Zoological Journal of the Linnean Society, 2023, **XX**, 1–16. With 3 figures.

Endemic lineages of spiny frogs demonstrate the biogeographic importance and conservational needs of the Hindu Kush–Himalaya region

SYLVIA HOFMANN[1,2,*], JOACHIM SCHMIDT[3], RAFAQAT MASROOR[4], LEO J. BORKIN[5], SPARTAK LITVINTCHUK[6], DENNIS RÖDDER[1], VLADIMIR VERSHININ[7,8], and DANIEL JABLONSKI[9]

[1]*LIB – Leibniz Institute for the Analysis of Biodiversity Change, Museum Koenig Bonn, Adenauerallee 160, 53113 Bonn, Germany*
[2]*UFZ – Helmholtz Centre for Environmental Research, Department of Conservation Biology, Permoserstrasse 15, 04318 Leipzig, Germany*
[3]*General and Systematic Zoology, Institute of Biosciences, University of Rostock, 18055 Rostock, Germany*
[4]*Pakistan Museum of Natural History, Garden Avenue, Shakarparian, Islamabad 44000, Pakistan*
[5]*Department of Herpetology, Zoological Institute, Russian Academy of Sciences, Universitetskaya nab. 1, 199034 St. Petersburg, Russia*
[6]*Institute of Cytology, Russian Academy of Sciences, Tikhoretsky pr. 4, 194064 St. Petersburg, Russia*
[7]*Institute of Plant and Animal Ecology, Russian Academy of Sciences, Ural Division, 620144 Ekaterinburg, Russia*
[8]*Ural Federal University, ul. Mira 19, Yekaterinburg, 620002 Russia*
[9]*Department of Zoology, Comenius University in Bratislava, Ilkovičova 6, Mlynská dolina, 84215 Bratislava, Slovakia*

Received 2 August 2022; revised 26 December 2022; accepted for publication 27 December 2022

The relict, endemic taxa *Allopaa* and *Chrysopaa* are key elements of the Hindu Kush–Himalayan amphibian fauna and potentially share a similar biogeographic evolution, making them important proxies for the reconstruction of the palaeoenvironmental and palaeotopographic history of the Himalaya–Tibet–Orogen. However, little is known about the taxonomy, phylogeography, genetic diversity and distribution of these taxa. We here provide new molecular data on Himalayan spiny frogs and species distribution models (SDMs) for *A. hazarensis* and *C. sternosignata*. The results reveal a better resolved phylogeny of these frogs compared to previous trees and strongly support the placement of *A. hazarensis* in the genus *Nanorana*. We further identify a so far unknown clade from the western Himalayas in *Nanorana*, apart from the subgroups *Chaparana*, *Paa* and the nominal *Nanorana*. In *A. hazarensis*, genetic diversity is relatively low. The results strengthen support for the recently proposed out-of-Tibet-into-the-Himalayan-exile hypothesis and a trans-Tibet dispersal of ancestral spiny frogs during the Palaeogene. Moreover, SDMs provide the first detailed distribution maps of *A. hazarensis* and *C. sternosignata* and strong evidence for distinct niche divergence among the two taxa. Our findings contribute to the knowledge about the distribution of these species and provide basic information for guiding future conservation management of them.

KEYWORDS: *Allopaa* – *Chrysopaa* – *Nanorana* – niche divergence – phylogeny.

INTRODUCTION

Understanding the palaeoenvironmental and topographic history of the northwestern Himalaya, also known as Kashmir or Indus Himalaya, and the adjacent Karakoram and Hindu Kush Mountain ranges is crucial for the reconstruction of the biogeographic history of species groups endemic to these parts of the Himalaya–Tibet–Orogen (HTO) (Schmidt *et al.*, 2012; Hofmann *et al.*, 2017; Jablonski *et al.*, 2021). For different groups of organisms, e.g. ground beetles, lazy toads and spiny frogs, the presence of

*Corresponding author. E-mail: s.hofmann@leibniz-lib.de

ancestral lineages in the northwestern Himalaya has been shown (Schmidt *et al.*, 2012; Hofmann *et al.*, 2017, 2019, 2021a). These lineages are most closely related to representatives in the eastern Himalaya or eastern central Himalaya, but they have no relatives across wide regions of the West and Central Himalaya. Such a paradoxical distributional pattern has been explained by a palaeo-Tibetan origin of Himalayan lineages and/or a trans-Tibet dispersal of ancestral lineages across central Tibet during the Late Palaeogene or Early Neogene when the Plateau was at significant lower elevation (Schmidt *et al.*, 2012; Hofmann *et al.*, 2017, 2019, 2021a).

In this context, the phylogenetic placement of the two westernmost dicroglossid frogs that occur in the HTO, *Allopaa hazarensis* (Dubois & Khan, 1979) from the Kashmir Himalaya and *Chrysopaa sternosignata* (Murray, 1885) from the Hindu Kush, have recently been addressed for the first time (Hofmann *et al.*, 2021a). While the basal position of the monotypic *Chrysopaa* Ohler & Dubois, 2006 relative to *Nanorana* Günther, 1896 and *Allopaa* Ohler & Dubois, 2006 was strongly supported, the placement of *A. hazarensis* from Pakistan in the genus *Nanorana* seemed reasonable, but remained less supported, presumably either due to a gap in taxon sampling and/or because it represents a lineage that diverged in an early stage of the evolutionary history of *Nanorana*. Both taxa, *Chrysopaa* and *Allopaa*, are endemic to the Hindu Kush–Himalayan area and were previously considered a single genus (*Paa*) (Khan, 2006). To the best of our knowledge, sympatric or syntopic records do not exist, although they might be possible and misidentification of the two frogs cannot be excluded. The two taxa are key elements of the regional thermophile fauna and potentially share a similar biogeographic origin, making them important proxies for the reconstruction of the palaeoenvironmental and palaeoelevational evolution of the Tibetan Plateau (Hofmann *et al.*, 2021a) and the evaluation of respective modern scenarios of the development of the HTO (Xiong *et al.*, 2022). Apart from molecular data, information about the current distribution and ecological niches of these relict taxa are crucial to assess their biogeographic history in the context of the evolution of the HTO.

Allopaa hazarensis can be found in boulder-rich streams (Dubois, 1975) or clear pools with flowing water (Khan *et al.*, 2008), often surrounded by subtropical or warm-temperate forest at elevations above 1000 m a.s.l. (own observation; Ahmed *et al.*, 2020). The species was described from near Datta (~34.30°N, 73.26°E), northern Pakistan (Manshera District, Hazera Division, about 1200 m a.s.l.; for molecular data from the type locality, see: Hofmann *et al.*, 2021a) and is known to occur in the Khyber Pakhtunkhwa Province of Pakistan, and in adjacent Kashmir, India (see: Frost, 2022). Apart from its still not fully resolved taxonomic position, little is known about the life history, ecology and distribution of *A. hazarensis*, and far less data are available about its genetic diversity across the known distribution range of the species.

On the other hand, *Chrysopaa sternosignata* has been originally described from Sindh Province in Pakistan (Murray, 1885). The syntypes, presumably at the Karachi Museum, are apparently lost (Frost, 2022); two further syntypes are noted in the BMNH (1947.2.1.21–22; Boulenger, 1920) with the type localities 'Mulleer (= Malir) near Kurrachee (= Karachi); Zandra and Quetta, in South Afghanistan' (now all localities in Pakistan). However, we consider the occurrence of the species at Malir as unlikely (own data; see also: Mertens, 1969) since the environmental conditions in the south-eastern edge of Pakistan (Sindh Province) are remarkably different compared to those of places where the species has its major distribution range (Balochistan, Afghanistan; Supporting Information, Fig. S1). *Chrysopaa* seems to be exclusively aquatic and has been reported to be common in pools and water-channels in the Quetta and Pishin districts of Balochistan Province in Pakistan from 1500 up to 1800 m a.s.l. (Boulenger, 1920), as well as in other areas of southern, western and central Afghanistan (Wagner *et al.*, 2016). Noteworthy, the species has never been recorded in Iran. It occupies predominantly regions in the colline zone with warm-temperate, arid climate. Similar to *A. hazaranesis*, there are virtually no studies on the ecology of the species and only a few molecular datasets for *C. sternosignata* are available. However, such information is essential not only for species conservation, but also for biogeography and molecular taxonomy.

Therefore, we aim (1) to place new samples of both species and additional *Nanorana* specimens onto the phylogeny of closely related spiny frogs in order to enhance phylogenetic resolution and to assess previous biogeographic hypotheses, (2) to provide information on the genetic diversity of *Allopaa hazarensis* and (3) to model the habitat suitability and potential distribution of the two taxa, which is vital for understanding the historical processes that shaped the disjunction of species and are probably important in guiding conservation planning. Our data represent the most detailed evaluation of molecular and occurrence data of these enigmatic frogs. The results encourage discussion on our understanding of their biogeographical history and might be important for conservation genetics.

MATERIAL AND METHODS

SAMPLING, LABORATORY PROTOCOLS AND DATA ACQUISITION

We used sequence data of the *16S* ribosomal RNA (rRNA), mitochondrial cytochrome c oxidase I (*COI*)

and nuclear recombination activating gene 1 (*Rag1*) region of spiny frogs compiled in our previous studies (Hofmann *et al.*, 2019, 2021a). We complemented these data with newly generated sequences from 22 specimens of *Allopaa hazarensis*, two *Chrysopaa* and four *Nanorana* museum samples from the West Himalaya (for details see Supporting Information, Table S1). The additional *Nanorana* samples could not be identified at species level; they represent taxa apparently adapted to warm-temperate climatic conditions, because the specimens were collected between 1400 and 2100 m a.s.l., in the lower cloud forest zone of the West Himalaya (Miehe, 1991). We further considered sequences of *A. hazarensis* and *Nanorana* taxa that have recently been uploaded to GenBank (mainly *16S*). Sampling was performed under the permit of the Pakistan Museum of Natural History, Islamabad, Pakistan (No. PMNH/EST-1[89]/05), according to the regulations for the protection of terrestrial wild animals. We followed the laboratory procedure using primers and polymerase chain reaction (PCR) conditions, as previously described (Hofmann *et al.*, 2019). Briefly, DNA was isolated from ethanol tissues using the DNeasy Blood & Tissue Kit (Qiagen, Venlo, Netherlands). Approximately 570 bp of the 16S, 539 bp of the *COI* and 1207 bp of the *Rag1* gene were amplified; PCR products were purified using the ExoSAP-IT enzymatic clean-up (USB Europe GmbH, Staufen, Germany) and the mi-PCR Purification Kit (Metabion, Planegg, Germany) and sequenced by Macrogen Inc. (Amsterdam, The Netherlands).

