-L.Z., J.-Q.J., Y.S., and J.C. collected the samples and performed physiological measurement; Y.-X.L., Y.S., and B.M. performed the gene expression experiments; B.-L.Z., H.-Y.C., and Y.-H.L. performed the histological experiment; F.G. and H.L. calculated the population-scaled recombination rates; G.-D.W. and B.-L.Z. analyzed the data; and G.-D.W., B.-L.Z., W.-Z.W., R.W.M., D.B.W., Y.-P.Z., and J.C. wrote the manuscript.

Reviewers: S.V.E., Harvard University; and C.M., Australian National University.

The authors declare no conflict of interest.

This open access article is distributed under Creative Commons Attribution Non-Commercial-NoDerivatives License 4.0 (CC BY-NC-ND).

The sequence data reported in this paper have been deposited in the genome sequence archive of Beijing Institute of Genomics, Chinese Academy of Sciences, gsa.big.ac.cn (accession nos. CRA000919 for *N. parkeri*, and CRA000918 for *N. plebeia*).

1. G.-D.W. and B.-L.Z. contributed equally to this work.

2. To whom correspondence may be addressed. Email: chej@mail.kiz.ac.cn, zhangyp@email.scri.ac.cn, or wakelab@berkeley.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1716257115/-/DCSupplemental.

Published online May 14, 2018.
How organisms adapt and diversify in high-altitude environments has attracted much attention in recent years (29–31). Because of its heterogeneous topography, inhospitable environment, and complex paleoclimate history, the Qinghai-Tibetan Plateau (QTP) is a natural laboratory for studying adaptation and speciation (32–35). Differences in elevation between valleys and mountaintops often exceed 2,000 m, presenting strong ecological gradients in abiotic variables, such as oxygen partial pressure (36), UV radiation (37), precipitation (38), and ambient temperatures (39). The periodicity of climatic oscillations affects diverse phenomena, such as dispersal, fluctuations in population size, and speciation (32, 40).

An endemic Tibetan frog, *Nanorana parkeri*, occurs in lentic environments of the southeastern QTP along the Yalu Zangbu River (YZR) drainage. It faces challenges that few other amphibians experience. Not unusual for frogs, it occurs at an elevation of 2,800 m, yet it is the only amphibian to also exist at 5,000 m (41), where oxygen is scarce and UV radiation is dangerously high. Western (W) and eastern (E) mitochondrial DNA matrilines of *N. parkeri* diverged in the middle Pleistocene and formed distinct entities with overlapping elevational ranges (42). Three nuclear DNA loci correspond to this pattern with limited admixture of E and W alleles in two localities near their geographic boundaries (43). However, neither the extent of gene flow nor the degree of isolation in the zone of admixture is known. No morphological differences between W and E have been noted (44). Clines in elevation, genetic-morphological discordance, and mtDNA introgression in the zone of admixture offer an unprecedented opportunity to investigate the role selection plays in driving rapid genetic change, environmental adaptation, and the evolution of species differences.

The annotated genome of *N. parkeri* (45) facilitates investigations into the evolutionary drivers of population divergence, and possibly speciation, by enabling the identification of selected genes and their functions. Herein, we report on whole-genome sequences of 63 *N. parkeri* individuals from across the range of the species and decipher the genetic mechanisms that may be responsible for driving population divergence. We reconstruct the evolutionary history of *N. parkeri* from a genomic perspective, investigate the impacts of isolation and gene flow on the process of speciation, and explore genomic consequences. We also evaluate ecological factors leading to speciation and identify candidate genes underlying the observed differentiation.

**Results**

**Sampling and Sequencing.** Collection localities ranged from 2,900 to 4,900 m in elevation and covered the entire documented distribution of *N. parkeri* (Fig. 1A and SI Appendix, Table S1).
The 45 E and W individuals were sequenced for 33.77 Gb (16.89-fold coverage) on average, and 18 individuals with mixed E and W heritage were sequenced for 11.92 Gb (5.96-fold coverage). Most reads were aligned (90.88% average mappable rate) to the reference genome of *N. parkeri* belonging to E (45). After genotyping and stringent quality-filtering, we retained 8.59 million single-nucleotide polymorphisms (SNPs) of E and W individuals and 6.44 million SNPs of mixed heritage.

**Population Structure, Phylogeny, andIntrogression.** Principal component analysis (PCA) and population structure analyses unambiguously identify five genetic clusters. The PCA plot separates populations W and E along the first eigenvector, which explains 57.76% of total genetic variance (Fig. 1B). The second eigenvector identifies subpopulations E1–E4 and explains 8.80% of the variance. E1 and E3 comprise samples from low elevations (∼2,900–3,300 m; localities 1–5; Fig. 1 A–C) near the Yalu Zangbu Grand Canyon, and no obvious geographic barrier separates them. Subpopulation E2 includes samples from high elevations (∼3,900–4,900 m; localities 6–8; Fig. 1 A–C). The remaining localities, at elevations from ∼3,700–4,500 m, form subpopulation E4. Phylogenetic analyses based on 1,000 neutral loci (Fig. 1D) and gene tree-based coalescent (*SI Appendix, Figs. S1 and S2*) methods display nearly identical topologies. Populations W and E split into distinct branches, with subpopulation E1 in the most basal position, subpopulation E2 following subpopulation E1, and then allopatric units E3 and E4. Population structure analysis corroborates the phylogenetic analyses and PCA, and identifies signs of nuclear gene flow from E to W at their boundary (Fig. 2). Two topology-based methods verify nuclear admixture among some of the matriline. The TreeMix analysis (*SI Appendix, Fig. S3*) and the D-statistic test (Table 1) show gene flow within sympatric E1 and E3. A significant signature of admixture is found between E2 and E4 ([|Z| > 3; Table 1]).

