



# Species groups distributed across elevational gradients reveal convergent and continuous genetic adaptation to high elevations

Yan-Bo Sun<sup>a,1</sup>, Ting-Ting Fu<sup>a,b,1</sup>, Jie-Qiong Jin<sup>a</sup>, Robert W. Murphy<sup>a,c</sup>, David M. Hillis<sup>d,2</sup>, Ya-Ping Zhang<sup>a,e,f,2</sup>, and Jing Che<sup>a,e,g,2</sup>

<sup>a</sup>State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, 650223 Kunming, China; <sup>b</sup>Kunming College of Life Science, University of Chinese Academy of Sciences, 650204 Kunming, China; <sup>c</sup>Centre for Biodiversity and Conservation Biology, Royal Ontario Museum, Toronto, ON M5S 2C6, Canada; <sup>d</sup>Department of Integrative Biology and Biodiversity Center, University of Texas at Austin, Austin, TX 78712; <sup>e</sup>Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, 650223 Kunming, China; <sup>f</sup>State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan University, 650091 Kunming, China; and <sup>g</sup>Southeast Asia Biodiversity Research Institute, Chinese Academy of Sciences, Yezin, 05282 Nay Pyi Taw, Myanmar

Contributed by David M. Hillis, September 7, 2018 (sent for review August 7, 2018; reviewed by John H. Malone and Fuwen Wei)

Although many cases of genetic adaptations to high elevations have been reported, the processes driving these modifications and the pace of their evolution remain unclear. Many high-elevation adaptations (HEAs) are thought to have arisen *in situ* as populations rose with growing mountains. In contrast, most high-elevation lineages of the Qinghai-Tibetan Plateau appear to have colonized from low-elevation areas. These lineages provide an opportunity for studying recent HEAs and comparing them with ancestral low-elevation alternatives. Herein, we compare four frogs (three species of *Nanorana* and a close lowland relative) and four lizards (*Phrynocephalus*) that inhabit a range of elevations on or along the slopes of the Qinghai-Tibetan Plateau. The sequential cladogenesis of these species across an elevational gradient allows us to examine the gradual accumulation of HEA at increasing elevations. Many adaptations to high elevations appear to arise gradually and evolve continuously with increasing elevational distributions. Numerous related functions, especially DNA repair and energy metabolism pathways, exhibit rapid change and continuous positive selection with increasing elevations. Although the two studied genera are distantly related, they exhibit numerous convergent evolutionary changes, especially at the functional level. This functional convergence appears to be more extensive than convergence at the individual gene level, although we found 32 homologous genes undergoing positive selection for change in both high-elevation groups. We argue that species groups distributed along a broad elevational gradient provide a more powerful system for testing adaptations to high-elevation environments compared with studies that compare only pairs of high-elevation versus low-elevation species.

*Nanorana* | *Phrynocephalus* | altitudinal gradient | high-elevation adaptation | DNA repair

Groups of organisms that are broadly distributed with respect to diverse environmental conditions provide an excellent opportunity to study the genetic basis of adaptation. High elevations offer extreme environments with low levels of oxygen, high levels of UV radiation, and cold temperatures, which challenge the survival of many species. Animals that have evolved in high-elevation environments typically exhibit numerous adaptations to counter these environmental challenges, such as altered developmental rates and body sizes, increased pulmonary ventilation, and increased hemoglobin oxygen affinity (1–6). In particular, the Qinghai-Tibetan Plateau and its surrounding slopes constitute a large high-elevation region with a mean elevation exceeding 4,000 m, as well as substantial spatial variation in both oxygen content and solar UV radiation (7), making it an ideal region for study of high-elevation adaptation (HEA).

Most previous studies of the genetic processes of HEA have compared species or populations from high elevations above 3,500 m with those from low elevations to identify sequence variation and/or expression shifts in the high-elevation group (8–15). In these studies, changes in genes involved in energy metabolism and responses to hypoxia have been routinely identified (8–15). However, “high elevation” is not a single point above sea level but, instead, represents a continuum. Despite this fact, studies along a broad elevational continuum are lacking, which has hindered our understanding of the genetic processes that result in HEA. Furthermore, the degree of evolutionary convergence in adaptation (e.g., refs. 16–19) among different species that live at similar elevations is unknown.

Multiple groups, especially of frogs and lizards, are distributed across a large range of elevations on the Qinghai-Tibetan Plateau and its adjacent slopes, and can be used to explore the genetic processes involved in HEA. For example, there are three species of the frog genus *Nanorana* (*Nanorana parkeri*, *Nanorana liebigii*, and *Nanorana phrynooides*) that occur up to elevations of

## Significance

Organisms living in extreme environments are useful for studying the process of adaptation. We studied two distantly related groups that are distributed across a broad elevational gradient on and near the Qinghai-Tibetan Plateau and identified molecular adaptations to increasing elevations. We show that high-elevation adaptation (HEA) emerged soon after a split from low-elevation lineages, and adaptations continue to evolve in species that inhabit increasingly high elevations. Genes related to DNA repair and energy metabolism evolved rapidly, suggesting a crucial role of these genes in HEA. Moreover, we observed common patterns of HEA for similar functions between distantly related lineages, although these functional changes often involved different specific genes.

Author contributions: D.M.H., Y.-P.Z., and J.C. designed research; Y.-B.S. and T.-T.F. analyzed data; J.-Q.J. did RNA extraction and prepared sampling; and Y.-B.S., T.-T.F., R.W.M., D.M.H., Y.-P.Z., and J.C. wrote the paper.

Reviewers: J.H.M., University of Connecticut; and F.W., Chinese Academy of Sciences.

The authors declare no conflict of interest.

Published under the PNAS license.

Data deposition: All the original RNA-sequencing data reported in this paper have been deposited into the Genome Sequence Archive, [gsa.big.ac.cn](http://gsa.big.ac.cn) (accession no. CRA000352).

<sup>1</sup>Y.-B.S. and T.-T.F. contributed equally to this work.

<sup>2</sup>To whom correspondence may be addressed. Email: [chej@mail.kiz.ac.cn](mailto:chej@mail.kiz.ac.cn), [zhangyp@mail.kiz.ac.cn](mailto:zhangyp@mail.kiz.ac.cn), or [dhillis@austin.utexas.edu](mailto:dhillis@austin.utexas.edu).

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1813593115/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1813593115/-DCSupplemental).

Published online October 22, 2018.

4,700 m, 3,500 m, and 2,400 m, respectively (20, 21), in addition to species related to *Nanorana* that are distributed at relatively low elevations. Studies of *N. parkeri* have identified several physiological and phenotypic adaptations that help this species cope with hypoxia and cold temperatures, including a slower developmental rate and a smaller body size (4). Similarly, three species of the lizard genus *Phrynocephalus* (*Phrynocephalus erythrurus*, *Phrynocephalus vlangalii*, and *Phrynocephalus putjatai*) occur at 4,300–5,200 m, 2,200–4,400 m, and 2,200–3,000 m, respectively, and, together with their low-elevation relatives, they are distributed across a broad elevational range. The high-elevation lizards have also evolved some physiological innovations, such as higher capability to carry oxygen (5). Given these broad elevational distributions, these two groups of ectotherms provide excellent models to explore the dynamics and convergent evolution of HEA across an elevational gradient.