Sequence alignment and phylogenetic reconstruction

The rRNA *16S* sequences were aligned based on their secondary structures using RNAsalsa v.0.8.1 (Stocsits *et al.*, 2009) and the ribosomal structure model of *Bos taurus* Linnaeus, 1758, provided with the RNAsalsa package. Protein-coding genes were aligned with MUSCLE (Edgar, 2004) using default settings in MEGA X (Kumar *et al.*, 2018). We found no ambiguities, such as deletions, insertions or stop codons, neither in the nucleotide-based nor amino acids alignments.

We performed phylogenetic analysis based on two datasets: (1) the *16S* sequence data, comprising a total of 88 samples and 572 nucleotide positions and (2) the concatenated mtDNA + nuDNA sequence alignment containing 2316 bp and 58 samples for which sequence data of at least two of the three loci were available. To reduce computational time, we considered only one sample of *Allopaa hazarensis* per locality in the concatenated sequence dataset (Supporting Information, Table S2).

We inferred a Bayesian inference (BI) tree for each of the two datasets using MrBayes v.3.2.6 (Ronquist *et al.*, 2012) considering stem and loop regions (*16S*) as partitions, as well as genes and codons. To provide secondary structure information for MrBayes we converted the RNAsalsa consensus structure output (a dot-bracket structure string) into a list of paired and unpaired positions and implemented them in the MrBayes input file. We assigned the doublet model (16 × 16; Schoniger & von Haeseler, 1999) to the rRNA stem pairs, and the standard 4 × 4 option with a Generalised time-reversible (GTR) evolutionary model to the remaining nucleotide positions. The site-specific rates were set variable. MrBayes was run for 10 million generations, sampling trees every 1000th generation. Inspection of the standard deviation of split frequencies, as well as an effective sample size value > 200 of the traces using TRACER v.1.7.1 (Rambaut *et al.*, 2018), indicated convergence of Markov chains. In all analyses, we used four parallel Markov chain Monte Carlo simulations with four chains and discarded the first 25% of the samples of each run as burn-in.

Based on the concatenated dataset we also inferred a maximum likelihood (ML) tree using RAxML-NG v.1.1.0 (Kozlov *et al.*, 2019). We used PartitionFinder v.1.1.1 (Lanfear *et al.*, 2012) to optimize the partition scheme with the following setting: branch lengths linked, corrected Akaike information criterion (AICc), greedy search algorithm, and the substitution models implemented in RAxML. RAxML-NG was then run with 20 random and 20 parsimony starting trees, 10 000 bootstrap replicates, and specifying the Felsenstein's bootstrap (FBP), as well as the recently introduced transfer bootstrap expectation (TBE; Lemoine *et al.*, 2018), as branch support metrics. Trees were visualized with FigTree v.1.4.3 (Drummond & Rambaut, 2007).

We also constructed minimum-spanning haplotype networks for *Allopaa hazarensis* for *16S* and *COI* using the software PopArt (http://popart.otago.ac.nz), and calculated genetic distances per locus between taxa with MEGA X. Sequences with > 25% missing data were excluded from haplotype networks. Nuclear heterozygote positions could not be phased because most populations were represented by only a few or single individuals which did not allow a robust statistical inference of haplotypes. Therefore, network reconstruction was not applied to *Rag1* sequence data.

Molecular dating

Using BEAST2 v.2.6.7 we reconstructed a dated phylogeny, based on the concatenated dataset, partitioned by genes and codons, and our previous calibration approach (Hofmann *et al.*, 2021a). Briefly, we imposed the following age constraints derived from fossil-calibrated divergence estimates (Hofmann *et al.*,

2019): most recent common ancestor (MRCA) of Paini 38.10 Mya, 28.70–47.50 (normal distribution, sigma: 4.80); split of Tibetan *Nanorana* and Himalayan *Paa* 12.59 Mya, 7.93–17.30 (normal distribution, sigma: 2.38); separation of the Plateau frog *Nanorana parkeri* (Stejneger, 1927) and *Nanorana ventripunctata* Fei & Huang, 1985 + *Nanorana pleskei* Günther, 1896 *c*. 6.35 Mya, 3.54–9.16 (normal distribution, sigma: 1.44). We performed three runs, each with a chain length of 50 million, a thinning range of 5000, a log-normal relaxed clock model, a birth–death tree prior, a random starting tree and the site models selected by the bModelTest package (Bouckaert & Drummond, 2017) implemented in BEAST2. Runs were then combined with BEAST2 LogCombiner v.2.6.7 by resampling trees from the posterior distributions at a lower frequency, resulting in 12 500 trees. Stationary levels and convergence of the runs were verified with TRACER based on the average standard deviation of split frequencies and ESS values > 200. The final tree was obtained with TreeAnnotator v.2.6.7 and visualized with FigTree.

Occurrence data and species distribution modelling (SDM) computation

We performed SDM to assess the geographical distribution of suitable climatic conditions for both taxa separately. A total of 38 individual records of *Allopaa hazarensis* and 43 of *Chrysopaa sternosignata* were available to us (Supporting Information, Table S2), including our own observations, data from literature (Ohler & Dubois, 2006; Akram *et al.*, 2022) and databases like VertNet (http://vertnet.org) and GBIF (https://gbif.org). Grids of 19 standard bioclimatic variables for the current climate (WorldClim v.2.1 climate data for 1970–2000) and elevation were downloaded at a resolution of 30 arc-seconds (~1 km grid cells at the equator) from http://www.worldclim.org (accessed on 24 July 2022; Fick & Hijmans, 2017). All layers were projected to WGS84 and clipped to the spatial modelling extent, covering the West Himalaya, Hindu Kush and adjacent regions. Prior to the modelling approach we carried out a principal components analysis (PCA) with the SDMtoolbox v.2.5 (Brown, 2014) available for ArcGIS v.10.8, reducing the clipped layers to three orthogonal principal components describing the majority (> 99%) of the variability in climate. We then used these three components to assess the climatic heterogeneity across the area of interest. To eliminate spatial clusters of species localities we filtered our presence data by Euclidian distances (min. 1 km, max. 5 km; three distance classes) according to climate heterogeneity, using the rarefying module in SDMtoolbox, resulting in the exclusion of 13 sites for the *A. hazarensis* and five for the *C. sternosignata* occurrence dataset. Because climatic variables are often highly correlated, we explored all climate variables and the elevation data for potential multicollinearity by calculating squared Pearson's correlation coefficients (r^2) using the respective python script in the SDMtoolbox and removed highly correlated variables. The remaining variables, namely BIO2 = mean diurnal temperature range, BIO6 = minimal temperature of coldest month, BIO8 = mean temperature of wettest quarter, BIO9 = mean temperature of driest quarter, BIO12 = annual precipitation and elevation had a mean intercorrelation of $r^2 = 0.27$ (range 0.012–0.811, SD = 0.295; Supporting Information, Table S3), and were than included in the model. We restricted background selection with a buffered minimum-convex polygons based on the known occurrences and using a buffer size of 300 km. Modelling was performed with MaxEnt v.3.4.3 (Phillips *et al.*, 2004, 2006), which implements the maximum entropy algorithm that is appropriate for analysing presence-only data and has proven high predictive accuracy compared with other modelling approaches (Elith *et al.*, 2006) and robustness for small sample sizes (Pearson *et al.*, 2006). We set different regularization multipliers (1; 1.5; 2; 2.5) to optimize model performance. Each model was generated using 80% of the species records for model training and 20% for model evaluation applying a bootstrap approach, specifying five spatial groups, and using the 'equal sensitivity and specificity threshold' as the minimum threshold above which the species is considered to be present. We also tested a minimum training presence (MTP) threshold to classify average continuous probabilities of the MaxEnt model into binary maps. The MTP value is frequently used in biomod2 approaches (see below); it represents the lowest value observed in the continuous prediction map at a presence location for a specific species.

Model performance and the importance of the environmental variables to the model were assessed using the mean area under the curve (AUC) of the receiver operating characteristics (ROC; Hanley & McNeil, 1982), which is a common measure of model accuracy (Swets, 1988), and jack-knife testing. Models with AUC values above 0.7 are considered potentially informative, good between 0.8 and 0.9, and excellent for AUC between 0.9 and 1 (Swets, 1988; Elith *et al.*, 2006; Préau *et al.*, 2018).

For reasons of comparison we also performed SDM modelling using an ensemble approach in biomod2 (Thuiller *et al.*, 2013, 2021) using the following algorithms: generalized linear models (GLM; McCullagh & Nelder, 1989), generalized additive models (GAM; Hastie & Tibishirani, 1990), generalized boosted models (GBM; Ridgeway, 1999), classification tree analysis (CTA, Breiman *et al.*, 1984) and artificial neural networks (ANN; Ripley, 1996). Model performance was internally tested in biomod2 via a ten-fold data-splitting approach (80% training/20% test)

and quantified using AUC (Hanley & McNeil, 1982), Cohen's Kappa (Monserud & Leemans, 1992) and the true skill statistic (TSS; Allouche *et al.*, 2006). The relative performance of the models was compared to derive consensus predictions based on all models with both AUC and TSS > 0.7, weighted by their proportional predictive performance on test data. The extrapolation areas were masked in the final maps and they were rescaled to 0–1, applying the MTP as threshold.