**Demography History.** We used the generalized phylogenetic coalescent sampler (G-PhoCS) (46) to infer ancestral population sizes, divergence times, and migration rates. We estimate the divergence time of W and E from ∼688.3–846.0 kya (K*a*) (Fig. 3A). The ancestral effective population size of E at this period was reduced by ∼56% (*SI Appendix, Table S2*). Pairwise sequentially Markovian coalescence (PSMC) results suggest that W was reduced by ∼50% (*SI Appendix, Fig. S4*). Population expansion occurred two- to threefold until the penultimate glaciation [Marine Isotope Stage 6, ∼100–200 K*a* (47)]. E1 diverged from E2–E4 at ∼181.7 K*a* [95% confidence interval (CI) = 158.0–205.4 K*a*]. The time of divergence for E2 relative to E3 and E4, and between the latter two, is estimated at 45.9 K*a* (95% CI: 36.9–55.6 K*a* ) and 39.3 K*a* (95% CI: 32.1–46.6 K*a*), respectively. We may underestimate the divergence time because the mutation rate [0.776e-09 per site per year (45)] is a conservative value. Demographic analysis shows evidence of genetic exchanges among all subpopulations of E as well as directional migration from E4 to W (Fig. 3A). Gene flow from the hypothesized ancestral population of E2–E4 into W occurred between 45.9 and 181.7 K*a* (G-PhoCS analyses). Coalescent simulations of gene flow between W and E under different migration scenarios suggest that the inferred demographic parameters were most concordant with the observed patterns of differentiation (Fig. 3B).

**Genomic Isolation Between West and East Matriline.** We used several statistical approaches to assess the extent of genomic differentiation between W and E. High levels of differentiation between W and E and within E are found using pairwise mean relative divergence (FST) and absolute sequence divergence (dxy). The mean FST = 0.4787 ± 0.0097 and dxy = 0.0030 ± 0.0002 values are more than threefold higher than those among subpopulations of E (FST: 0.1405 ± 0.0322; dxy: 0.0009 ± 0.0001; Fig. 4 and *SI Appendix, Tables S3 and S4*). Levels of divergence using the density of fixed differences per site between populations (df) are about 100-fold higher than those within subpopulations of E (*SI Appendix, Table S5*). The genomic landscape of divergence using the df statistic finds large numbers of loci across the genome fixed between W and subpopulations of E (Fig. 4A and *SI Appendix, Fig. S5*).

We used a modified population branch statistic (PBS) approach to measure the extent of genomic differentiation between W and all four subpopulations of E (*Materials and Methods*). The mean PBS value of W (PBSW) is 0.5583. Neutral coalescent simulations based on the inferred demographic model find the top 2.5% of the observed PBSW values (ca. PBSW ≥ 0.8906; Fig. 5A) to be significantly higher than simulated neutral values (PBS ≤ 2.2e-16, two-tailed Mann–Whitney test; Fig. 6). Because all five populations of the species expanded recently (based on PSMC analysis), we performed the simulation taking this expansion into account; results were similar to those under the G-PhoCS model (*SI Appendix, section 1.1*). Accordingly, we can identify highly divergent regions (HDRs) of the genome with the observed PBSW ≥ 0.8906; there are 579 regions with a total length of 39.60 Mb. The longest fragment is 350 kb, but most (76.34%) are only a one-window-length (50 kb) fragment (*SI Appendix, Fig. S6*). FST and dxy values are significantly higher in HDRs than those in the background genome (P < 2.2e-16, two-tailed Mann–Whitney test; Fig. 5B and *SI Appendix, Table S6*). The correlation between FST and dxy is significantly positive (Pearson’s R > 0.3, P < 2.2e-16; *SI Appendix, Fig. S7*), reflecting reduced gene flow in these regions. A contrasting pattern between W and E occurs with respect to nucleotide diversity (π) (*SI Appendix, Table S6*). The value in HDRs of W is no less than for the rest of the genome (0.00078 vs. 0.00067) but is reduced significantly (P < 2.2e-16) in the corresponding regions of subpopulations of E (Fig. 5B and *SI Appendix, Table S6*). The reduced level of intrapopulation diversity in HDRs suggests that selection has affected E. Relative

### Table 1. Admixture signatures from D-statistic tests

| Test | D statistic | Z score |
|------|-------------|---------|
| W, E1, E3, E4 | -0.247 | -50.868 |
| W, E1, E2, E3 | -0.264 | -61.376 |
| W, E1, E4 | -0.209 | -2.354 |
| W, E2, E3, E4 | 0.117 | 21.657 |
| E1, E2, E3, E4 | 0.260 | 78.181 |

Populations with gene flow are denoted in boldface.
to the genomic background, these regions also contain a significantly high derived allele frequency (DAF) ($P < 2.2e-16$).