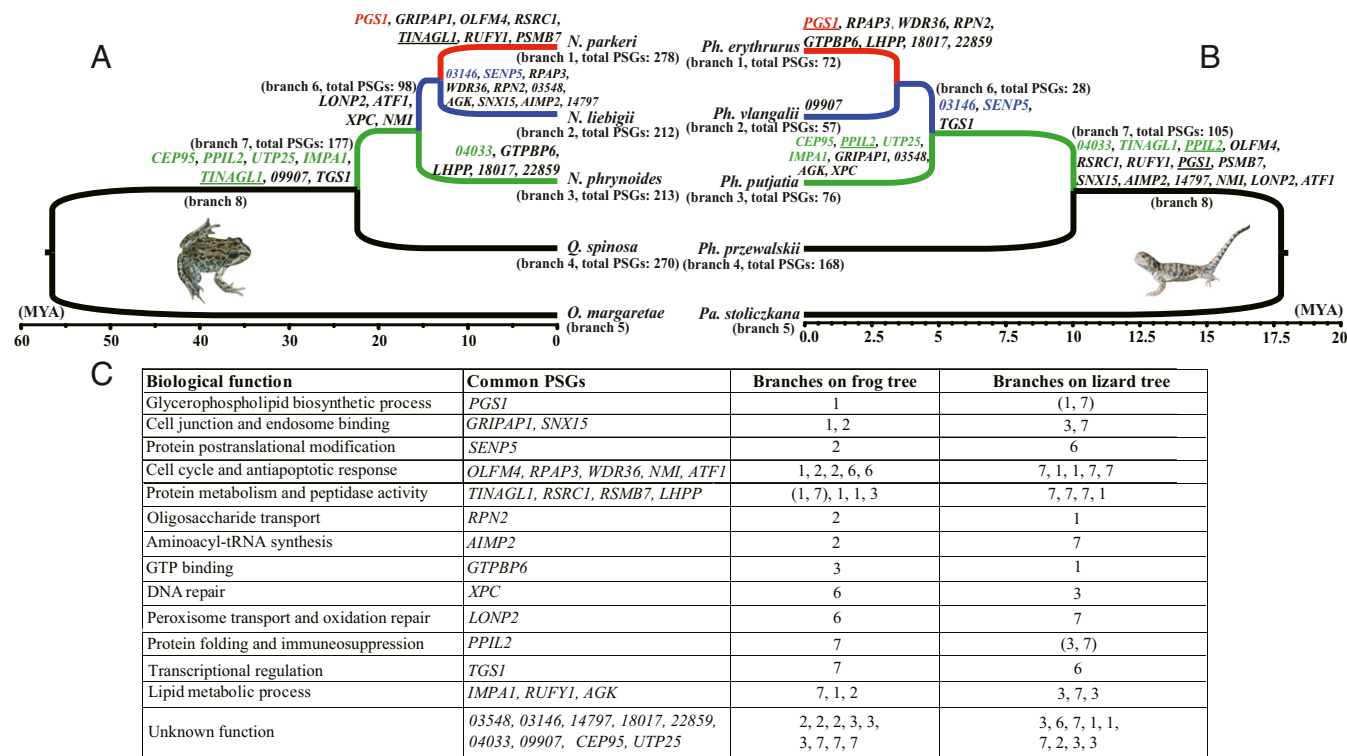
Advances in high-throughput sequencing, especially RNA sequencing (RNA-Seq) (22), enable the collection of gene sequences for nonmodel animals. These advances greatly facilitate comparative evolutionary studies (23). We analyzed the transcriptomes of the three high-elevation *Nanorana* and *Phrynocephalus* species noted above, as well as two low-elevation (~1,000 m) relatives of each genus. Multitissue transcriptomic sequencing allowed construction of >7,000 high-quality orthologous genes for both groups. Evolutionary analyses involved the detection of positive selection among individual genes on different branches, inferences of the timing of these changes with respect to changes in elevational distribution, and identification of convergent evolution of functional adaptations and specific genes.

## Results and Discussion

**Sampling and Sequencing.** For the frog group, four closely related species, including *N. parkeri*, *N. liebigii*, *N. phrynooides*, and *Quasipaa spinosa*, were collected for RNA-Seq. Among them, *N. parkeri* is the highest distributed species (specimens were collected at elevations >4,000 m). *N. liebigii* and *N. phrynooides* were collected at middle elevations (~3,000 m and ~1,800 m, respectively), and *Q. spinosa* was collected at low elevations (<1,000 m) (Fig. 1A and SI Appendix, Fig. S1). The low-elevation species *Odorrana margaritae* (24) was collected as an outgroup for comparative analysis.

Multiple tissues (brain, heart, liver, skin, ovary, and testis) were collected from each species of frog to maximize the number of expressed genes that were examined. *N. parkeri*, *N. liebigii*, *N. phrynooides*, and *Q. spinosa* had 282,884,179, 197,229,347, 96,028,690, and 287,795,501 high-quality reads, respectively (SI Appendix, Table S1). The final assemblies showed both high-contig N50 values and read-mapping ratios (SI Appendix, Table S1). All of the original RNA-Seq data were deposited into the open-access Genome Sequence Archive ([gsa.big.ac.cn](http://gsa.big.ac.cn)) under accession no. CRA000352.

For each species, we generated a de novo transcriptomic assembly, from which we examined a total of 8,138 orthologous genes among the five frogs. The average mean identity score (MIS) of these gene alignments is 93%, indicating reliable predictions of orthology. Based on the “supergene” alignment that was constructed from concatenations of all of the high-quality orthologous genes, we constructed a maximum likelihood (ML) tree (Fig. 1A). The divergence time at the root was estimated to



**Fig. 1.** Time-scaled phylogenetic trees for frogs (A) and lizards (B) distributed across an elevational gradient of the Qinghai-Tibetan Plateau. Black lineages represent lowland distributions, green lineages represent distributions up to about 2,000 m, blue lineages represent distributions from roughly 2,000–3,500 m, and red lineages represent distributions from about 3,500–4,500 m. The branch number and the number of PSGs are indicated beside each tip or branch of the two trees, and PSGs that evolved in parallel are shown on the respective branches of each tree. (C) All of the parallel PSGs that were identified in both groups of high-elevation species, and their functions (where known), are shown. Parallel PSGs that evolved at the same change in elevational distribution are shown in the respective color of that lineage (green, blue, or red, respectively). Underlined genes showed positive selection on more than one branch of the frog or lizard tree. The functional unknown PSGs are represented by the last five digits of the anole lizard’s ENSEMBL gene identifiers.

be 56.5 (29.1–92.8) Mya, which provides an appropriate time interval for identifying episodic positive selection across the group (25).