To test for environmental niche divergence among the two species we performed the niche identity test implemented in ENMTOOLS v.1.4.4 (Warren *et al.*, 2010), based on Schoener's *D* (Schoener, 1968), as recommended by Rödder & Engler (2011), and a measure derived from Hellinger distance called *I* (Warren *et al.*, 2008). These metrics measure the similarity of two distribution models as an indicator of niche overlap, ranging from zero (no overlap) to one (complete overlap). We tested against the null hypothesis of SDMs being identical by randomly generating a distribution of niche overlap values with unknown species identities to which the observed overlap of Schoener's *D* and *I* is compared (Miller & Franklin, 2002; McIntyre, 2012). The null hypothesis was rejected if the observed value of niche overlap between two taxa falls outside the 95% confidence limits ($P < 0.05$) of the simulated values. Statistical significance against the null hypothesis was determined by 100 pseudoreplicates, and the number of background points were set to 10 000. For niche quantification, only the MAXENT models were used.

RESULTS

PHYLOGENY AND GENETIC DIVERSITY

The BI and ML gene trees for the partitioned concatenated dataset (mtDNA + nuDNA) were well resolved with almost identical topologies, while in the *16S* tree most clades remained weakly supported (Fig. 1; Supporting Information, Fig. S2). However, consistent with previous findings, *Chrysopaa* was placed basally relative to the genus *Nanorana* and *Allopaa hazarensis*. All analysis recovered a placement of *A. hazarensis* within *Nanorana*, rendering that genus in the sense of Frost *et al.* (2022) paraphyletic. In fact, *A. hazarensis* formed a highly supported monophyletic clade that constitutes the sister-position to the *Chaparana* subgroup (concatenated analysis) or, based on *16S* and only weakly supported, to *Chaparana* + *Nanorana liebigii* (Günther, 1860) (*Paa*) + the clade with the unidentified specimens from the West Himalaya. The so far unidentified specimens from the West Himalaya formed a separate clade within *Nanorana*, beside *Chaparana*, *Paa* and the nominal *Nanorana* (Fig. 1). Noteworthy, species from the eastern Himalaya [*Nanorana chayuensis* (Ye, 1977), *Nanorana conaensis* (Fei & Huang, 1981), *Nanorana maculosa* (Liu *et al.*, 1960) and *Nanorana medogenesis* (Fei & Ye, 1999)] clustered together with lineages from the West Himalaya, except for the unidentified specimens, while all lineages from the central Himalaya formed an own distinct clade. Moreover, it appears that several undescribed species may exist within Himalayan *Paa*, and some taxa branched in multiple subclades, e.g. *N. liebigii*, *Nanorana vicina* (Stoliczka, 1872), indicating high intraspecific variation or potential cryptic diversity (Fig. 1). Also, our *16S* phylogeny (Supporting Information, Fig. S2), as well as the estimates of genetic distances (Supporting Information, Table S4), showed the taxonomical misclassification of the species *Nanorana arunachalensis* (Saikia, Sinha & Kharkongor, 2017) within *Nanorana*.

The average genetic distances between *Allopaa hazarensis* or *Chrysopaa sternosignata* and *Nanorana* species ranged between ~5–8% (*16S*), ~13–21% (*COI*) and ~2–4% (*Rag1*), except for *N. arunachalensis*, which was more than 14% (*16S*) distant from all taxa (Supporting Information, Tables S4, S5a, b). Within *A. hazarensis*, genetic distances were present but relatively low (Supporting Information, Table S6a–c), ranging up to 1.5% (*16S*), 0.8% (*COI*) and 0.3% (*Rag1*). Similarly, overall nucleotide variability was small. Haplotype networks for the three genes were geographically less structured, with eight (*16S*) and six (*COI*) haplotypes (Fig. 2). Populations east and west of the Indus did not share any *COI* haplotype. However, in *16S* the haplotype h7 was present on both sides of the Indus River (Fig. 2), indicating at least some exchange between localities. Most *COI*-haplotypes were divergent by one or two mutations, indicating a more recent origin and no major divergence among these sequences.

The four distinct phylogenetic groups (*Allopaa*, *Chaparana*, *Nanorana* and *Paa*) were close to each other genetically with genetic distances between them of similar magnitudes (14–17%, *COI*; 2.3–3.6% *Rag1*; Supporting Information, Table S7a, b). In contrast, the pairwise distances between the *Nanorana* clades (including *Allopaa*) and *Chrysopaa* were slightly higher in *COI*, ranging between 17 and 21%, but not in the nuclear *Rag1* (2.1–3.1%). Since main clades were less supported in the *16S* tree, we did not estimate genetic distances for this locus.

DIVERGENCE TIME

Separation of *Quasipaa* Dubois, 1992 (southern China, South-East Asia) and *Chrysopaa* from the other Paini lineages of the HTO occurred in the Mid-Oligocene between *c.* 28 Mya (20.7–35.5 Mya) and 30 Mya (22.2–37.6 Mya) (Supporting Information, Fig. S3). The warm-temperate sister-clades *Allopaa* (Kashmir Himalaya) and *Chaparana* (eastern margin of the

Figure 1. Bayesian inference (BI; left) and maximum likelihood tree (ML; right) based on concatenated mtDNA and nDNA sequence data (*16S* + *COI* + *Rag1*) of the tribe Paini. Numbers at branch nodes refer to posterior probabilities ≥ 0.9 (BI tree), as well as Felsenstein's bootstrap values ≥ 70% and transfer bootstrap expectation ≥ 0.9 (ML tree). Branches of *Allopaa hazarensis* are indicated red, while *Chrysopaa sternosignata* is highlighted green. Species names are followed by voucher number (if available). Coloured shaded boxes indicate subgroups of *Nanorana* and the new clade (in yellow) with so far unidentified specimens.

HTO) split from the Himalayan *Paa* and Tibetan *Nanorana* in the Early Miocene, around 19 Mya (14.1–23.7 Mya), and diverged only slightly later *c.* 16 Mya (11.4–20.4 Mya). A similar age (17 Mya; 13.9–21.5 Mya) was estimated for the newly discovered clade from the West Himalaya, while the ancestral Plateau lineage (*Nanorana*) appeared around 14 Mya (11.1–17.8 Mya), although that lineage diversified only in the Late Miocene and Pliocene (7–4 Mya). The clade comprising lineages from western and eastern Himalaya (*Paa*) split from the central Himalayan clade around 11 Mya (7.7–13.4 Mya). Diversification in both of these clades had then taken place continuously during the whole Late Cenozoic. Our estimated divergence times (Supporting Information, Fig. S3) were consistent with previous results (Hofmann *et al.*, 2019, 2021a).

Species distribution model

All records are located in montane regions between *c.* 700 and 2100 m a.s.l. (*Allopaa hazarensis*), and between *c.* 1000 and 3000 m a.s.l. (*Chrysopaa sternosignata*), respectively. These regions are characterized by a heterogeneous climate (Supporting Information, Fig. S4), particularly those along the Himalaya. The average performance of our MaxEnt models was considered significantly better than random (mean $AUC_{Allopaa}$ = 0.978; mean $AUC_{Chrysopaa}$ = 0.802; Supporting Information, Table S8). The highest