**Divergence of W and E Matrilines in Relation to Reproductive Isolation.** We used the annotated genome of *N. parkeri* to infer the functions of candidate target genes within the HDRs that might relate to speciation. We annotated 365 protein-coding genes in the HDRs of PBSW. Gene ontology (GO) evaluations identified 51 functional classes that were significantly overrepresented ($P < 0.05$; SI Appendix, Table S7). In total, 21 genes distributed among 14 GO terms have functions related to reproduction, including, as examples, reproductive developmental process (GO:003006), sexual reproduction (GO:0019953), and spermatogenesis (GO:0007283) (Table 2 and SI Appendix, Table S7).

**Little Representation of E Matrilines in the Nuclear Genome Within the Zone of Admixture.** Individuals in the zone of admixture are more closely related to W than E with respect to the nuclear genome (*SI Appendix, Figs. S8 and S9*). However, all mixed individuals have the mitogenome of E (*SI Appendix, Fig. S10*). We used PCAdmix to identify genomic regions from E in the genomes of the mixed individuals. The low level of E ancestry detected ($<5.9–11.2\%$; *SI Appendix, Fig. S11*) is not caused by systematic errors ($<2\%$; *SI Appendix, Table S8,*). Regions of E ancestry occur in no less than 67% (24 of 36 haploids) of mixed individuals. Such regions, about 1.6 Mb in size, are dispersed randomly across the genome, with a mean length of 62 Kb. These regions contain 31 protein-coding genes, one relevant to mitochondria [*VDAC3 (48)*], but none exhibit reduced levels of polymorphism (0.00059 > 0.00046) or significant variation in intergroup and intragroup nonsynonymous/synonymous ratios ($P > 0.05$, McDonald–Kreitman test (49)). We detect no signal of selection on the mitogenome in the zone of admixture (Tajima D = $-1.49$, $P > 0.10$).

**Incomplete Isolation and Gene Flow Within the Geographic Range of E Matrilines.** Levels of whole-genomic differentiation among E1–E4 are comparatively small. Interpopulation $F_{ST}$ values range from 0.0974 to 0.1825, and dxy values range from 0.0007 to 0.0010 (*SI Appendix, Tables S3 and S4,*). Mean $d_f$ values within the subpopulations range from $1.27 \times 10^{-16}$ to $6.02 \times 10^{-07}$ (*SI Appendix, Figs. S12 and S13 and Table S5*), which corresponds to ~1.27–6.02 fixed differences for every 10 Mb of sequence data in compared populations. We applied the PBS statistic to identify HDRs and then examined the features of these regions. Similar patterns occur for each subpopulation as follows: (i) most outlier windows (>72%) are discontinuous and one window in length (50 kb) (*SI Appendix, Fig. S6*); (ii) $F_{ST}$ and dxy values are significantly higher than the background genome ($P < 2.2 \times 10^{-16}$, two-tailed Mann–Whitney test; Fig. 5D and *SI Appendix, Fig. S14 and Table S6*); (iii) $\pi$ is significantly lower in HDRs relative to the rest of the genome, although the level varies among the subpopulations (Fig. 5D and *SI Appendix, Fig. S14 and Table S6*); and (iv) E1–E3 are skewed toward high frequencies of derived variants (Fig. 5D and *SI Appendix, Fig. S14 and Table S6*).

**Environmental, Histological, and Physiological Divergence Affecting E Matrilines.** In a PCA of all bioclimatic variables associated with frog sampling sites (*SI Appendix, Fig. S15*), PC1 explains 49.13% of the variation. This value differs significantly between low-elevation clades E1 and E3 and high-elevation clades E2 and E4 ($P = 3.3e-4$). Histological sections of middorsal skin show that the frogs from high elevations have a significantly greater number of granular glands than frogs from low elevations ($P < 0.05$, two-tailed t test; Fig. 7A); such glands may function in response to environmental stimuli (50). Hemoglobin (Hb) levels in peripheral blood are correlated with elevation ($P < 0.01$; Fig. 7B). In contrast to Hb levels, muscular oxygen content in relatively low-elevation populations is significantly higher than muscular oxygen content in populations from higher elevations ($P < 0.01$; Fig. 7C); this corresponds to the decreased oxygen content in air.

**Genes, Ecological Factors, and the Evolution of E Matrilines.** Genes from the HDRs of each subpopulation were evaluated for functional categories that related to environmental factors. GO enrichment analyses revealed many overrepresented functional categories that appear to associate with adaptation to the environment of the QTP (*SI Appendix, Tables S9–S12*). For instance, Kyoto Encyclopedia of Genes and Genomes pathways and GO categories related to metabolism (fatty acid metabolism, hsa00771), blood circulation (circulatory system process, GO:0003013), motility (muscle cell differentiation, GO:0042692), and immune response (inflammatory response, GO:0006954) characterize low-elevation E1 (*SI Appendix, Table S9*). GO categories associated with apoptosis
and isolating barriers (RIBs) N. parkeri distribution between E3 and E4 and their landscape of genomic divergence measured by the df. Both statistics are measured from 50-kb nonoverlapping windows. Scaffolds are concatenated to reveal the whole-genomic divergence pattern. (B) FST distribution between E1 and E2 and their landscape of genomic divergence. (C) F ST distribution between E3 and E4 and their landscape of genomic divergence.

Fig. 4. Genomic divergence associated with species formation. (A) FST distribution between W and E1 and their landscape of genomic divergence measured by the df. Both statistics are measured from 50-kb nonoverlapping windows. Scaffolds are concatenated to reveal the whole-genomic divergence pattern. (B) FST distribution between E1 and E2 and their landscape of genomic divergence. (C) FST distribution between E3 and E4 and their landscape of genomic divergence.