For the lizard group, all of the transcriptomes of *P. przewalskii*, *P. putjataia*, *P. vlangalii*, and *P. erythrurus* were obtained from previous studies (13, 14). *Paralaudakia stoliczkana*, another low-elevation species (occurs at 100–1,700 m; [www.zoology.csd.cn/page/index.vpage](http://www.zoology.csd.cn/page/index.vpage)) that is closely related to the *Phrynocephalus* species, was collected (at ~1,050 m) in this study for outgroup comparison. The full-length transcripts of *P. stoliczkana* were generated in this study through the PacBio sequencing technology. Using the same procedures used for the frog transcriptomes, a total of 7,288 orthologous genes (MIS = 96%) were analyzed and an ML phylogeny was constructed from the concatenated transcripts of the lizards (Fig. 1B). This phylogeny showed shallower evolutionary divergences for the lizards than the frogs, indicating a later colonization of high elevations in these species.

**HEA Is Evident in the Earliest Stages of Elevational Expansion.** Previous studies have shown that the highest elevation lineages (>4,000 m) in each group have many HEAs, especially at the molecular level (14, 26). However, because these studies compared only pairs of high-elevation versus low-elevation species, it is unclear if these adaptations accumulated gradually at increasing elevations or if they evolved at a particular elevational threshold. The two distantly related species groups studied here (Fig. 1A and B) provide opportunities to address these questions through comparative evolutionary analyses. To address these issues, we examined evidence for positive selection along each of the branches of the two phylogenies.

We first estimated the dN/dS ratio on each branch of the two phylogenies. For both groups, 87–96% of the genes display a dN/dS ratio <1 (Dataset S1), indicating that most genes are undergoing purifying selection. This is true for species at all elevational distributions. To detect episodic positive selection in this general background of purifying selection, we used the branch-site likelihood-based test (25, 27), which has been widely used to find signals of positive selection along specific branches of a phylogenetic tree (8–15). Under this model, we were able to identify genes under positive selection on each branch in the phylogeny (Fig. 1A and B), and thereby assign genes under selection to particular elevational shifts in distribution. By comparison with the lowland species, we removed genes that were under positive selection in both low and high elevations, and then identified specific genes that were under positive selection at each increase in elevational distribution.

We identified a total of 835 genes under positive selection exclusively on high-elevational lineages of the frog tree and 301 genes under positive selection exclusively on high-elevational lineages of the lizard tree (Fig. 1A and B). The smaller number of positively selected genes (PSGs) identified in the lizards is not accounted for by their shallower evolutionary divergences compared with the frogs, because lower sequence divergence has little or no correlation with the mean dN/dS ratio (28). Therefore, other biological factors likely account for the fewer PSGs identified in the lizards; one possibility is interspecific hybridization (29), which can dilute signals of positive selection. Nonetheless, we found 32 genes under positive selection on high-elevational lineages of both the frog and lizard trees. These 32 genes are shown along their respective branches in Fig. 1A and B, and the genes are shown with their associated functions (where known) in Fig. 1C.

In the first stages of HEA on the frog tree (Fig. 1A; green lineages, branches 7 and 3), PSGs are significantly over-represented by functions such as regulation of endocytosis and exocytosis [gene ontology (GO):0048261, GO:0045056, and GO:0045921], positive regulation of cellular amine metabolic process (GO:0033240), positive regulation of the HIF-1-alpha

signaling pathway (GO:1902073), respiratory chain complex IV assembly (GO:0008535), response to radiation and postreplication repair (GO:0009314 and GO:0006301), and vasculogenesis (GO:0001570) (SI Appendix, Tables S2–S5). Each of these functions is closely related to the physiological stresses produced by hypoxia or strong UV radiation (UV). For example, endocytosis could be regulated by many genetic and posttranslational alterations, and such regulation is important for cell survival under hypoxia (30). Based on the enrichment results (SI Appendix, Tables S3 and S5), even the frogs distributed at 2,000 m appear to have evolved multiple adaptations related to endocytosis, including modifications of membrane docking and negative regulation of the p38 MAPK cascade (30). This highlights the potential roles of endocytosis and exocytosis in HEA. Further, biogenic amines, such as dopamine, norepinephrine, and serotonin, act as neurotransmitters; are involved in numerous functional pathways; and are particularly susceptible to the effects of hypoxia (31–33). Adaptations related to cellular biogenic amine metabolic processes are important to maintain physiological homeostasis.

The rapid modification and adaptation of several functions mentioned above, especially the response to radiation, cellular responses to hypoxia, and vasculogenesis, are also observed in other high-elevation animals, such as the snub-nosed monkey (15) and Tibetan antelope (8), suggesting they are major targets of natural selection to deal with cellular stresses associated with high-elevation existence, and thus show convergence in function among different groups of high-elevation animals.

The above results strongly suggest that even frogs at moderate elevations (up to about 2,000 m) have already evolved molecular adaptations to reduced oxygen content and/or increased UV radiation. Similar results were also obtained from the distantly related group of lizards (SI Appendix, Tables S6–S9). On the lizard phylogeny, 178 PSGs were identified on the moderate-elevation (up to ~2,000 m) lineages (branches 7 and 3). Of these, there are three PSGs (*PPIL2*, *CPLX1*, and *NPHS2*) undergoing positive selection on both branches 7 and 3 on the lizard phylogeny. The 178 lizard PSGs were significantly enriched in GO terms overrepresented by functions like peroxisomal transport (GO:0043574), DNA damage recognition and repair (GO:0000715, GO:0000717, GO:0006303, and GO:0000726), cellular response to oxidative stress (GO:0034599), response to amine (GO:0014075), and cellular carbohydrate biosynthetic and metabolic process (GO:0034637 and GO:0044262), whose modifications are also closely related to adaptations to high elevations and have been widely observed in other high-elevation animals (8–15).

**Continued Evolution of HEAs at Increasing Elevations.** Given the early emergence of positive selection consistent with HEA (at elevational expansions to just 2,000 m), we next asked if there was continuing adaptation with increasing elevations or, alternatively, if the adaptations to moderate elevations simply allow expansion of the group to higher elevations? Our results showed that all of the higher elevation species showed positive selection for genes associated with HEA beyond the changes observed in the moderate-elevation species, suggesting a continuous increase in HEA with increasing elevational distribution.

In the frog species with an elevational distribution around 3,000 m (*N. liebigii*), the identified PSGs were further enriched in categories of carbohydrate transport (GO:0008643), angiogenesis involved in coronary vascular morphogenesis (GO:0060978), heart morphogenesis and development (GO:0003247 and GO:0007507), mitochondrial DNA repair (GO:0043504), and generation of precursor metabolites and energy (GO:0006091) (SI Appendix, Tables S10–S13). Among these functional categories, angiogenesis is well known for responding to hypoxia by increasing the oxygen supply, a function that has also been observed in other high-elevation animals (8, 9, 15). Combining

the early adaptations evolved by *N. phrynooides*, such as the functions associated with vasculogenesis, our results support a similar role of angiogenesis in high-elevation frogs.