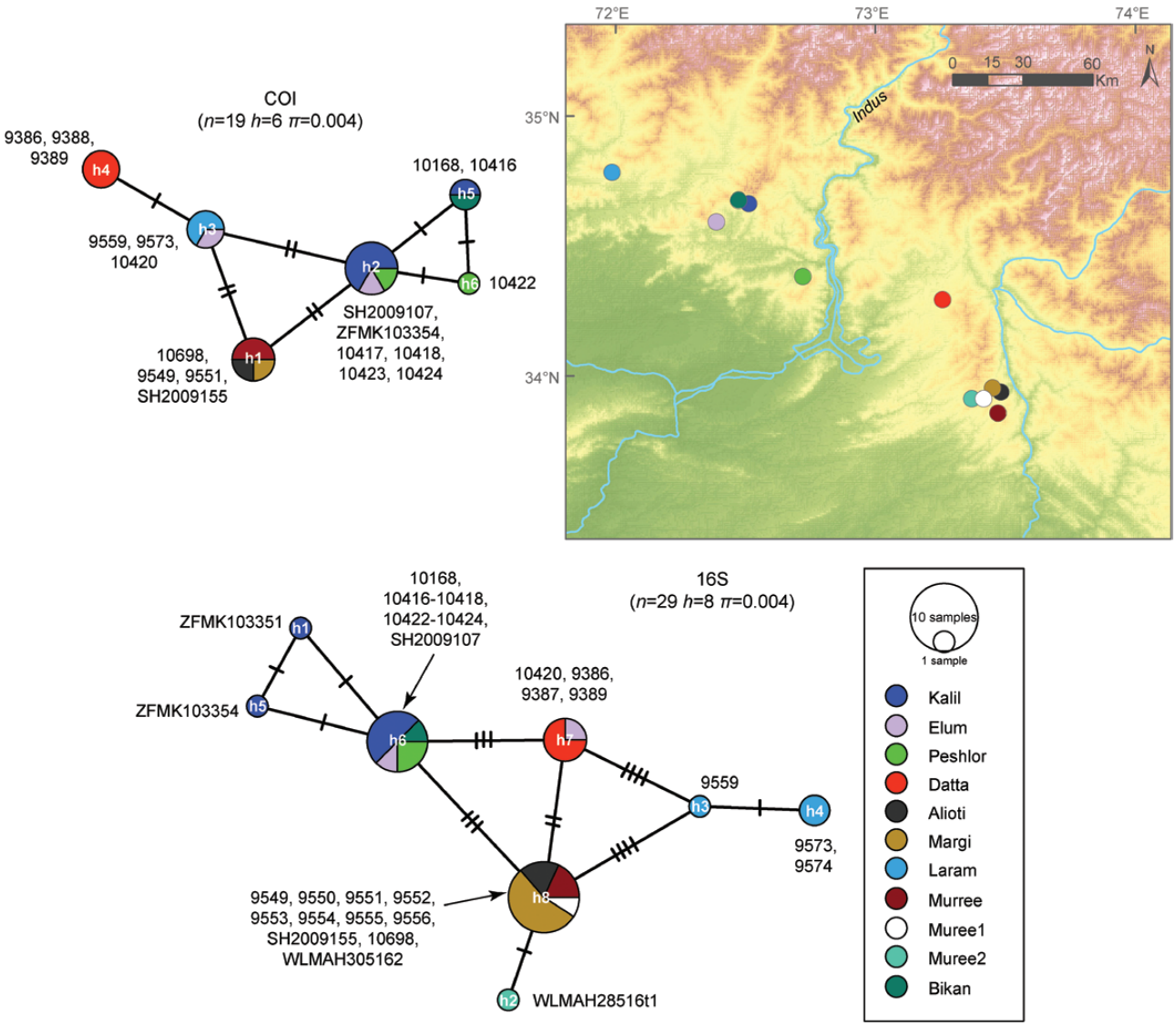

Figure 2. Minimum-spanning haplotype networks of *Allopaa hazarensis* generated for *16S* and *COI* sequence data with the number of used sequences, detected haplotypes, and the level of nucleotide variability. Symbol sizes reflect haplotype frequencies, and a small black line between two haplotypes corresponds to one mutation step. Sequence-IDs are indicated for each haplotype (h1–h8). Map shows the localities from where the respective haplotypes originate.

probability for the distribution of *A. hazarensis* is indicated for the southern foothills of the northwestern Himalaya across northern Pakistan, India and Nepal, with a warm-temperate climate (Fig. 3A), while *C. sternosignata* occurs in arid regions of the Hindu Kush and its southern and western extensions, e.g. the Paropamisus Mountains and Sulaiman mountain ranges (Fig. 3B). According to the MAXENT model, a distribution of *A. hazarensis* in the southwestern area of the Hindu Kush is also plausible, e.g. along the Spīn Ghar mountain range. The potential distribution range of *Chrysopaa* connects to the Iranian plateau and Zagros Mountains, where the species may find favourable conditions (note: the species has never been recorded from there). When applying the MTP threshold to the MAXENT results, the potential distribution range of *Chrysopaa* also encompasses parts of the Karakoram Range, and the Tibetan Plateau (Supporting Information, Fig. S5).

The performance of the biomod2 ensembles was likewise excellent, showing high-performance values, as well as sensitivity and specificity ($AUC_{Allopaa}$ = 0.998; mean $AUC_{Chrysopaa}$ = 0.977; Supporting Information, Tables S9, S10). Compared to the MAXENT models, the ensembles predicted a basically similar potential distribution for *C. sternosignata* and a more restricted range

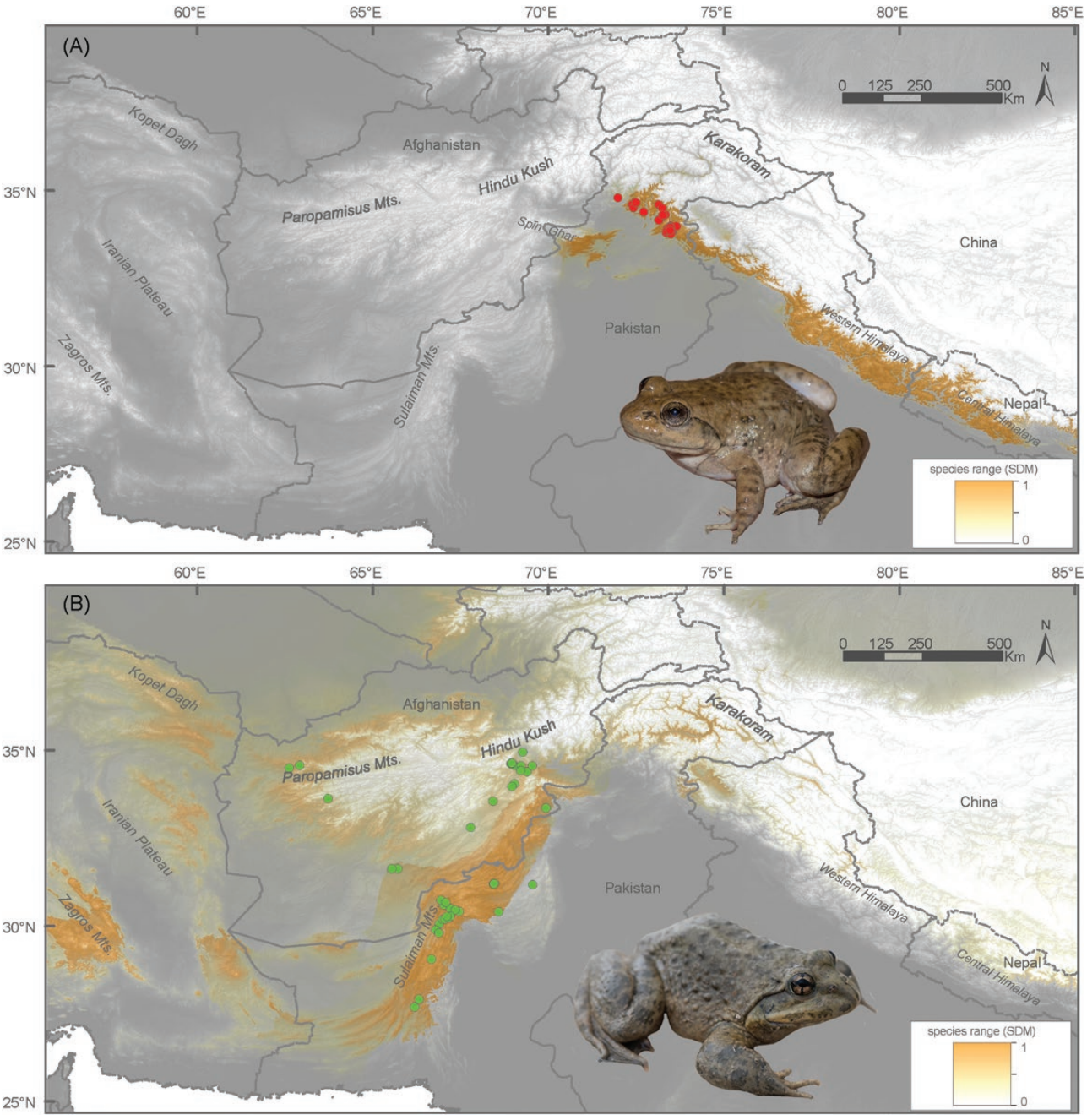

Figure 3. Distribution map for *Allopaa hazarensis* (A) and *Chrysopaa sternosignata* (B) derived from species distribution model (SDM) using MAXENT, including known records of the species (red = *A. hazarensis*, green = *C. sternosignata*). Photo credit: D. Jablonski.

with high distribution probability of *A. hazarensis* in northern Pakistan (Supporting Information, Fig. S6).

The evaluation of the variable contribution in the MAXENT models implies that for both *Allopaa hazarensis* and *Chrysopaa sternosignata*, BIO8 (mean temperature of wettest quarter) was most important to the MAXENT model (Supporting Information, Table S8). Raster values for the original occurrence points were highly significant between the two species (two-tailed test $P < 0.001$, d.f. = 138; Supporting Information, Fig. S7).

Biomod2 results showed that for *Allopaa hazarensis* BIO12 (annual precipitation) followed by BIO6 (minimal temperature of coldest month), and for *Chrysopaa sternosignata* elevation were most important to the ensemble model (Supporting Information, Table

S10). This is well in concert with the results from the MAXENT modelling, where these variables had the most useful information by themselves, since they had the highest gain when used in isolation.

Overall, the species did not show equivalent niche patterns under current climatic conditions in the western HTO ($D = 0.028$, $I = 0.094$). The niche identity tests revealed significant environmental divergence in the SDM comparison, since the observed values of overlap were smaller than the null distributions of background divergence, indicating that the niches of *A. hazarensis* and *C. sternosignata* are different ($P < 0.05$; Supporting Information, Table S11; Fig. S8).

DISCUSSION

HISTORICAL BIOGEOGRAPHY OF HINDU KUSH–HIMALAYAN SPINY FROGS

Consistent with former studies (Che *et al.*, 2010; Hofmann *et al.*, 2019), the South-East Asian *Quasipaa* is sister to all other spiny frogs. Our results also confirm the basal phylogenetic placement of *Chrysopaa* from the Hindu Kush mountains relative to *Allopaa* and *Nanorana* (Hofmann *et al.*, 2021a), and strongly support *A. hazarensis* nested in *Nanorana*, rendering *Nanorana* paraphyletic. In previous work, the paraphyletic nature of *Nanorana* has been already indicated, although its monophyly could not be rejected (Akram *et al.*, 2021; Hofmann *et al.*, 2021b). Phylogenetic resolution and support in weakly supported parts of the tree can be increased by adding more data for a single taxon without adding more characters/genes (San Mauro *et al.*, 2012). Here, we included for the first time a number of additional *A. hazarensis* and *Nanorana* samples from the West Himalaya in the tree, which enhanced support for previously less robust internal branches.