(Induction of apoptosis, GO:0006917) were significantly enriched in high-elevation E2 (SI Appendix, Table S10), and response to radiation (GO:0009314) and light stimulus (GO:0009416) were significantly enriched in high-elevation E4 (SI Appendix, Table S11). GO categories associated with blood circulation (GO:0008015) and metabolism process (response to lipid, GO:0033993) are present in low-elevation E3 (SI Appendix, Table S12).

To verify the function of the genes in HDRs, we performed real-time quantitative PCR on seven candidate genes from three different tissues (heart, lung, and brain). These seven candidate genes were identified based upon what they potentially contribute to high-altitude adaptation (SI Appendix, Table S13). Four of these differ in their levels of expression between low-elevation and high-elevation lineages (Fig. 7D), including two related to cardiac function in the heart [Troponin C1 Slow (TNNC1) (51) and Adenosine A1 Receptor (ADORA1) (52)] and two hypoxia-related genes [DnaJ Homolog Subfamily C Member 8 (DNAJC8) in the brain (53) and Laminin Subunit Beta 3 (LAMB3) in the lung (54)].

Discussion

Genomic Isolation and Speciation Between E and W Matrilines. Population genomic analyses find substantial genomic isolation between W and E, with limited directional gene flow. Gene flow has been found only in one geographically restricted area, where no obvious geographic barrier (e.g., mountains) exists that could restrict dispersal. Further, the genetic patterns indicate that frogs from both E and W cross the YZR within their ranges and reject the null hypothesis of panmixia within a single species. Neutral processes, including divergence, changes of current, and ancestral effective population size (Ne), and levels of gene flow have the potential to explain the broad pattern of differentiation between W and E. However, the HDRs between W and E exhibit significantly more genetic differentiation than our simulation of male reproductive genes, for instance, in primates (55), birds (19), fruit flies (56), and amphibians (57). Taken together with the fact that W, E2, and E4 occupy similar climatic environments (SI Appendix, Fig. S15) and display no differences in morphology, we suggest that endogenous selection is a dominant factor building reproductive isolation, resulting in speciation between W and E matrilines.

Populations from the geographic region of admixture between E and W matrilines occur along a narrow valley containing a tributary of the YZR. They have the mitogenome of E, whereas nuclear genes of W strongly dominate. Such mitonuclear discordance is common in vertebrates (58); however, here, a cohort of N. parkeri genomes offers an unusual opportunity to decipher the genomic pattern and the genetic mechanism of the phenomenon. Population genomic analyses reveal the E introgressed regions in hybrids are small and randomly dispersed in the genome. Those short introgressed segments suggest to us that they formed by hybridization, and were subsequently broken into even smaller sizes by recombination operating over many generations (59, 60). We detect no signal of selection in either the regions of introgression or the mitochondrial genome, and no significant enrichment of the pathways relevant to mitochondria or of the GOs of genes in introgressed regions. Thus, the mitonuclear discordance is not caused by natural selection driven by environmental factors (29, 61, 62). A likely scenario is establishment of reproductive isolating barriers (RIBs) resulting from secondary contact with historical and limited introgression after mid-Pleistocene divergence (63, 64). We hypothesize that hybrids formed historically only between females of E and males of W upon secondary contact, with hybrid females continuing to backcross to the males of W, leading to postzygotic, prezygotic, or both categories of RIBs.

Ecological Adaptation Within E Matrilines. E Tibetan frogs primarily segregate into high-elevation (>3,700 m, E2 and E4) and low-elevation (<3,700 m, E1 and E3) populations (Fig. 14). This pattern likely reflects an ecological stratification in the southeastern QTP, including oxygen partial pressure (36), UV radiation (37), precipitation (38), and ambient temperature (39). If ecological selection is one of the forces that drives differentiation of the populations of E, one expects to find an imprint on genomic regions involved in ecological adaptation (27, 65).
The HDRs of each subpopulation of E exhibit signals of selection compared with the background, including low \( \pi \), increased dxy between populations, and DAF. GO enrichment analyses point to environmental drivers of divergence. No clear geographic barriers separate the low-elevation sympatric pair E1 and E3. Although they were formed at different times (Fig. 3A), they share several GO functional categories, such as blood circulation and metabolic processes. Given the phylogeny and population history, these functions likely evolved independently in each lineage. The histological, physiological, and expressional evidence (Fig. 7) corresponds to the scenario that ecological stratification in the southeastern QTP has greatly promoted adaptive population diversification in situ.

In summary, \( N. parkeri \) is a useful model to investigate the processes and genetic bases of speciation along geographic and environmental gradients. We argue that natural selection plays important roles in driving continuing divergence within the species, and even in maintaining it. The extreme environments of the Tibetan Plateau can drive the rapid evolution of species [e.g., yak (33)]. Given the rapidity of changes and the challenging environment, the area is a natural laboratory for studying how selection drives adaptation; how environments influence evolutionary history; and, in some cases, how speciation can occur.