The middle-high elevational lizard (*P. vlangalii*) showed further changes consistent with HEA, with positive selection in genes responsible for nitric oxide-mediated transduction and signaling pathway (GO:0007263 and GO:0038060), cellular response to nitric oxide (GO:0071732 and GO:0071732), cellular glucan metabolic process (GO:0006073), and generation of precursor metabolites and energy (GO:0006091) (*SI Appendix, Tables S14–S17*). This is consistent with a previous study that focused on the HEAs of *P. vlangalii* (13).

The increasing adaptation to high elevations continued in the species with the highest elevational distributions (>4,000 m; frog: *N. parkeri*; lizard: *P. erythrurus*; branch 1 in Fig. 1 *A* and *B*) (for both frogs and lizards). There were continued selective changes, with positive selection in frog genes responsible for regulation of carbohydrate derivative biosynthesis and transport (GO:1901137 and GO:1901264), lipid biosynthesis (GO:0008610), response to oxygen levels (GO:0070482), and cellular response to calcium ions (GO:0071277) (*SI Appendix, Tables S18 and S19*). The lizard PSGs at this stage mainly function as a response to radiation and DNA damage (GO:0010165, GO:0030330, and GO:0042770), as well as in multiple categories associated with DNA repair (GO:0006281, GO:0000726, and GO:0045002) (*SI Appendix, Tables S20 and S21*).

By examining the GO terms associated with PSGs from all of the three high-elevation lineages (branches colored green, blue, and red in Fig. 1), we observed several candidate HEA functions that appear to be evolving continuously with increasing elevational expansion in both groups, including functions associated with response to UV and respiratory electron transport chain in frogs (Table 1 and *SI Appendix, Table S22*) and functions associated with response to radiation and DNA repair in lizards (Table 1 and *SI Appendix, Table S23*).

One interesting finding is the continuous evolution of genes related to UV responses across increasing elevational distributions in frogs (GO:0009411; Table 1 and *SI Appendix, Table S22*). UV is a major environmental stress factor at high elevations, and UV exposure is especially problematic for exposed amphibian skin, where it can produce oxidant stress and DNA damage. Different genes related to responses to UV exposure appear to be under positive selection in different branches on the tree, suggesting a stepwise adaptation to UV exposure. Some of these genes showed the highest expression levels in the frog skin (the tissue with the highest level of UV exposure); however, the majority of the PSGs showed the highest expression levels in the

brain (*SI Appendix, Table S24*). Another interesting finding is the continuous evolution of the respiratory electron transport chain in frogs (GO:0022904 for ATP generation; *SI Appendix, Table S22*). Given that the high-elevation *Nanorana* species have slower development rates (4), both strategies, that of reducing energy consumption and that of increasing energy generation, probably shaped the metabolic adaptations of *Nanorana* to high elevations.

When we mapped the PSGs onto Kyoto Encyclopedia of Genes and Genomes (KEGG)-annotated pathways, we found additional evidence of continuous evolution of HEA in both groups. For example, at each elevational stage in the frogs, we found multiple PSGs that are involved in the oxidative phosphorylation pathway (with PSGs of *SDHA* and *LHPP* on the lineage from low elevation to ~2,000 m; *SDHA* and *NDUFA6* from ~2,000 m to ~3,000 m; and *NDUFB9*, *ATP6API*, *COX4II*, *ATP6V0A1*, and *ATP6VIH* from ~3,000 m to ~4,500 m), further supporting the role of ATP generation (energy metabolism) in HEA. We observed similar patterns in several other pathways, including nucleotide excision repair and the calcium signaling pathway (*SI Appendix, Table S25*). Although our analysis of the KEGG-annotated pathways in the lizards identified different pathways from those identified in the frogs, these analyses nonetheless supported the role of energy metabolism in HEA. Specifically, we found evidence of positive selection acting on the pathways of glycolysis, fructose and mannose metabolism, and galactose metabolism with increasing elevational stages of the lizards (detailed information is provided in *SI Appendix, Table S26*).

**Convergent Evolution of HEA Between Frogs and Lizards.** Although the two groups studied here (*Nanorana* and *Phrynocephalus*) are distantly related, they face similar environmental conditions and challenges at comparable elevations. To assess the degree of convergence in HEA between the two groups, we examined the GO terms associated with the PSGs from both groups at each elevational stage. This represents convergence at the functional level, rather than convergence of specific nucleotide substitutions in genes or amino acid replacements in proteins.

Our analyses identified 103, 35, and 35 GO terms associated with PSGs at elevational expansions to 2,000 m, 3,500 m, and 4,500 m, respectively (*SI Appendix, Tables S27–S29*). These included a number of functions that have evolved convergently in frogs and lizards (Table 2). For example, the genes *GTSE1* and *NBN* in the frogs and the genes *HIPK2* and *SMC1A* in the lizards are all associated with a response to DNA damage (GO:0042770) at the highest elevation species (branch 1; Table 2). Moreover, the

**Table 1. Number of PSGs associated with specific GO terms across all three elevational expansions in frogs and lizards**

Taxa	GO ID	P value	Term	Green*	Blue <sup>†</sup>	Red <sup>‡</sup>
Frogs	GO:0005996	0.038	Monosaccharide metabolic process	4	2	5
	GO:0009411	0.039	Response to UV	1	4	4
	GO:0046165	0.003	Alcohol biosynthetic process	2	3	5
	GO:0043255	0.006	Regulation of carbohydrate biosynthetic process	5	1	3
	GO:0022904	0.037	Respiratory electron transport chain	2	1	2
Lizards	GO:0006302	0.001	Double-strand break repair	4	1	4
	GO:0044262	0.001	Cellular carbohydrate metabolic process	2	2	4
	GO:0010212	0.021	Response to ionizing radiation	2	1	2
	GO:0006073	0.002	Cellular glucan metabolic process	1	2	2
	GO:0000726	0.000	Nonrecombinational repair	3	1	2

ID, identifier.

\*Branches 7 and 3 on both phylogenies.

<sup>†</sup>Branches 6 and 2 on both phylogenies.

<sup>‡</sup>Branch 1 on both phylogenies.

**Table 2. Convergent GO terms between frogs and lizards at comparable elevations**

Branch	GO term	P value	GO term definition
1	GO:0006109	0.002529437	Regulation of carbohydrate metabolic process
	GO:0006953	0.006084761	Acute-phase response
	GO:0042770	0.025243823	Signal transduction in response to DNA damage
	GO:0032355	0.043205949	Response to estradiol
2 and 6	GO:0045216	0.044599456	Cell–cell junction organization
	GO:0034976	0.019758571	Response to endoplasmic reticulum stress
	GO:0006457	0.026225839	Protein folding
	GO:0043467	0.029021199	Regulation of generation of precursor metabolites and energy
	GO:0043547	0.029612275	Positive regulation of GTPase activity
3 and 7	GO:0036119	0.046455853	Response to platelet-derived growth factor
	GO:0051660	0.000308287	Establishment of centrosome localization
	GO:0006457	0.004496455	Protein folding
	GO:0043547	0.012808641	Positive regulation of GTPase activity
	GO:0072524	0.013543229	Pyridine-containing compound metabolic process
	GO:0030397	0.024392786	Membrane disassembly
	GO:0032846	0.034964282	Positive regulation of homeostatic process
	GO:0051591	0.034999104	Response to cAMP
	GO:0019318	0.03834842	Hexose metabolic process
	GO:0048872	0.044892596	Homeostasis of number of cells
1, 6, and 7	GO:0032355	0.000587452	Response to estradiol
	GO:0051660	0.000692194	Establishment of centrosome localization
	GO:0006096	0.002883309	Glycolytic process
	GO:0048520	0.013978057	Positive regulation of behavior
	GO:0043547	0.024198176	Positive regulation of GTPase activity
	GO:0001666	0.030533707	Response to hypoxia
	GO:2000352	0.035373391	Negative regulation of endothelial cell apoptotic process