So far, three subgroups are distinguished in *Nanorana*, namely *Chaparana* from montane regions of the southeastern margin of the Tibet Plateau (TP) and mountains of north-eastern China, *Paa* from montane to high-montane regions of the West, Central and East Himalaya, and nominal *Nanorana* from high-montane and alpine regions of the TP and its eastern margin. The strongly supported finding that *Allopaa* is phylogenetically most closely related to *Chaparana*, which occurs at the diametrically opposite end of the HTO near the ancestral area of spiny frogs (Che *et al.*, 2010; Hofmann *et al.*, 2019), strengthens support for the recently proposed hypothesis of a trans-Tibet dispersal of ancestral lineages during the Palaeogene (Hofmann *et al.*, 2021a). Accordingly, it is assumed that the ancestor of *Allopaa* (and *Chrysopaa*) appeared elsewhere near the eastern margin of the developing mountains during the Late Oligocene–Early Miocene (between 28 and 19 Mya; Supporting Information, Fig. S3) and expanded their range up to the western margin of the Himalayan–Tibetan orogenic system. This movement must have been facilitated by a moderately elevated corridor in the Late Oligocene–Early Miocene 'Tibet' (this area should not be perceived like the alpine plateau today) with subtropical and warm-temperate climates and associated sufficient humidity. The current climatic niche differentiation of *A. hazarensis* and *Chrysopaa* fits to this hypothesis. Our scenario is also consistent with subtropical to warm-temperate fossil floras in significant parts of Tibet's interior between 26 and 19 Mya (Ding *et al.*, 2014; Sun *et al.*, 2014; Ai *et al.*, 2019), and with modern geoscientific models of the HTO (Spicer *et al.*, 2021; Xiong *et al.*, 2022). The position and deep divergence of the newly discovered clade with some unidentified specimens endemic to a small area in the West Himalaya (Fig. 1; Supporting Information, Fig. S9) further supports the out-of-Tibet-into-the-Himalayan-exile hypothesis (Schmidt *et al.*, 2012; Hofmann *et al.*, 2019). Because species of this clade are adapted to the warm-temperate zone, as is also observed for *Allopaa*, the existence of large-scale, warm-temperate environments north of the greater Himalaya during the early Neogene has to be assumed (~17 Mya, Supporting Information, Fig. S3; see also: Hofmann *et al.*, 2019, 2021a). The well-supported sister-group position of the newly discovered clade to the Tibetan *Nanorana* and Himalayan *Paa* indicates that this group evolved in the southern parts of palaeo-Tibet and subsequently diversified in the West Himalaya using transverse valleys as immigration corridors. This scenario is similar to the distributional history of wingless ground beetle lineages, which share habitat preferences and distributional patterns, as seen in spiny frogs (Schmidt *et al.*, 2012). Due to the extreme topographic dynamics of the Greater Himalaya and low dispersal capacity of both wingless ground beetles and spiny frogs, these organisms are unable to disperse paralleling this mountain chain. Therefore, most lineages remain endemic to, and diversify within, restricted parts of the Himalaya. The new clade also confirms our previous expectation of the existence of additional, so far unknown lineages endemic to the Kashmir and West Himalaya, which may contribute to resolve the evolution of the HTO (Hofmann *et al.*, 2019).

Similarly, in the *Paa* subgroup, the western Himalayan lineages, including *Nanorana vicina* (Stoliczka, 1872), cluster together with species from the eastern parts of the Himalaya and Transhimalaya [*Nanorana chayunesis* (Ye, 1977), *Nanorana conaensis* (Fei & Huang, 1981), *Nanorana maculosa* (Liu *et al.*, 1960), *Nanorana medogensis* (Fei & Ye, 1999); MRCA 10.5 Mya], but there are no relatives in the vast intermediate area covered by the Central Himalaya

(Supporting Information, Fig. S9). Such a paradoxical distributional pattern can be most parsimoniously explained by the scenario described above, that is, by the evolution of ancestral lineages in palaeo-Tibet during the Miocene, and dispersal and diversification of descendant lineages in the course of the uplifting Greater Himalaya. Based on geoscientific studies, there is high uncertainty with respect to the time of uplift of certain parts of the Greater Himalaya, ranging between *c*. 15 Mya (Gébelin *et al.*, 2013) and relatively recent (Wang *et al.*, 2006). Timing of the Himalayan uplift is one of the most controversially debated aspects in the geosciences of the area and needs to be unravelled (Mulch & Chamberlain, 2006). Based on our phylogeny and previous work (Hofmann *et al.*, 2017, 2019, 2021a), the Himalaya is a young geological feature.

Genetic diversity of *Allopaa hazarensis*

In contrast to the deeply divergent clades in the phylogeny of the spiny frogs, genetic diversity of *Allopaa hazarensis* across its known range is small and without a clear distribution pattern of haplotypes, according to their geographic origin. Major haplotypes are present at multiple localities. However, the Indus seems to act as physical barrier, limiting gene flow between populations. The relatively high haplotype and low nucleotide diversity suggest a recent expansion of the species. This is supported by the higher number of unique haplotypes in relation to all haplotypes (Slatkin & Hudson, 1991; Fu, 1997), although this is influenced by sample sizes. Most likely, all of these haplotypes originated from the respective predominant ancestral haplotypes after the expected population expansion. Such a pattern and interpretation has been reported before for other amphibians (Garcia-Gonzalez *et al.*, 2012; Greenwald *et al.*, 2020). Overall, neither genetic distances, nor haplotype networks and nucleotide variability indicate potential cryptic diversity within *A. hazarensis*. In general, the networks suffer from the low number of sampling sites and, therefore, we consider these results only as preliminary. Because amphibians are predominantly site-loyal and of low vagility (Vences & Wake, 2007; da Fonte *et al.*, 2019), and since *Allopaa* is highly adapted for a semi-aquatic lifestyle in mountain streams, we assume that colonization by *A. hazarensis* is mainly facilitated through (small) aquatic corridors (e.g. by rafting), not via terrestrial dispersal routes. Movement of *A. hazarensis* in natural habitats has recently been addressed using radio transmitters, suggesting almost no overland dispersal (Akram *et al.*, 2022) and movement distances of only a few meters. However, in this study data collection was limited to only eight days in September, and almost 40% of the transmitters were lost during that time, rendering the results less conclusive.

Noteworthy, according to some authors [e.g. Frost (2022) and references therein], the genus *Allopaa* consists of two species, *A. hazarensis* and *Allopaa barmoachensis* (Khan & Tasnim, 1989), the latter originally described as *Rana barmoachensis* Khan & Tasnim, 1989. Based on morphological examination of the holotype, Dubois (1992) and repeatedly Ohler and Dubois (2006) considered this taxon as a junior synonym of *hazarensis* (as *Paa*). Given these data, as well as the geographical proximity of the type locality of *A. barmoachensis* to the main distribution range of *A. hazarensis* (and an even similar elevation of the species records), we tentatively agree with the opinion that *A. barmoachensis* is a synonym of *A. hazarensis*. However, molecular and additional morphological data are required to verify that conclusion.

Taxonomic placement of *Nanorana arunachalensis*

Besides, our *16S* phylogeny and estimates of genetic distances revealed the misclassification of *Nanorana arunachalensis*. This species had been described as *Odorrana arunachalensis* (Saikia *et al.*, 2017) but was recently reassigned to *Nanorana* based on morphological considerations (Qi *et al.*, 2019). We here used the *16S* sequence data of two vouchers (ZSIS-M37: MN496464 and ZSIS-M40: MN636773) from the type locality that were uploaded to GenBank (https://www.ncbi.nlm.nih.gov/genbank/) in September 2019 by Saikia and colleagues, probably in response to the reassignment proposed by Qi *et al.* (2019). Based on our data, the species should be reclassified into *Odorrana* Fei *et al.*, 1990 as long as there are no further molecular and morphological data that support a classification into another genus.

Niche differentiation confirms phylogeny

Our species distribution models, for the first time, present the potential geographic distribution range of the relict taxa *Allopaa hazarensis* and *Chrysopaa sternosignata*, which are geographically separated with an allopatric distribution pattern. Although our MaxEnt model predicted the occurrence of *A. hazarensis* across the West and Central Himalaya (Fig. 3A), this must be rejected. Suitable habitats for the species might be available there, but there is no evidence about the presence of the genus in these areas; especially since the biomod2 ensemble model does not predict this either (Supporting Information, Fig. S6A). On the other hand, such environmental suitability supports our scenario described above and the hypothesis of a trans-Tibet dispersal by ancestral lineages of *A. hazarensis* (and *Chrysopaa*) during the Miocene/Oligocene from the eastern into

the northwestern regions of the HTO. Consequently, a colonization of suitable habitats in the western and central parts of the Himalaya by those lineages must be excluded. In fact, there are no records of *A. hazarensis* in other than the northwestern part of the Himalaya, although that taxon is easy to detect if present due to its strictly semi-aquatic natural history. However, it remains possible that *A. hazarensis* and *Chrysopaa* may occur in other untapped areas, where suitable habitats are present, i.e. the Afghan part of the Hindu Kush and eastern Iran, respectively. Both areas are zoologically least-investigated, thus, the presence of the two taxa there cannot be excluded.

Note that there was a trend for over-estimating suitable areas when using the less restrictive MTP threshold in the MaxEnt model to produce the binary habitat map, in particular for *Chrysopaa sternosignata* (Supporting Information, Fig. S5B). It is well known that SDM predictions, particularly those that are based on small occurrence data and thresholded by MTP values, can result in higher estimates (Vale *et al.*, 2013; Kass *et al.*, 2021).