Material and Methods

Sample Collection and Sequencing. Tibetan frogs (45 in total) were selected for our genomic study based on matrilineal (mtDNA) genealogies (42). An additional 18 individuals come from the area of mixed clades in a tributary of the YZR (localities 25, 26, and 27; Fig. 1). One individual of \( Nanorana pleskei \) was used as the outgroup taxon (66). All collections were made according to

Fig. 5. Identification of HDRs. (A) Distribution of \( F_{ST} \)-based statistic of PBS\( W \). HDRs in top 2.5% PBS distribution, light blue; outside HDRs, gray. (B) Comparisons of HDRs of PBS\( W \) in terms of \( F_{ST} \), \( \pi \), dxy, and DAF versus the genomic background. (C) Distribution of PBS\( E_1 \). (D) Comparisons of HDRs of PBS\( W \) with the genomic background. Asterisks designate levels of significance between HDRs and outside HDRs by a two-tailed Mann–Whitney test (* \( P < 0.01 \); ** \( P < 1e-8 \); *** \( P < 2.2e-16 \)).

Fig. 6. Distribution of observed genome-wide top 2.5% PBS\( W \) values compared with the simulated PBS\( W \) values under the full model.
Table 2. GO analysis of genes located in regions that strongly differentiated W from all four subpopulations of E

| Category                     | Term                                      | No. of genes | P value |
|------------------------------|-------------------------------------------|--------------|---------|
| Cluster 1                    | GO:0048538-thymus development             | 5            | 0.00    |
|                              | GO:0048534-hemopoietic or lymphoid organ  | 12           | 0.01    |
|                              | development process                       |              |         |
|                              | GO:0002520-immune system development      | 12           | 0.01    |
| Cluster 2                    | GO:0003006-reproductive developmental      | 15           | 0.00    |
|                              | process                                   |              |         |
|                              | GO:0048610-reproductive cellular process   | 11           | 0.00    |
|                              | GO:0048609-reproductive process in a       | 18           | 0.01    |
|                              | multicellular organism                     |              |         |
|                              | GO:0032504-multicellular organism         | 18           | 0.01    |
|                              | reproduction                               |              |         |
|                              | GO:0019953-sexual reproduction             | 17           | 0.01    |
|                              | GO:0007281-germ cell development           | 7            | 0.01    |
|                              | GO:0007276-gamete generation              | 15           | 0.01    |
|                              | GO:0048232-male gamete generation         | 12           | 0.02    |
|                              | GO:007283-spermatogenesis                 | 12           | 0.02    |
| Cluster 3                    | GO:0035270-endocrine system development   | 6            | 0.01    |
|                              | GO:0030325-adrenal gland development       | 3            | 0.01    |

Annotation clusters with an enrichment score of ≥2 are shown.

To infer admixture events, we applied the D-statistic using the qpDstat module in the ADMIXTOOLS package (78). We also adopted a tree-based approach, which was implemented in TreeMix (79), to verify the existence of gene flow through modeling the migration values set from 0 to 4 with a block of 5,000 SNPs.

Demographic Analysis Using G-PhoCS and PSMC. G-PhoCS (46) was employed to infer the complete demographic history for N. parkeri, including population divergence times, ancestral population size, and migration rates based on 1,000 neutral loci (80). The parameters were inferred in a Bayesian manner using Markov Chain Monte Carlo to jointly sample model parameters and genealogies of the input loci (46). Migration scenarios were added by combining results of D-statistic tests, Frappe, and TreeMix because G-PhoCS often have limited ability to characterize complex migration scenarios (81). Additionally, two postdivergence migration bands were added to test if gene flow occurred between W and E since they separated. Each Markov chain was run for 2,000,000 generations while sampling parameter values every 20th iteration. Burn-in and convergence of each run were determined with TRACER 1.5 (82). More information about the control file of G-PhoCS is provided in SI Appendix, section 1.2. Divergence times in units of years, effective population sizes, and migration rates were calibrated by the estimates of generation time and neutral mutation rate from previous studies (45, 83). We repeated the G-PhoCS analysis with four separate runs to obtain reliable and stable estimates for the demographic parameters. To validate the G-PhoCS inferences and test if the divergent gene flow to W was correctly identified, we used ms (84) to perform simulations under different migration scenarios between W and E (SI Appendix, section 1.3). We used the inferred demographic parameters to produce 50 Mbp of sequences and compared the observed patterns of differentiation for W-E1 with those under different migration scenarios.

The trajectory of demographic histories for the five populations of Tibetan frogs was inferred by the PSMC model (85). Because PSMC has high false-negative rates at low sequence coverage, we restricted this analysis to the individual in each group with highest coverage (≥15x). In addition, a correction factor (N4) was invoked to correct for the false-negative rate caused by the shallow sequence depth (80). The PSMC analysis was set as the following parameters: -N25 -115 -r5 -b -p. A bootstrapping approach with 100 replicates was performed to assess the variation in the inferred Ne trajectories. A generation time of 5 y and a neutral mutation rate of 0.776e-09 per site per year were used to convert the population sizes and scaled time into real sizes and time (45, 83).

The program fastsimcoal2 (86) was used to estimate the extent of population expansion within 30,000 y based on the model of G-PhoCS. Fifty time simulations were performed, and the result with the highest likelihood was kept. For each run, demographic estimates were obtained from 100,000 simulations (~100,000) and 40 expectation/conditional maximization cycles (~L 40) per parameter file.

Inference of Mitochondrial Phylogeny and Hybrid Ancestry Assignment. We used BWA-ALN (v.0.7.4) to map all raw sequence reads from each individual to the previously assembled complete mitochondrial genome of N. pleskei (87), the closest relative of our Tibetan frog. SNPs were identified and filtered based on the same nuclear genome. The matrilineal (mitochondrial) genealogy was inferred using the ML method based on the GTRGAMMA model implemented in PAML (v.4.9) (88). The PAML analysis was set as the following parameters: -L 40 -t 15 -r 5 -b -p. A bootstrapping approach with 100 replicates was performed to assess the variation in the inferred Ne trajectories. A generation time of 5 y and a neutral mutation rate of 0.776e-09 per site per year were used to convert the population sizes and scaled time into real sizes and time (45, 83).