gene *PPP1R3B* in the frogs and the genes *SIRT6*, *PP1F*, and *ENSACAG00000024726* in the lizards are all associated with regulation of generation of precursor metabolites and energy (GO:0043467) at the middle-high elevational lineages (branches 6 and 2; Table 2). When we considered the entire pathway leading to the species with the highest elevational distributions (branches 7, 6, and 1 in Fig. 1 *A* and *B*, leading to *N. parkeri* in the frogs and *P. erythrus* in the lizards), the PSGs from both groups were associated with many terms related to hypoxia adaptation, such as the glycolytic process, which is a primary energy pathway during hypoxia (Table 2 and *SI Appendix*, Table S30).

If all elevational stages are considered, there is additional functional convergence evident in the GO terms associated with PSGs of the frogs and lizards, such as terms associated with energy metabolism, response to UV radiation, and DNA repair (*SI Appendix*, Tables S22, S23, S25, and S26). However, there are some differences in the DNA repair-related pathways between the two groups. In the frogs, changes in DNA repair functions were evident mainly in the earliest elevational shifts (up to 2,000 m); these mainly included changes to postreplication repair. In contrast, among the lizards, several different types of repair functions exhibited continual evolution along the elevational gradient, including double-strand break repair and nonrecombinational repair.

Although there are a large number of functions showing convergent evolution in the high-elevation frogs and lizards, many of these functions are associated with different sets of PSGs. However, there are 32 PSGs in common between the frogs and lizards (Fig. 1). Among these overlapping PSGs, *XPC* is involved in the damage recognition of nucleotide excision repair and plays an important role in the early steps of global genome nucleotide excision repair, especially in damage recognition, open complex formation, and repair protein complex formation (34). Other important functions associated with the overlapping PSGs include peroxisomal transport, lipid metabolism, and glycerophospholipid biosynthesis (Fig. 1C).

KEGG pathway analysis further highlights the roles of energy metabolism and DNA repair in HEA. The related energy metabolism pathways associated with the PSGs of both the frogs and lizards include pathways connected to glycolysis, fructose and mannose metabolism, and oxidative phosphorylation (*SI Appendix*, Table S31). The pathways associated with DNA repair in both groups are connected to nucleotide excision repair, homologous recombination, and nonhomologous end-joining (*SI Appendix*, Table S31). Moreover, categories related to both energy metabolism and DNA repair had a higher proportion of PSGs and more genetic changes than even the typically fast-evolving spermatogenesis and immune systems (*SI Appendix*, Fig. S2).

#### **Convergent Functional Adaptation to Hypoxia Involves Different Sets of Genes.**

Because many PSGs appeared to be associated with adaptation to hypoxia in both frogs and lizards, we further examined this convergent functional adaptation by comparing the frog and lizard branch-specific PSGs with the genes showing significant expression alterations under hypoxia (35). We expected strong changes in expression levels under hypoxia to be a strong indication that the genes are involved in the core hypoxia-responsive pathways (36). Accordingly, we examined all genes from HypoxiaDB (a highly curated database with more than 72,000 entries) that show expression changes under hypoxia (35) and constructed a “reference” list of hypoxia-responsive genes with a series of filtering analyses (*Methods*). Although the reference genes are from studies of humans, given the functional conservation of many genes across tetrapods, we expected the comparisons to provide useful information about genes involved in hypoxic responses in frogs and lizards. The final reference list of 1,191 genes contained 317 associated GO terms, which cover a wide range of functions, including immunity, neurodevelopment, energetic metabolism, and related aspects (*SI Appendix*, Table S32). We show the association of

hypoxia-associated GO terms with PSGs and particular shifts in elevational distribution on the frog and lizard trees in Fig. 2.

We next examined the GO terms between the frog and lizard PSGs and the reference list of hypoxia-responsive genes, and conducted statistical analyses to test whether the frogs and lizards showed more selection among hypoxia-responsive categories than expected at random. We identified 22 hypoxia-responsive categories along the frog branches 7 and 3, a number significantly higher than the mean expected value that was obtained from 1,000 randomly selected gene sets of the same size (22 versus 7.8;  $P < 2.2e-16$ ), strongly suggesting the emergence of hypoxia adaptation at an elevation of about 2,000 m. However, the lizard PSGs at this stage did not show significantly higher similarity with the hypoxia-associated reference list compared with the mean expected value (7 versus 6.8;  $P = 0.142$ ). At the elevational expansion to 3,500 m, both groups showed significantly higher similarity with the reference list than the mean expected value (frogs: 10 versus 6.7 with  $P < 2.2e-16$ ; lizards: 9 versus 5.7 with  $P < 2.2e-16$ ). The same pattern was observed at the highest elevational expansion (frogs: 9 versus 7.2 with  $P < 2.2e-16$ ; lizards: 8 versus 5.6 with  $P < 2.2e-16$ ). Therefore, it appears that much of the adaptation to hypoxia may occur at somewhat higher elevations in lizards than in frogs. Of course, some of the evolution of functions shown in Fig. 2 may also be related to factors other than hypoxia. There are many other ecological and physiological factors that are different among the