Conservational needs of relict Hindu Kush–Himalaya lineages

Amphibians are highly vulnerable to changes in thermal and hydric environments due to their ectothermic physiology, unshelled eggs, highly permeable skin and biphasic life-cycles [Araujo *et al.* (2006) and references therein]. Thus, temperature and precipitation are among the most important factors that determine the geographical distribution and abundance of amphibian species (Carey & Alexander, 2003). In both *Allopaa hazarensis* and *Chrysopaa sternosignata*, BIO8 (temperature of the wettest month) was important to the model and was significantly different between the localities of the two species (Supporting Information, Fig. S7), probably simply because the precipitation in northern Pakistan is highest during July/August, while in western Pakistan and adjacent Afghanistan most of the annual precipitation occurs between December and March. The niche identity test analyses revealed a clear divergence of the environmental niches of the two species and indirectly confirmed the different habitat requirements of *A. hazarensis* and *Chrysopaa*. While *A. hazarensis* grows in warm-temperate and wetter places in the foothills to lower montane zone (comparable to those of its sister group *Chaparana*; Ohler *et al.*, 2000; Che *et al.*, 2010), *Chrysopaa* occurs in higher, warmer and drier environments. Reproduction of *Chrysopaa* starts in April, after the rain season; for *A. hazarensis* breeding starts probably later in June with the first summer rain (Dubois & Khan, 1979).

According to the IUCN Red List, both *Allopaa hazarensis* and *Chrysopaa sternosignata* are considered as 'least concern' (Khan *et al.*, 2004, 2008). The two species are flagged with a stable population trend: except from prolonged drought, there are supposedly no threats to *A. hazarensis*, while *Chrysopaa* faces threats from aquatic pollution and over-harvesting for science teaching and research in schools and colleges in Pakistan. However, endemic species inhabiting mountain systems are particularly exposed to the effects of global climate change [Cordier *et al.* (2019) and references therein]. Given their topographic characteristics, the environmental gradients of mountain systems are especially heterogeneous (see also Supporting Information, Fig. S4). Increasing temperatures and decreasing precipitation along these gradients could lead to a significant upward shift of suitable climatic conditions for amphibians and certain other organisms. Logically, this displacement affects species in higher elevations, like *A. hazarensis* and *C. sternosignata*, more negatively because of the reduction of their geographically explicit climatic niche (Blaustein *et al.*, 2010). Consequently, if these species are not able to adapt, they successively will become locally lost by extinction. In fact, glaciers in the northwestern Himalayan region of Jammu and Kashmir have been reported to be retreating at higher rates as compared to other parts of the Himalayan arc [Rashid *et al.* (2021) and references therein], potentially impacting severely the stream-flow regimes. Therefore, we encourage future studies related to the status, trends and fragmentation of *A. hazarensis* and, particularly, *Chrysopaa* populations, in order to determine the appropriate conservation status for these species. Compared to the geographic distribution maps in the most recent status assessment for the IUCN Red List from 2008, our predicted potential distribution of the two frogs differs from these data by a significantly better resolution and higher precision. Our results contribute to the knowledge about the distribution of these species and may provide basic information for guiding future management of them, e.g. by surveying and protecting suitable habitats.

CONCLUSIONS

The Hindu Kush–Himalaya region represents a unique area with a biodiversity that has a potentially high informative value for phylogeographic research against the context of the evolution of the HTO. However, until now, its endemic faunal (and floristic) diversity has remained least-investigated (Jablonski *et al.*, 2021). Our phylogenetic results demonstrate the close relationship of northwestern Himalayan spiny frogs to representatives in the eastern parts of the

HTO, supporting the recently proposed out-of-Tibet-into-the-Himalayan-exile hypothesis and a trans-Tibet dispersal of ancestral spiny frogs during the Palaeogene. Given the current taxonomy (Fig. 1) and a distribution strictly linked to the Himalaya (see also: Hofmann *et al.*, 2019), the genus *Allopaa* [type species *Rana* (*Paa*) *hazarensis* Dubois & Khan, 1979] could be treated under *Nanorana* (type species *Nanorana pleskei* Günther, 1896) rather than as distinct genus. If so, the genus *Allopaa* should be synonymized with *Nanorana*.

Alternatively, *Chaparana* could be elevated to genus level (and consequently also *Paa*), maintaining *Allopaa* at genus level and resolving the paraphyletic nature of *Nanorana*. However, this would imply that each monophyletic Himalayan or Hindu Kush lineage that is identified in the future (what to expect) might be erected to genus level, potentially resulting in a significant taxonomic inflation across *Nanorana*. Considering the shared biogeographic history of these spiny frogs, the present and the new distinct clades to be discovered in the future should all be assigned to *Nanorana s.l.*, until a systematic review of this group is available. Changes in supraspecific classification (and especially at the genus level) should only be carried out on the basis of highly stable phylogenies and if clearly supported by additional information (Vences *et al.*, 2013), e.g. morphological, ecological or behavioural data. Similarly, the new *Nanorana* clade from the West Himalaya with the so far unidentified specimens has to be validated by morphological and additional molecular data. We further suggest excluding *Nanorana arunachalensis* from *Nanorana* and reassigning that species to the genus *Odorrana*.

SDM findings show distinct environmental niches and non-overlapping distribution ranges of *Allopaa hazarensis* and *Chrysopaa sternosignata* and provide crucial information for guiding potential future management of the two species. In view of the rapid global change of today, it might be expected that the species track their niches toward higher elevations or become locally extinct. Currently, understanding climate effects and consequences of other potential threats on *Allopaa* and *Chrysopaa* is hindered by limitations of existing data. Our study sets a baseline to the completion of the overall picture of the species distribution patterns and may aid our understanding to target conservation actions.

ACKNOWLEDGEMENTS

We are grateful to Svetlana D. Vershinina (†2021) for assistance in the field. We thank Jana Poláková for her work in the lab. We also thank the anonymous referees and the editors Jinzhong Fu and Maarten Christenhusz for their constructive comments. This work was supported by the German Research Foundation (DFG, grant no. HO 3792/8-1) to SH. DJ was supported by the Slovak Research and Development Agency under the contract no. APVV-19-0076. LJB was supported by the Zoological Institute within the state assignment of the Ministry of Science and Higher Education, Russian Federation, No. 122031100282-2.

DATA AVAILABILITY

All new sequences were uploaded to GenBank; details on individual samples and accession numbers are available in the Supporting Information, Tables S1 and S2.

REFERENCES

Ahmed W, **Rais M**, **Saeed M**, **Akram A**, **Khan IA**, **Gill S. 2020**. Site occupancy of two endemic stream frogs in different forest types in Pakistan. *Herpetological Conservation and Biology* **15**: 506–511.

Ai K, **Shi G**, **Zhang K**, **Ji J**, **Song B**, **Shen T**, **Guo S. 2019**. The uppermost Oligocene Kailas flora from southern Tibetan Plateau and its implications for the uplift history of the southern Lhasa terrane. *Palaeogeography, Palaeoclimatology, Palaeoecology* **515**: 143–151.

Akram A, **Rais M**, **López-Hervas K**, **Tarvin RD**, **Saeed M**, **Bolnick DI**, **Cannatella DC. 2021**. An insight into molecular taxonomy of bufonids, microhylids, and dicroglossid frogs: first genetic records from Pakistan. *Ecology and Evolution* **11**: 14175–14216.

Akram A, **Rais M**, **Saeed M**, **Ahmed W**, **Gill S**, **Haider J. 2022**. Movement paradigm for Hazara torrent frog *Allopaa hazarensis* and Murree Hills frog *Nanorana vicina* (Anura: Dicroglossidae). *Biodiversity Data Journal* **10**: e84365.

Allouche O, **Tsoar A**, **Kadmon R. 2006**. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* **43**: 1223–1232.

Araujo MB, **Thuiller W**, **Pearson RG. 2006**. Climate warming and the decline of amphibians and reptiles in Europe. *Journal of Biogeography* **33**: 1712–1728.

Blaustein AR, **Walls SC**, **Bancroft BA**, **Lawler JJ**, **Searle CL**, **Gervasi SS. 2010**. Direct and indirect effects of climate change on amphibian populations. *Diversity* **2**: 281–313.

Bouckaert RR, **Drummond AJ. 2017**. bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC Evolutionary Biology* **17**: 1–11.

Boulenger GA. 1920. A monograph of the South Asian, Papuan, Melanesian and Australian frogs of the genus *Rana*. *Records of the Indian Museum* **20**: 1–226.

Breiman L, **Friedman J**, **Olshen RA**, **Stone CS. 1984**. *Classification and regression trees*. Chapman and Hall.

Brown JL. 2014. SDMtoolbox: a python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. *Methods in Ecology and Evolution* **5**: 694–700.

Carey C, **Alexander MA. 2003**. Climate change and amphibian declines: is there a link? *Diversity and Distributions* **9**: 111–121.

Che J, **Zhou WW**, **Hu JS**, **Yan F**, **Papenfuss TJ**, **Wake DB**, **Zhang YP. 2010**. Spiny frogs (Paini) illuminate the history of the Himalayan region and Southeast Asia. *Proceedings of the National Academy of Sciences of the USA* **107**: 13765–13770.

Cordier JM, **Lescano JN**, **Rios NE**, **Leynaud GC**, **Nori J. 2019**. Climate change threatens micro-endemic amphibians of an important South American high-altitude center of endemism. *Amphibia-Reptilia* **41**: 1–11.

Ding L, **Xu Q**, **Yue Y**, **Wang HJ**, **Cai F**, **Li S. 2014**. The Andean-type Gangdese Mountains: paleoelevation record from the Paleocene–Eocene Linzhou Basin. *Earth and Planetary Science Letters* **392**: 250–264.

Drummond AJ, **Rambaut A. 2007**. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* **7**: 214.

Dubois A. 1975. A new sub-genus (*Paa*) and three new species of the genus *Rana*. Remarks on the phylogeny of Ranidae (Amphibia, Anura) [translated from French]. *Bulletin du Muséum National d'Histoire Naturelle Zoologie* **231**: 1093–1115.

Dubois A. 1992. Notes sur la classification des Ranidae (Amphibiens Anoures). *Bulletin Mensuel de la Société Linnéenne de Lyon* **61**: 305–352.

Dubois A, **Khan MS. 1979**. A new species of frog (genus *Rana*, subgenus *Paa*) from northern Pakistan (Amphibia, Anura). *Journal of Herpetology* **13**: 403.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**: 1792–1797.