The program fastsimcoal2 (86) was used to estimate the extent of population expansion within 30,000 y based on the model of G-PhoCS. Fifty time simulations were performed, and the result with the highest likelihood was kept. For each run, demographic estimates were obtained from 100,000 simulations (~100,000) and 40 expectation/conditional maximization cycles (~L 40) per parameter file.

Annotation clusters with an enrichment score of ≥2 are shown.

To infer admixture events, we applied the D-statistic using the qpDstat module in the ADMIXTOOLS package (78). We also adopted a tree-based approach, which was implemented in TreeMix (79), to verify the existence of gene flow through modeling the migration values set from 0 to 4 with a block of 5,000 SNPs.

Demographic Analysis Using G-PhoCS and PSMC. G-PhoCS (46) was employed to infer the complete demographic history for N. parkeri, including population divergence times, ancestral population size, and migration rates based on 1,000 neutral loci (80). The parameters were inferred in a Bayesian manner using Markov Chain Monte Carlo to jointly sample model parameters and genealogies of the input loci (46). Migration scenarios were added by combining results of D-statistic tests, Frappe, and TreeMix because G-PhoCS often have limited ability to characterize complex migration scenarios (81). Additionally, two postdivergence migration bands were added to test if gene flow occurred between W and E since they separated. Each Markov chain was run for 2,000,000 generations while sampling parameter values every 20th iteration. Burn-in and convergence of each run were determined with TRACER 1.5 (82). More information about the control file of G-PhoCS is provided in SI Appendix, section 1.2. Divergence times in units of years, effective population sizes, and migration rates were calibrated by the estimates of generation time and neutral mutation rate from previous studies (45, 83). We repeated the G-PhoCS analysis with four separate runs to obtain reliable and stable estimates for the demographic parameters. To validate the G-PhoCS inferences and test if the divergent gene flow to W was correctly identified, we used ms (84) to perform simulations under different migration scenarios between W and E (SI Appendix, section 1.3). We used the inferred demographic parameters to produce 50 Mbp of sequences and compared the observed patterns of differentiation for W-E1 with those under different migration scenarios.

The trajectory of demographic histories for the five populations of Tibetan frogs was inferred by the PSMC model (85). Because PSMC has high false-negative rates at low sequence coverage, we restricted this analysis to the individual in each group with highest coverage (≥15x). In addition, a correction factor (N4) was invoked to correct for the false-negative rate caused by the shallow sequence depth (80). The PSMC analysis was set as the following parameters: -N25 -115 -r5 -b -p. A bootstrapping approach with 100 replicates was performed to assess the variation in the inferred Ne trajectories. A generation time of 5 y and a neutral mutation rate of 0.776e-09 per site per year were used to convert the population sizes and scaled time into real sizes and time (45, 83).

The program fastsimcoal2 (86) was used to estimate the extent of population expansion within 30,000 y based on the model of G-PhoCS. Fifty time simulations were performed, and the result with the highest likelihood was kept. For each run, demographic estimates were obtained from 100,000 simulations (~100,000) and 40 expectation/conditional maximization cycles (~L 40) per parameter file.

Inference of Mitochondrial Phylogeny and Hybrid Ancestry Assignment. We used BWA-ALN (v.0.7.4) to map all raw sequence reads from each individual to the previously assembled complete mitochondrial genome of N. pleskei (87), the closest relative of our Tibetan frog. SNPs were identified and filtered based on the same nuclear genome. The matrilineal (mitochondrial) genealogy was inferred using the ML method based on the GTRGAMMA model implemented in PAML (v.4.9).
RAxML (v.8.1.15) (74). PCAdmix (88) was used to identify the ancestries in each of the 18 hybrid individuals. We used W and E4 as the ancestral populations because they were the geographically closest populations and showed the highest signal of admixture in the $f_3$ test. To prevent high-linkage blocks from having excessive influence on the inferred ancestry of a region, SNPs were thinned with $r^2 > 0.80$ in all ancestral and admixed groups. PCA was performed using a window of 40 SNPs. A default calling threshold of 0.9 was used as the criterion to assign ancestry. However, systematic errors, such as incomplete lineage sorting and estimation errors, may have contributed to similar introgression signals in the PCAdmix analysis. To estimate these, we reanalyzed the PCAdmix analysis using non-E4 individuals as hybrids.

Inference HDRs. Genomic differentiation was calculated based on the PBS under the topological transformation described previously (89). The PBS value estimates the amount of sequence change along a population branch since its divergence from other branches of a population tree. First, pairwise $F_{ST}$ statistics were calculated using Weir and Cockerham’s method (90), implemented in VCFtools v0.1.11 (91), with nonoverlapping 50-kb genomic windows. Negative values of $F_{ST}$ was treated as 0. Windows less 20 kb in length were excluded for further analysis. A log-transformation $F_{ST}$ was used for the PBS calculation (92). The length of the branch leading to W since the divergence from all subpopulations of E was estimated as follows:

$$PBS_{\text{West}} = \frac{3T_{(\text{West}, E1)} + T_{(\text{West}, E2)} + T_{(\text{West}, E3)} + T_{(\text{West}, E4)} - T_{(E1, E2)} - T_{(E1, E3)} - T_{(E1, E4)}}{6}.$$ 

Calculations for branch lengths leading to subpopulations E2, E3, and E4 were similar to the equation for E1. HDRs were defined as the upper 2.5% of each PBS distribution. To avoid the potential assembly or mapping errors at the ends of scaffolds, we further removed high divergence peaks less than two consecutive windows in both ends of scaffolds in both sides.