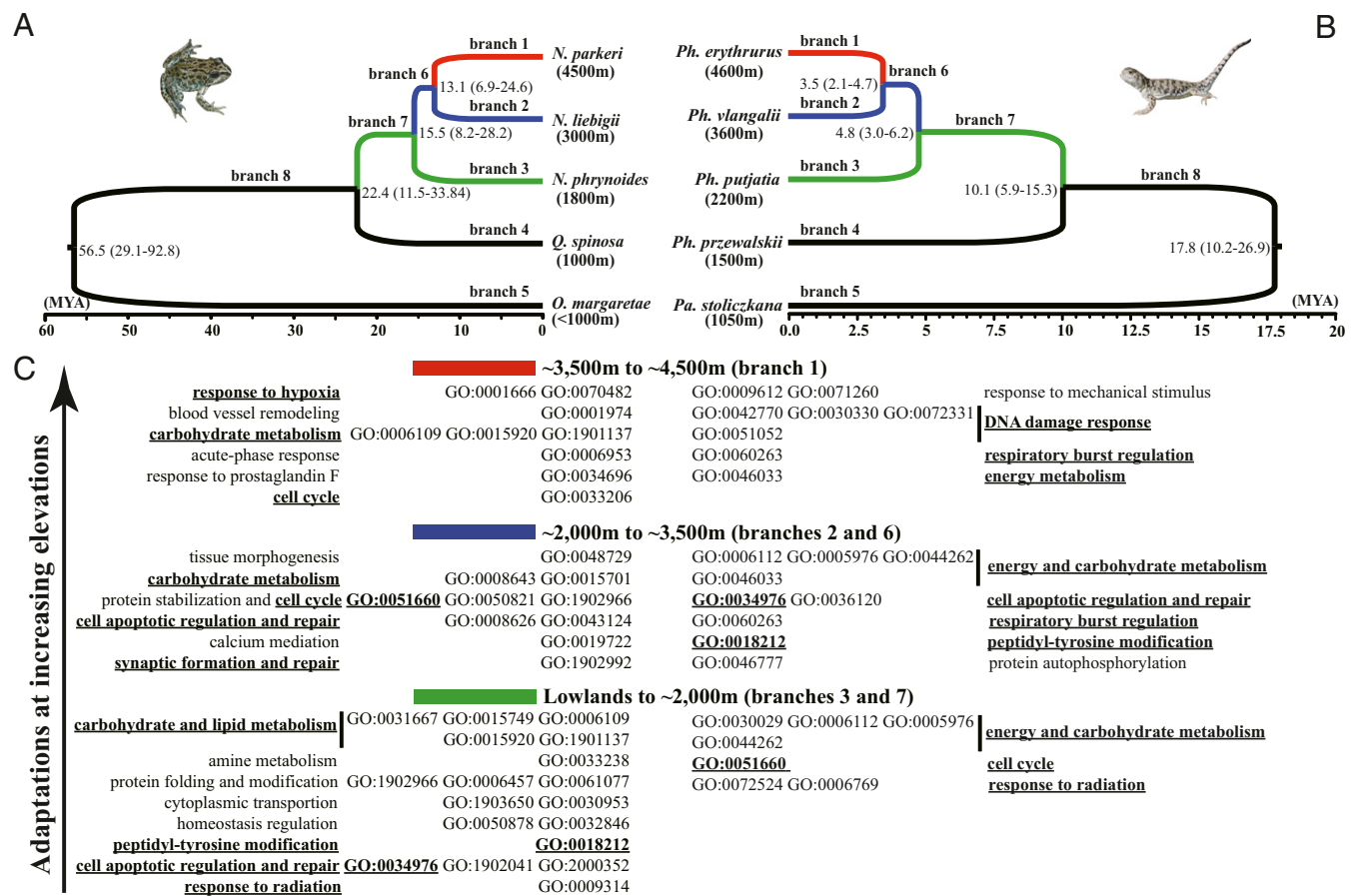
lizards, including reproductive mode (i.e., oviparity versus viviparity).

Although many functions associated with adaptation to hypoxia evolved in parallel in frogs and lizards (Fig. 2C), the PSGs associated with these functions do not contain any overlap (Table 3). Therefore, there appears to be considerable parallel adaptation to hypoxia, but involving completely different sets of genes.

## Methods

**Transcriptome Sequencing and Assembly.** The project approval (BBCJ-2014-001) was issued by Kunming Institute of Zoology (KIZ). Sample collections followed the animal use protocols approved by the Animal Care and Ethics Committee of KIZ. Total RNA for each sample was extracted using the SV Total RNA Isolation System (Promega). At least three tissues among the brain, heart, liver, skin, ovary, and testes were sampled from each species of *Nanorana* to allow the isolation of as many expressed genes as possible. All samples were immediately placed in LN<sub>2</sub> until RNA extraction. For each tissue, a single paired-end cDNA library was constructed for poly (A)<sup>+</sup> RNAs as described in the Illumina manual and sequenced using the HiSeq 2000 sequencing platform with a read length of 101 bp. Each tissue was sequenced with a high depth in the sequencing step. All of the sequencing information for the samples is summarized in *SI Appendix, Table S1*.

Raw sequencing reads obtained from each tissue were first cleaned with Cutadapt version 1.1.2 (37) to remove potential adapter sequences and low-quality regions with options “-q 20.” For each species, all cleaned reads from the different tissues were combined together and used as a single input for transcript assembly. De novo assembly was generated by Trinity version 20140717 (38) with its default parameters. CD-HIT version 4.6 (39) was then



**Fig. 2.** Convergent evolution of functions associated with adaptation to hypoxia in high-elevation frogs (A) and lizards (B). The numbers at each node show the estimated time of divergence for the corresponding split in the tree, with 95% confidence limits for the split shown in parentheses. (C) Example GO terms associated with each elevational shift that include functions that have been associated with adaptation to hypoxia. Functions that evolved in parallel in high-elevation lineages of both frogs and lizards are shown in boldface and are underlined.

**Table 3. Branch-associated PSGs related to adaptations to hypoxia (Fig. 2) in high-elevation frogs and lizards**

Branch	Elevational shift	Frog PSGs	Lizard PSGs
1	~3,500 to ~4,500 m	<i>FMN2, KIAA0196, BMPR2, TYW5, B4GALT7, B3GAT2, IMPDH1, ADCY4, RRM2B, GMDS, GPAA1, ENSACAG00000002886, AP2A1, PPP4R3A, ADIPOR1, GPER1, PLCD1, SCARB1, ALDOA, HIPK2, F7, POSTN, ADAM17, FN1, ITIH4, FLNA</i>	<i>BAG3, RAC1, ENSACAG000000008899, ZRANB3, DSCC1, NBN, GTSE1</i>
2 and 6	~2,000 to ~3,500 m	<i>EPHA2, PTK7, BMP4, RXRA, PBX1, COL3A1, STARD13, EZR, SLC26A4, CA12, SPHK1, PTGFR, PPP3CB, SLC37A3, SLC37A2, SLC35F6, UBE4B, IPO9, FLOT2, CRTAP, CEP83, NLRX1, ENSACAG000000025660</i>	<i>ADK, PPP1R1A, PPP1R3B, JAK2, FYN, CDKL5, AARS, ATF6, RAC1</i>
3 and 7	Lowlands to ~2,000 m	<i>XRCC6, FBXL3, PRKAA1, CLK2, ENSACAG000000013018, COL3A1, MAPK8, MEN1, CCKAR, FGG, TNFAIP3, STX4, SCARB1, ENSACAG00000002091, ENSACAG00000002125, ENSACAG00000009367, SERPINC1, F5, MBD1, ENSACAG000000015535, ENSACAG000000015364, FKBP4, HSPH1, ENSACAG000000017064, HSPG2, COASY, CANT1, MAT2B, TYW5, ALG3, UPP2, FGF1, EZR, CEP120, PPIL2, PDIA4, PRDX4, CCT4, TBCE, PDIA3, PPP1R15B, PLA2G6, FLOT1, GATA1, THY1, GAS6, KEAP1, TPR, SEMA4D, CTNND1, VEGFA, NOC2L, PLS1</i>	<i>APOA1BP, PGM2, IMPA1, PPP1R2, PAFAH1B1, NPHS2, MYL6, FRMD1</i>

applied on the final assemblies to eliminate redundant contigs, which helped in subsequent homolog prediction. Transrate (40) was used to evaluate the qualities of the de novo transcriptome assemblies based on common statistical indexes, such as average contig length, contig N50 values, and read mapping ratios. The same filtering and assembling methods were used for all samples.