Elith J, **Graham CH**, **Anderson RP**, **Dudı´k M**, **Ferrier S**, **Antoine G**, **Hijmans RJ**, **Huettmann F**, **Leathwick JR**, **Lehmann A**, **Li J**, **Lohmann LG**, **Loiselle BA**, **Manion G**, **Moritz C**, **Nakamura M**, **Nakazawa Y**, **Overton JM**, **Peterson AT**, **Phillips SJ**, **Richardson K**, **Scachetti-Pereira R**, **Schapire RE**, **Sobero´n J**, **Williams S**, **S WM**, **Zimmermann NE. 2006**. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* **29**: 129–151.

Fick SE, **Hijmans RJ. 2017**. WorldClim 2: new 1 km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37**: 4302–4315.

da Fonte LFM, **Mayer M**, **Lötters S. 2019**. Long-distance dispersal in amphibians. *Frontiers of Biogeography* **11**: e44577.

Frost DR. 2022. *Amphibian species of the world: an online reference, v. 6.1. Electronic database*. New York: American Museum of Natural History.

Fu YX. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* **147**: 915–925.

Garcia-Gonzalez C, **Campo D**, **Pola IG**, **Garcia-Vazquez E. 2012**. Rural road networks as barriers to gene flow for amphibians: species-dependent mitigation by traffic calming. *Landscape and Urban Planning* **104**: 171–180.

Gébelin A, **Mulch A**, **Teyssier C**, **Jessup MJ**, **Law RD**, **Brunel M. 2013**. The Miocene elevation of Mount Everest. *Geology* **41**: 799–802.

Greenwald K, **Stedman A**, **Mifsud D**, **Stapleton M**, **Larson K**, **Chellman I**, **Parrish DL**, **Kilpatrick CW. 2020**. Phylogeographic analysis of mudpuppies (*Necturus maculosus*). *Journal of Herpetology* **54**: 78–86.

Hanley JA, **McNeil BJ. 1982**. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**: 29–36.

Hastie TJ, **Tibishirani R. 1990**. *Generalized additive models*. London: Chapman and Hall.

Hofmann S, **Stoeck M**, **Zheng Y**, **Ficetola FG**, **Li JT**, **Scheidt U**, **Schmidt J. 2017**. Molecular phylogenies indicate a Paleo-Tibetan origin of Himalayan lazy toads (*Scutiger*). *Scientific Reports* **7**: 3308.

Hofmann S, **Baniya CB**, **Litvinchuk SN**, **Miehe G**, **Li JT**, **Schmidt J. 2019**. Phylogeny of spiny frogs *Nanorana* (Anura: Dicroglossidae) supports a Tibetan origin of a Himalayan species group. *Ecology and Evolution* **9**: 14498–14511.

Hofmann S, **Jablonski D**, **Litvinchuk SN**, **Masroor R**, **Schmidt J. 2021a**. Relict groups of spiny frogs indicate Late Paleogene-Early Neogene trans-Tibet dispersal of thermophile faunal elements. *PeerJ* **9**: e11793.

Hofmann S, **Masroor R**, **Jablonski D. 2021b**. Morphological and molecular data on tadpoles of the westernmost Himalayan spiny frog *Allopaa hazarensis* (Dubois & Khan, 1979). *ZooKeys* **1049**: 67–77.

Jablonski D, **Basit A**, **Farooqi J**, **Masroor R**, **Böhme W. 2021**. Biodiversity research in a changing Afghanistan. *Science* **6549**: 1402.

Kass JM, **Meenan SI**, **Tinoco N**, **Burneo SF**, **Anderson RP. 2021**. Improving area of occupancy estimates for parapatric species using distribution models and support vector machines. *Ecological Applications* **31**: e02228.

Khan MS. 2006. *Amphibians and reptiles of Pakistan*. Malabar, FL: Krieger Publishing Company.

Khan MS, **Papenfuss T**, **Anderson S**, **Kuzmin S**, **Dutta S**, **Ohler A**, **Sengupta S. 2004**. Chrysopaa sternosignata. *The IUCN red list of threatened species* **2004**: e.T58440A11781495.

Khan MS, **Dutta S**, **Ohler A. 2008**. *Allopaa hazarensis*. *The IUCN red list of threatened species* **2022**: e.T58426A166103757.

Kozlov AM, **Darriba D**, **Flouri T**, **Morel B**, **Stamatakis A. 2019**. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**: 4453–4455.

Kumar S, **Stecher G**, **Li M**, **Knyaz C**, **Tamura K. 2018**. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution* **35**: 1547–1549.

Lanfear R, **Calcott B**, **Ho SY**, **Guindon S. 2012**. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* **29**: 1695–1701.

Lemoine F, **Domelevo Entfellner JB**, **Wilkinson E**, **Correia D**, **Davila Felipe M**, **De Oliveira T**, **Gascuel O. 2018**. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* **556**: 452–456.

McCullagh P, **Nelder JA. 1989**. *Generalized linear models*. London: Chapman and Hall.

McIntyre PJ. 2012. Polyploidy associated with altered and broader ecological niches in the *Claytonia perfoliata* (Portulacaceae) species complex. *American Journal of Botany* **99**: 655–662.

Mertens R. 1969. Die Amphibien und Reptilien West-Pakistans. *Stuttgarter Beiträge zur Naturkundeaus dem Staatlichen Museum für Naturkunde in Stuttgart* **197**: 1–96.

Miehe G. 1991. Der Himalaya, eine multizonale Gebirgsregion. In: Walter H, Breckle S-W, eds. *Ökologie der Erde, Band 4. Spezielle Ökologie der Gemäßigten und Arktischen Zonen außerhalb*. Euro-Nordasiens. Stuttgart: Gustav Fischer, 181–230.

Miller J, **Franklin J. 2002**. Modeling the distribution of four vegetation alliances using generalized linear models and classification trees with spatial dependence. *Ecological Modelling* **157**: 227–247.

Monserud RA, **Leemans R. 1992**. Comparing global vegetation maps with Kappa statistics. *Ecological Modelling* **62**: 275–293.

Mulch A, **Chamberlain CP. 2006**. Earth science - The rise and growth of Tibet. *Nature* **439**: 670–671.

Murray JA. 1885. A new frog (*Rana sternosignata*) from Sind. *Annals and Magazine of Natural History* **5**: 120–121.

Ohler A, **Dubois A. 2006**. Phylogenetic relationships and generic taxonomy of the tribe Paini (Amphibia, Anura, Ranidae, Dicroglossinae), with diagnoses of two new genera. *Zoosystema* **28**: 769–784.

Ohler A, **Marquis O**, **Swan S**, **Grosjean S. 2000**. Amphibian biodiversity of Hoang Lien Nature Reserve (Lao Cai Province, northern Vietnam) with description of two new species. *Herpetozoa* **13**: 71–87.

Pearson RG, **Raxworthy CJ**, **Nakamura M**, **Peterson AT. 2006**. Original article: predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography* **34**: 102–117.

Phillips SJ, **Dudık M**, **Schapire RE. 2004**. A maximum entropy approach to species distribution modeling. Twenty-First International Conference on Machine Learning, 4–8 July. New York, NY, USA, 655–662.

Phillips SJ, **Anderson RP**, **Schapire RE. 2006**. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190**: 231–259.

Préau C, **Trochet A**, **Bertrand R**, **Isselin-Nondedeu F. 2018**. Modeling potential distributions of three European amphibian species comparing ENFA and MaxEnt. *Herpetological Conservation and Biology* **13**: 91–104.

Qi S, **Zhou ZY**, **Lu YY**, **Li JL**, **Qin HH**, **Hou M**, **Zhang Y**, **Ma J**, **Li PP. 2019**. A new species of *Nanorana* (Anura: Dicroglossidae) from southern Tibet, China. *Russian Journal of Herpetology* **26**: 159–174.

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

Rashid I, **Majeed U**, **Najar N**, **Bi A. 2021**. Retreat of Machoi Glacier, Kashmir Himalaya between 1972 and 2019 using remote sensing methods and field observations. *Science of the Total Environment* **785**: 147376.

Ridgeway G. 1999. The state of boosting. *Computing Science and Statistics* **31**: 172–181.

Ripley BD. 1996. *Pattern recognition and neural networks*. Cambridge: Cambridge University Press.

Rödder D, **Engler JO. 2011**. Quantitative metrics of overlaps in Grinnellian niches: advances and possible drawbacks. *Global Ecology and Biogeography* **20**: 915–927.

Ronquist F, **Teslenko M**, **van der Mark P**, **Ayres DL**, **Darling A**, **Hohna S**, **Larget B**, **Liu L**, **Suchard MA**, **Huelsenbeck JP. 2012**. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61**: 539–542.

Saikia B, **Sinha B**, **Kharkongor IJ. 2017**. *Odorrana arunachalensis*: a new species of cascade frog (Anura: Ranidae) from Talle Valley Wildlife Sanctuary, Arunachal Pradesh, India. *Journal of Bioresources* **4**: 30–41.

San Mauro D, **Gower DJ**, **Cotton JA**, **Zardoya R**, **Wilkinson M**, **Massingham T. 2012**. Experimental design in phylogenetics: testing predictions from expected information. *Systematic Biology* **61**: 661–674.

Schmidt J, **Opgenoorth L**, **Holl S**, **Bastrop R. 2012**. Into the Himalayan exile: the phylogeography of the ground beetle *Ethira* clade supports the Tibetan origin of forest-dwelling Himalayan species groups. *PLoS One* **7**: e45482.

Schoener TW. 1968. The *Anolis* lizards of Bimini – Resource partitioning in a complex fauna. *Ecology* **49**: 704–726.

Schoniger M, **von Haeseler A. 1999**. Toward assigning helical regions in alignments of ribosomal RNA and testing the appropriateness of evolutionary models. *Journal of Molecular Evolution* **49**: 691–698.