Characterization of HDRs in Terms of $d_{xy}$, $d_{df}$, $\pi$, and DAF. In addition to $F_{ST}$ and PBS, we applied several population genomic parameters to quantify and compare the outlier windows with the background genome, including the following: $d_{xy}$ between populations (93), $d_{df}$ between populations (19), within-population $\pi$ level (19, 23, 91), DAF, and population-scaled recombination rates ($\rho$). We treated N. pleskei as the outgroup when comparing differences in DAF between W and E, and treated three nonadmixture W individuals (Fig. 2) as the outgroup when comparing DAF differences among subpopulations of E.

Fig. 7. Morphological and physiological changes associated with elevation. (A) Bar plots of the differences in numbers of granular glands in the middorsal skin between low-elevation (2,968 m, E1) and high-elevation (4,859 m, E4) populations (two-tailed test: $P < 0.05$). (B) Hb levels (grams per deciliter) at different elevations. Red lines show the best-fit regression line based on a third-order polynomial equation. The 95% confidence interval is shown in gray. (C) TOC (µmol/L) in low (E1) and high (E4) elevations. (D) Expression level of DNAJC8 in the brain, TNNC1 and ADORA1 in the heart, and LAMB3 in the lung from low-elevation (E1) and high-elevation (E4) populations of E, respectively. Eight replicates were performed for each group. Statistically significant differences in differential expression are indicated by asterisk(s) (two-tailed t test: *$P < 0.05$; **$P < 0.01$).
GO Enrichment Analysis. GO enrichment analysis was performed using DAVID (Database for Annotation Visualization and Integrated Discovery) (94). GO terms with p-value <0.05 than two genes were then excluded from further analysis. GO terms with a P value <0.05 were considered to be significantly enriched.

Phenotype Data from Physiological and Morphological Detection. Hb levels and tissue oxygen content (TOC) were collected in the field during June and July 2015. Measurements were made in five communities in E along an elevation gradient: Nyingchi (29.61 °N, 94.36°E, 2,968 m above sea level), Lhasa (29.67°N, 90.88°E, 3,671 m above sea level), Zhaxiagang (29.75°N, 91.95°E, 4,011 m above sea level), Ruduo (29.70°N, 92.23°E, 4,368 m above sea level), and Milashankou (29.80°N, 92.34°E, 4,859 m above sea level). At least 10 individuals were measured for each community. Hb level in peripheral blood was measured from 53 adult frogs, and muscular oxygen content was measured from 69 adult frogs. Hb was determined using Mission Plus Hb, blood was measured from 53 adult frogs, and muscular oxygen content was measured from 69 adult frogs. Hb was determined using Mission Plus Hb, blood was measured from 53 adult frogs, and muscular oxygen content was measured from 69 adult frogs.

RNA Isolation and Reverse Transcriptase PCR Assay. Total RNAs from populations at low (Nyingchi, 2,968 m) and high (Milashankou, 4,859 m) elevations of E were extracted using the TRIzol total RNA extract kit (Tiangen). Reverse transcriptase cDNA was carried out using the First Strand cDNA synthesis kit (Fermentas) to prepare templates for real-time quantitative PCR. The primers of candidate genes used for real-time PCR are displayed in Table S14. Actin, Beta (ACTB) was used as a loading control.

Acknowledgments. We thank Michael W. Nachman (University of California, Berkeley), Weivi Zhai (Genome Institute of Singapore), and two reviewers for helpful comments and suggestions. We also thank Na Lin of Anhui University for assistance in file preparation. This work was supported by the Strategic Priority Research Program (B) (Grant XDB13020200 to J.C.) of the Chinese Academy of Sciences (CAS), National Natural Science Foundation of China (Grants 91431105 and 31622052 to J.C.), and the Animal Branch of the Germplasm Bank of Wild Species of CAS (Large Research Infrastructure Funding) (J.C.). W.-W.Z. was supported by Strategic Priority Research Program (B) (Grant XDB03030107) of CAS and the National Key Research and Development Program of China (Grant 2017YFC0505202). G.-D.W. was supported by the 13th Five-year Informatization Plan of CAS (Grant No. XHH13503-05). J.C. and G.-D.W. are supported by the Youth Innovation Promotion Association, CAS.