**Orthology Determination and Alignment Construction.** We applied a modified reciprocal best hits method (“crb-blast”) to improve the accuracy of ortholog assignments between each assembly and a reference gene dataset (41). The reference dataset was the total protein sequences of *Anolis carolinensis*, which is available at the ENSEMBL genome database ([www.ensembl.org](http://www.ensembl.org)). Before performing BLAST analyses, the coding regions of each assembled contig were determined using TransDecoder version 3.01 (42). This program first identified ORFs that were at least 50-aa long (using TransDecoder. LongOrfs), and then verified those peptides with BLAST or domain hits to the public protein databases (using TransDecoder.Predict). CD-HIT (39) was used again to remove redundant protein sequences, especially when there were more than one ORF predicted for the transcript. The final consensus 1:1 orthologous gene set was then obtained based on the ortholog predictions generated between the query species and the reference gene set.

For each gene, Prank (43) was first used to generate an alignment at the codon level (with options “-F -codon”). The codon alignments were then analyzed by FasParser (44) with the “Trim” function to strap low-quality aligned columns. Final alignments with a length of less than 120 bp were discarded to reduce false-positive findings in subsequent evolutionary analyses. To evaluate the quality of our set of predicted orthologous genes, we estimated the mean identity ratio (MIR) for each alignment with FasParser. Briefly, for each alignment, we first calculated the identity ratio (IR) between each pair of sequences and then averaged the MIR by averaging all of the IR values. Generally, a high MIR was assumed to indicate a more accurate orthologous prediction due to the conservative nature of orthologous genes.

**Evolutionary Analyses.** Based on the corresponding topology of the frogs (20, 21) and lizards (45), codeml (version 4.9d) (46) was used to estimate dN, dS, and dN/dS with the free ratio model (“model = 1, NSsites = 0”) to obtain a general evolutionary pattern of selective pressure along the lineages. Mean dN/dS ratios were estimated by excluding genes with very small estimates of dS (<0.001, which would always result in a very large dN/dS). Genes showing evidence of positive selection along each branch were identified with the improved branch-site model (27), during which the targeted branch(es) were individually assigned as the foreground branch and others as background branches. Finally, a likelihood ratio test was used to compare a model (“model = 2, NSsites = 2, omega = 0.5|1.5, fix\_omega = 0”) of positive selection on the foreground branch with a null model (“model = 2, NSsites = 2, omega = 1, fix\_omega = 1”) where no positive selection occurred on the foreground branch and to calculate a *P* value.

The divergence time for each ancestral node on the phylogenetic tree was estimated using MCMCTree (46). All alignments were first concatenated to a supergene alignment by FasParser, and the ML tree was then constructed by

RAxML version 8.1.15 (47) with a GTR +  $\Gamma$  model of sequence evolution. For divergence time estimates, the calibration times were obtained from the TimeTree project ([www.timetree.org](http://www.timetree.org)) and/or previous studies (21, 45).

**Functional Annotation and Enrichment Analysis.** Functional annotations were obtained using the Trinotate pipeline ([trinotate.github.io](http://trinotate.github.io)). Briefly, we used several methods of functional annotation for each gene set as follows: BLAST+ (48) was used to search for homologies in SwissProt, HMMER (49) was used to search the PFAM protein domain database, SignalP (50) was used to predict protein signal peptides, and tmHMM (51) was used to identify transmembrane domains. All search hits were used for comparison with curated annotation databases (52, 53). Finally, all functional annotation data were then integrated into an SQLite database for fast, efficient searching for terms to create a whole annotation report for each transcriptome assembly.

GO enrichment analyses were conducted with Goseq (54). For determining the foreground branch-specific enrichments, we additionally applied enrichment analysis using the total PSGs from both foreground and background branches, and then chose the categories exclusively containing the foreground branch PSGs to represent the foreground-specific enrichments. To avoid noise in the enrichments, categories containing more than 500 entries (which are likely to represent very general biological functions) were discarded, after which all of the categories were filtered with REVIGO (55) to remove redundancy.

**Construction of the Reference Hypoxia-Responsive Pathways.** All genes showing expression shifts under hypoxia were collected from the HypoxiaDB database (35), as was their expression information. To gain a more precise reference dataset of hypoxia-responsive pathways, these genes were filtered according to the following steps: (i) For genes with quantitative measurements of expression change, there had to be at least a twofold change in expression in at least one study for retention, and (ii) for genes with no quantitative expression shift measurement, the description for the expression change had to include qualitative descriptions such as “dramatically” or “rapidly” affected. After this filtration, 713 strongly up-regulated genes and 478 strongly down-regulated genes were obtained from 3,500 differentially regulated genes.

With these retained genes, functional enrichments were constructed using the Goseq program (54). We discarded the categories with more than 500 entries to avoid nonspecific functions. To avoid nonmeaningful categories that might be introduced due to the pleiotropy of genes, we also deleted the categories that were also associated with the PSGs of the lowland species (*SI Appendix, Tables S33–S36*). Further, we removed the redundancy of these categories with the REVIGO method (55). Finally, only categories with a *P* value lower than 0.01 were retained.

**ACKNOWLEDGMENTS.** We thank Peng Shi for helpful discussions; Ke Jiang, Fang Yan, Hong-Man Chen, Jun-Xiao Yang, Xiao-Long Tu, Kai Wang, Bao-Lin Zhang, and Zi-Jie Zhang for great assistance in collections; and Lu Wang, Chun-Ling Zhu, and Li Zhong for help in RNA-Seq. This research was supported by the Strategic Priority Research Program (B) (Grant XDB13020200) of the

Chinese Academy of Sciences (CAS); National Natural Science Foundation of China Grants 91431105, 31671326, and 31871275; and the Animal Branch of the Germplasm Bank of Wild Species of the CAS (Large Research Infrastructure

Funding). J.C. and Y.-B.S. were supported by the Youth Innovation Promotion Association of the CAS. D.M.H. was funded by the CAS President's International Fellowship Initiative (Grant 2018VBA0039).