Slatkin M, **Hudson RR. 1991**. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129**: 555–562.

Spicer RA, **Su T**, **Valdes PJ**, **Farnsworth A**, **Wu FX**, **Shi G**, **Spicer TEV**, **Zhou Z. 2021**. Why 'the uplift of the Tibetan Plateau' is a myth. *National Science Review* **8**: nwaa091.

Stocsits RR, **Letsch H**, **Hertel J**, **Misof B**, **Stadler PF. 2009**. Accurate and efficient reconstruction of deep phylogenies from structured RNAs. *Nucleic Acids Research* **37**: 6184–6193.

Sun J, **Xu Q**, **Liu WZ**, **Zhang Z**, **Xue L**, **Zhao P. 2014**. Palynological evidence for the latest Oligocene–early Miocene paleoelevation estimate in the Lunpola Basin, central Tibet. *Palaeogeography, Palaeoclimatology, Palaeoecology* **399**: 21–30.

Swets JA. 1988. Measuring the accuracy of diagnostic systems. *Science* **240**: 1285–1293.

Thuiller W, **Georges D**, **Engler R. 2013**. Biomod2: ensemble platform for species distribution modeling. *R package v.2.7*.

Thuiller W, **Georges D**, **Gueguen M**, **Engler R**, **Breiner FT. 2021**. Biomod2: ensemble platform for species distribution modeling. *R package v.3.5.1*.

Vale CG, **Tarroso P**, **Brito JC. 2013**. Predicting species distribution at range margins: testing the effects of study area extent, resolution and threshold selection in the Sahara–Sahel transition zone. *Diversity and Distributions* **20**: 20–33.

Vences M, **Wake DB. 2007**. Speciation, species boundaries and phylogeography of amphibian. In: Heatwole HH, Tyler

M, eds. *Amphibian biology, 6, systematics*. Chipping Norton: Surrey Beatty & Sons, 2613–2669.

Vences M, **Guayasamin JM**, **Miralles A**, **De La Riva I. 2013**. To name or not to name: criteria to promote economy of change in Linnaean classification schemes. *Zootaxa* **3636**: 201–244.

Wagner P, **Bauer AM**, **Leviton AE**, **Wilms TM**, **Böhme W. 2016**. A checklist of the amphibians and reptiles of Afghanistan – Exploring herpetodiversity using biodiversity archives. *Proceedings of the California Academy of Sciences* **63**: 457–565.

Wang Y, **Deng T**, **Biasatti D. 2006**. Ancient diets indicate significant uplift of southern Tibet after ca. 7 Ma. *Geology* **34**: 309–312.

Warren DL, **Glor RE**, **Turelli M. 2008**. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* **62**: 2868–2883.

Warren DL, **Glor RE**, **Turelli M. 2010**. ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography* **33**: 607–611.

Xiong Z, **Liu X**, **Ding L**, **Farnsworth A**, **Spicer RA**, **Xu Q**, **Valdes P**, **He S**, **Zeng D**, **Wang C**, **Li Z**, **Guo X**, **Su T**, **Zhao C**, **Wang H**, **Yue Y. 2022**. The rise and demise of the Paleogene Central Tibetan Valley. *Science Advances* **8**: eabj0944.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article on the publisher's website.

Table S1. GenBank accession numbers of sequences used in this study. Taxa with newly obtained sequences are indicated by an asterisk.

Table S2. Locations (Loc) of the two species (Sp) *Allopaa hazarensis* (A) and *Chrysopaa sternosignata* (C) as shown in Figure 3. Some coordinates and elevations are approximations from descriptions in the original publications. Lat(itude) and Long(itude) are given in decimal degree. Ctr = country (Pakistan [PK], Afghanistan [AFG]), Alt = elevation. Samples CUHC10278 and CUHC11352 could not be georeferenced unambiguously due to lack of information.

Table S3. Pairwise correlation of variables used for species distribution modelling: Pearson's r^2 (upper triangular) and Spearman's ρ (lower triangular).

Table S4. Uncorrected genetic distances (%; lower left matrix), including standard error estimates (upper right matrix) between *16S* sequences of taxa as shown in the Supporting Information, Figure S2. A = *Allopaa hazarensis*, C = *Chrysopaa sternosignata*. *Nanorana* species names were abbreviated after five characters and are as follows: *phrynoides*, *sichuanensis*, *zhaoermii*, *arunachalensis*, *vicina* (prefix WLMNV), *xuelinensis*, *aenea*, cf. *blanfordii*, cf. *polunini*, *rarica*, cf. *rostandi*, *chayuensis*, *conaensis*, *liebigii*, *maculosa*, *medogensis*, *pleskei*, *quadranus*, sp. (2Bahn_RAS), cf. *vicina* (2Pul_RAS, 782_RAS), sp. (A1966/13_NME), *taihangnica*, *unculuanus*, *ventripunctata*, *yunnanensis*, cf. *ercepea*, *parkeri*, *Quasipaa boulengeri*, sp. (VV5.1/8.1/11.1_RAS), cf. *vicina* (VV9.1_RAS). A star after the abbreviated species name refers to 'cf. '. Grey shaded cells indicate the large genetic distance between *N. arunachalensis* and other spiny frogs, including *Nanorana* species, showing the taxonomic misclassification of that species. Values in coloured cells are distances < 2.5%.

Tables S5. a,b Uncorrected genetic distances (%; lower left matrix), including standard error estimates (upper right matrix), between taxa used in this study for *COI* (a) and *Rag1* (b). A star after the species name refers to 'cf. '. Bold values in grey shaded cells highlight distances lower than 5% (*COI*) and lower than 0.1% (*Rag1*). A = *Allopaa hazarensis*, C = *Chrysopaa sternosignata*, N = *Nanorana*, Q = *Quasipaa boulengeri*.

Table S6. a–c Uncorrected pairwise genetic distances between sequences of *Allopaa hazarensis* for *16S* (a), *COI* (b) and *Rag1* (c). Distances ≥ 1% (*16S*) and ≥ 0.5% (*COI*) and ≥ 0.2% (*Rag1*) are highlighted bold. [#]CUHC = Comenius University Herpetological Collection; *=prefix 'WLM:AH'; [$]=ZFMK(SH20) (Museum Koenig, Bonn).

Table S7. a,b Uncorrected genetic distances (%; lower left matrix) with standard error (upper right matrix) between *Allopaa hazarensis* (A), *Chrysopaa* (C), *Quasipaa* (Q) and subgenera of *Nanorana* (*Chaparana* [Ch], *Nanorana* [N], *Paa* [P]), including the unknown clade from the Western Himalaya (WH) for *COI* (a) and *Rag1* (b).

Table S8. SDM performance and evaluation for *Allopaa hazarensis* (A) and *Chrysopaa sternosignata* (C). Information on the model performance and evaluation, and the variable contribution are given. Test AUC values indicate performance as follows: > 0.9 excellent, > 0.8 good, and > 0.7 useful discrimination ability of the model. High values in variable contribution are highlighted bold.

Table S9. Performance evaluation scores for the final ensemble classifier combining biomod2 algorithms.

Table S10. Statistics for single models specified in biomod2; (a) for *Allopaa hazarensis* and (b) *Chrysopaa sternosignata*.

Table S11. Observed niche overlap values and results of niche identity test. Empirical overlap values smaller than the null distribution support niche divergence. Asterisk denote significance at $^{*}P < 0.05$.

Figure S1. PCA of 19 WorldClim v.2.1 variables, showing climate space: the more similar the colours the more similar values. Records of *Chrysopaa sternosignata* are indicated as green filled circles (for details see Supporting

Information, Table S1). The arrow points to Malir, Pakistan, that is noted as type locality [Mulleer (= Malir) near Kurrachee (= Karachi)] for the syntypes BMNH 1947.2.1.21, 1947.2.1.22 of *C. sternosignata*. Malir has apparently substantially different climate conditions compared to the known distribution range of the species. Thus, we consider that type locality as erroneously defined.

Figure S2. Bayesian inference tree based on *16S* rRNA sequence data. Numbers at branch nodes refer to posterior probabilities ≥ 0.9. The clade of *Allopaa hazarensis* is indicated red, while *Chrysopaa sternosignata* is highlighted green. Species names are followed by voucher number (if available).

Figure S3. Ultametric time-calibrated phylogeny generated with BEAST2 based on the concatenated sequence data of spiny frogs. Grey bars specify the 95% HPD for the respective nodes; ages are shown for nodes that are supported by Bayesian posterior probability ≥ 0.95.

Figure S4. Climate heterogeneity raster based on recent WorldClim v.2.1 data; warm colours depict high areas of climatic heterogeneity.

Figure S5. Distribution map for *Allopaa hazarensis* (A) and *Chrysopaa sternosignata* (B) derived from species distribution model (SDM) using MAXENT and a minimum training presence threshold. Maps include known records of the species (red = *A. hazarensis*, green = *C. sternosignata*).

Figure S6. Distribution map for *Allopaa hazarensis* (A) and *Chrysopaa sternosignata* (B) derived from species distribution model (SDM) using biomod2, including known records of the species (red = *A. hazarensis*, green = *C. sternosignata*).

Figure S7. Environmental variable BIO8 (mean temperature of wettest quarter) across the modelled area based on WorldClim v.2.1 climate data for 1970–2000. Records of *Allopaa hazarensis* and *Chrysopaa sternosignata* are indicated by red and green circles, respectively.

Figure S8. Sample output from ENMTOOLS v.1.4.4 for identity test for all three implemented niche overlap metrics. Histograms represent the distribution of overlaps for each metric from the null distribution, while the dashed vertical line represents the overlap between the models built using the empirical data.

Figure S9. Simplified Bayesian inference tree with the main (sub)genera mapped to High Asia with the Indus and Brahmaputra River systems. Topology and colour codes of the clades match Figure 1.