References

1. Seehausen O, et al. (2014) Genomics and the origin of species. Nat Rev Genet 15: 176–192.
2. Coyne J, Orr H (2004) Speciation (Sinauer, Sunderland, MA).
3. Mayr E (1999) Speciation and the Origin of Species (Belknap Press of Harvard Univ Press, Cambridge, MA).
4. Dobzhansky T (1937) Genetics and the Origin of Species (Columbia Univ Press, New York).
5. Hewitt GM (2000) Is it possible to derive evolutionary consequences of climatic fluctuations in the Quaternary. Philos Trans R Soc Lond B Biol Sci 355:183–195, discussion 195.
6. Avice JC (2001) Phylogeography: The History and Formation of Species (Harvard Univ Press, Cambridge, MA).
7. Funk DJ (1998) Isolating a role for natural selection in speciation: Host adaptation and sexual isolation in Neohamulus bebbianus leaf beetles. Evolution 52:1744–1759.
8. Schluter D (2000) The Ecology of Adaptive Radiation (Oxford Univ Press, Oxford).
9. Nosil P (2012) Ecological Speciation (Oxford Univ Press, Oxford).
10. Turelli M, Barton NR (2001) Evolutionary consequences of speciation. Trends Ecol Evol 16:330–342.
11. Nosil P, Feder JL (2012) Genetic divergence during speciation: Causes and consequences. Philos Trans R Soc Lond B Biol Sci 367:332–342.
12. Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. Genet 283:362–370.
13. Nosil P, Feder JL, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous speciation in the Hawaiian honeycreeper Drepanis rubricauda. Evolution 63:1541–1554.
14. Nosil P (2008) Speciation with gene flow: Models and data. Annu Rev Ecol Evol Syst 41:215–230.
15. Mayr E (1966) Animal Species and Evolution (Belknap Press of Harvard Univ Press, Cambridge, MA).
16. Hörber AW, Tautz D (2010) Understanding the onset of hybrid speciation. Trends Genet 26:54–58.
17. Mayr E (1999) Systematics and the Origin of Species from the Viewpoint of a Zoologist (Harvard Univ Press, Cambridge, MA).
18. Nosil P, Funk BJ, Vermeiren D (2014) Genomic islands of speciation separate cichlid ecomorphs in Lake Malawi. PLoS One 9:e89323.
Malone JH, Michalak P (2008) Physiological sex predicts hybrid sterility regardless of genotype. Science 319:59.

Toews DP, Breisford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. Mol Ecol 21:3907–3930.

Pool JE, Nielsen R (2009) Inference of historical migration changes in rate from the lengths of migrant tracts. Genetics 181:711–719.

Harris K, Nielsen R (2013) Inferring demographic history from a spectrum of shared haplotype lengths. PLoS Genet 9:e1003521.

Irvin DE (2012) Local adaptation along smooth ecological gradients causes phylogeographic breaks and phenotypic clustering. Am Nat 180:35–49.

Currat M, Ruedi M, Petit RJ, Excoffier L (2008) The hidden side of invasions: Massive introgression by local genes. Evolution 62:1908–1920.

Chan KM, Levin SA (2005) Leaky prezygotic isolation and porous genomes: Rapid introgression of maternally inherited DNA. Evolution 59:720–729.

Collard C, et al. (2010) Strong reproductive barriers in a narrow hybrid zone of West-Mediterranean green toads (Bufo viridis subgroup) with Plio-Pleistocene divergence. BMC Evol Biol 10:232.

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. Mol Ecol 14:671–688.

Che J, et al. (2010) Spiny frogs (Paini) illuminate the history of the Himalayan region and Southeast Asia. Proc Natl Acad Sci USA 107:13765–13770.

Sambrook J, Russel DW (2012) Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press, Plainview, NY).

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

Li H, et al.; 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079.

DePristo MA, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491–498.

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.

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.

Tang H, Peng J, Wang P, Kisch NJ (2005) Estimation of individual admixture: Analytical and study design considerations. Genet Epidemiol 28:289–301.

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

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.

Liu L, Yu L, Pearl DK, Edwards SV (2009) Estimating species phylogenies using coalescence times among sequences. Syst Biol 58:468–477.

Shaw TI, Ruan Z, Glenn TC, Liu L (2013) STRAWE: Species TRee analysis web server. Nucleic Acids Res 41:W238–W241.

Patterson N, et al. (2012) Ancient admixture in human history. Genetics 192:1065–1093.

Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet 8:e1002967.

Wang GD, et al. (2016) Out of southern East Asia: The natural history of domestic dogs across the world. Cell Res 26:21–33.

Freedman AH, et al. (2014) Genome sequencing highlights the dynamic early history of dogs. PLoS Genet 10:e1004166.

Rambaut A, Drummond AJ (2007) Tracer 1.5. Available at tree.bio.ed.ac.uk/software/tracer. Accessed July 3, 2011.

Ma X, Lu X (2009) Sexual size dimorphism in relation to age and growth based on skeletochronological analysis in a Tibetan frog. Amphiibia-Reptilia 30:351–359.

Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics 18:337–338.

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

Excoffier L, Dupanloup I, Huerta-Sánchez E, Sousa VC, Foll M (2013) Robust demographic inference from genomic and SNP data. PLoS Genet 9:e1003905.

Chen G, Wang B, Liu J, Xie F, Jiang J (2011) Complete mitochondrial genome of Nanorana pleskei (Amphibia: Anura: Dicroglossidae) and evolutionary characteristics. Curr Zool 57:S75–805.

Brisbin A, et al. (2012) PCAdmix: Principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. Hum Biol 84:334–364.

Lachance J, Tishkoff SA (2014) Biased gene conversion skews allele frequencies in human populations, increasing the disease burden of recessive alleles. Am J Hum Genet 95:408–420.

Weir BS, Cockerman CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370.

Danecek P, et al.; 1000 Genomes Project Analysis Group (2011) The variant call format and VCFtools. Bioinformatics 27:2156–2158.

Yi X, et al. (2010) Sequencing of 50 human exomes reveals adaptation to high altitude. Science 329:75–78.

Al H, et al. (2015) Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. Nat Genet 47:217–225.

Huang W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4:44–57.