- Ruiz G, Rosenmann M, Veloso A (1983) Respiratory and hematological adaptations to high-altitude in Telmatobius frogs from the Chilean Andes. *Comp Biochem Physiol A* 76:109–113.
- Althuler DL, Dudley R (2006) The physiology and biomechanics of avian flight at high altitude. *Integr Comp Biol* 46:62–71.
- Navas CA, Chaui-Berlinck JG (2007) Respiratory physiology of high-altitude anurans: 55 years of research on altitude and oxygen. *Respir Physiol Neurobiol* 158:307–313.
- Ma X, Lu X, Merila J (2009) Altitudinal decline of body size in a Tibetan frog. *J Zool* 279:364–371.
- Lu S, et al. (2015) Differences in hematological traits between high- and low-altitude lizards (genus *Phrynocephalus*). *PLoS One* 10:e0125751.
- Storz JF, Scott GR, Cheviron ZA (2010) Phenotypic plasticity and genetic adaptation to high-altitude hypoxia in vertebrates. *J Exp Biol* 213:4125–4136.
- Norsang G, Kocbach L, Stammes J, Tsoja W (2011) Spatial distribution and temporal variation of solar UV radiation over the Tibetan Plateau. *Appl Phys Res* 3:37–46.
- Ge RL, et al. (2013) Draft genome sequence of the Tibetan antelope. *Nat Commun* 4: 1858.
- Qiu Q, et al. (2012) The yak genome and adaptation to life at high altitude. *Nat Genet* 44:946–949.
- Qu Y, et al. (2013) Ground tit genome reveals avian adaptation to living at high altitudes in the Tibetan plateau. *Nat Commun* 4:2071.
- Yang L, Wang Y, Zhang Z, He S (2014) Comprehensive transcriptome analysis reveals accelerated genetic evolution in a Tibet fish, *Gymnoditychus pachycheilus*. *Genome Biol Evol* 7:251–261.
- Yang W, Qi Y, Bi K, Fu J (2012) Toward understanding the genetic basis of adaptation to high-elevation life in poikilothermic species: A comparative transcriptomic analysis of two ranid frogs, *Rana chensinensis* and *R. kukunoris*. *BMC Genomics* 13:588.
- Yang W, Qi Y, Fu J (2014) Exploring the genetic basis of adaptation to high elevations in reptiles: A comparative transcriptome analysis of two toad-headed agamas (genus *Phrynocephalus*). *PLoS One* 9:e112218.
- Yang Y, et al. (2015) Comparative transcriptomic analysis revealed adaptation mechanism of *Phrynocephalus erythrurus*, the highest altitude lizard living in the Qinghai-Tibet Plateau. *BMC Evol Biol* 15:101.
- Yu L, et al. (2016) Genomic analysis of snub-nosed monkeys (*Rhinopithecus*) identifies genes and processes related to high-altitude adaptation. *Nat Genet* 48:947–952.
- Zhang Z, et al. (2016) Convergent evolution of rumen microbiomes in high-altitude mammals. *Curr Biol* 26:1873–1879.
- Yeaman S, et al. (2016) Convergent local adaptation to climate in distantly related conifers. *Science* 353:1431–1433.
- Hu Y, et al. (2017) Comparative genomics reveals convergent evolution between the bamboo-eating giant and red pandas. *Proc Natl Acad Sci USA* 114:1081–1086.
- Footo AD, et al. (2015) Convergent evolution of the genomes of marine mammals. *Nat Genet* 47:272–275.
- Che J, et al. (2009) Phylogeny of the Asian spiny frog tribe Paini (Family Dicroglossidae) sensu Dubois. *Mol Phylogenet Evol* 50:59–73.
- 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.
- Wilhelm BT, Landry JR (2009) RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. *Methods* 48:249–257.
- Gerstein MB, et al. (2014) Comparative analysis of the transcriptome across distant species. *Nature* 512:445–448.
- Qiao L, Yang W, Fu J, Song Z (2013) Transcriptome profile of the green odoriferous frog (*Odorranana margaretae*). *PLoS One* 8:e75211.
- Yang Z, Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol* 19:908–917.
- Sun YB, et al. (2015) Whole-genome sequence of the Tibetan frog *Nanorana parkeri* and the comparative evolution of tetrapod genomes. *Proc Natl Acad Sci USA* 112: E1257–E1262.
- Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol* 22: 2472–2479.
- Dos Reis M, Yang Z (2013) Why do more divergent sequences produce smaller non-synonymous/synonymous rate ratios in pairwise sequence comparisons? *Genetics* 195: 195–204.
- 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.
- Wang Y, Ohh M (2010) Oxygen-mediated endocytosis in cancer. *J Cell Mol Med* 14: 496–503.
- Davis JN, Carlsson A (1973) The effect of hypoxia on monoamine synthesis, levels and metabolism in rat brain. *J Neurochem* 21:783–790.
- Eldridge FL, Millhorn DE (1981) Central regulation of respiration by endogenous neurotransmitters and neuromodulators. *Annu Rev Physiol* 43:121–135.
- McNamara MC, Gingras-Leatherman JL, Lawson EE (1986) Effect of hypoxia on brainstem concentration of biogenic amines in postnatal rabbits. *Brain Res* 390: 253–258.
- Sugasawa K, et al. (1998) Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. *Mol Cell* 2:223–232.
- Khurana P, Sugadev R, Jain J, Singh SB (2013) HypoxiaDB: A database of hypoxia-regulated proteins. *Database (Oxford)* 2013:bat074.
- Khurana P, Tiwari D, Sugadev R, Sarkar S, Singh SB (2016) A comprehensive assessment of networks and pathways of hypoxia-associated proteins and identification of responsive protein modules. *Netw Model Anal Health Inform Bioinform* 5:1–13.
- Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17:10–12.
- Grabherr MG, et al. (2011) Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat Biotechnol* 29:644–652.
- Fu L, Niu B, Zhu Z, Wu S, Li W (2012) CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* 28:3150–3152.
- Smith-Unna R, Boursnell C, Patro R, Hibberd JM, Kelly S (2016) TransRate: Reference-free quality assessment of de novo transcriptome assemblies. *Genome Res* 26: 1134–1144.
- Aubry S, Kelly S, Kumpers BM, Smith-Unna RD, Hibberd JM (2014) Deep evolutionary comparison of gene expression identifies parallel recruitment of trans-factors in two independent origins of C4 photosynthesis. *PLoS Genet* 10:e1004365.
- Haas BJ, et al. (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8:1494–1512.
- Löytynoja A, Goldman N (2005) An algorithm for progressive multiple alignment of sequences with insertions. *Proc Natl Acad Sci USA* 102:10557–10562.
- Sun YB (2017) FasParser: A package for manipulating sequence data. *Zool Res* 38: 110–112.
- 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.
- Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24:1586–1591.
- Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
- Camacho C, et al. (2009) BLAST+: Architecture and applications. *BMC Bioinformatics* 10:421.
- Finn RD, Clements J, Eddy SR (2011) HMMER web server: Interactive sequence similarity searching. *Nucleic Acids Res* 39:W29–W37.
- Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat Methods* 8:785–786.
- Krogh A, Larsson B, von Heijne G, Sonnhammer EL (2001) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J Mol Biol* 305:567–580.
- Powell S, et al. (2012) eggNOG v3.0: Orthologous groups covering 1133 organisms at 41 different taxonomic ranges. *Nucleic Acids Res* 40:D284–D289.
- Ashburner M, et al.; The Gene Ontology Consortium (2000) Gene ontology: Tool for the unification of biology. *Nat Genet* 25:25–29.
- Young MD, Wakefield MJ, Smyth GK, Oshlack A (2010) Gene ontology analysis for RNA-seq: Accounting for selection bias. *Genome Biol* 11:R14.
- Supek F, Bošnjak M, Škunca N, Šmuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6:e21800.