

DR. YANG LIU (Orcid ID : 0000-0003-4580-5518)

DR. ALEXANDER SUH (Orcid ID : 0000-0002-8979-9992)

PROF. GERALD HECKEL (Orcid ID : 0000-0002-0162-323X)

DR. MANUEL SCHWEIZER (Orcid ID : 0000-0002-7555-8450)

Article type : Original Article

Seasonal migration patterns and the maintenance of evolutionary diversity in a cryptic bird radiation

Running title: Cryptic radiation of pale sand martins

Qindong Tang^{1,2}, Reto Burri³, Yang Liu⁴, Alexander Suh^{5,6}, Gombobaatar Sundev⁷, Gerald Heckel^{1,*} Manuel Schweizer^{1,2,*}

¹Institute of Ecology and Evolution, University of Bern, Switzerland

²Natural History Museum, Bern, Switzerland

³Schweizerische Vogelwarte, 6204 Sempach, Switzerland

⁴State Key Laboratory of Biocontrol, College of Ecology School of Life Science, Sun Yat-sen University, Guangzhou, China

⁵School of Biological Sciences—Organisms and the Environment, University of East Anglia, Norwich, UK

⁶Department of Organismal Biology – Systematic Biology, Evolutionary Biology Centre (EBC), Uppsala University, Uppsala, Sweden

⁷National University of Mongolia and Mongolian Ornithological Society, Ulaanbaatar, Mongolia

* Manuel Schweizer and Gerald Heckel are joint senior authors

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the [Version of Record](#). Please cite this article as [doi: 10.1111/MEC.16241](https://doi.org/10.1111/MEC.16241)

This article is protected by copyright. All rights reserved

Corresponding author:

Manuel Schweizer

Natural History Museum, Bern, Switzerland

Institute of Ecology and Evolution, University of Bern, Switzerland

Email: manuel.schweizer@nmbe.ch

Accepted Article

Abstract

Morphological differentiation associated with evolutionary diversification is often explained with adaptive benefits but the processes and mechanisms maintaining cryptic diversity are still poorly understood. Using genome-wide data, we show here that the pale sand martin *Riparia diluta* in Central and East Asia consists of three genetically deeply differentiated lineages which vary only gradually in morphology but broadly reflect traditional taxonomy. We detected no signs of gene flow along the eastern edge of the Qinghai-Tibetan plateau between lowland south-eastern Chinese *R. d. fohkienensis* and high-altitude *R. d. tibetana*. Largely different breeding and migration timing between these low and high altitude populations as indicated by phenology data suggests that allochrony might act as prezygotic isolation mechanism in the area where their ranges abut. Mongolian populations of *R. d. tibetana*, however, displayed signs of limited mixed ancestries with Central Asian *R. d. diluta*. Their ranges meet in the area of a well-known avian migratory divide, where western lineages take a western migration route around the Qinghai-Tibetan plateau to winter quarters in South Asia, and eastern lineages take an eastern route to Southeast Asia. This might also be the case between western *R. d. diluta* and eastern *R. d. tibetana* as indicated by differing wintering grounds. We hypothesize that hybrids might have non-optimal intermediate migration routes and selection against them might restrict gene flow. Although further potential isolation mechanisms might exist in the pale sand martin, our study points towards contrasting migration behaviour as an important factor in maintaining evolutionary diversity under morphological stasis.

Key words: Sand martin, *Riparia diluta*, cryptic diversification, phylogeography, population genomics, allochrony

1 Introduction

How biodiversity is generated and maintained remains one of the main questions in evolutionary biology. Studies on the evolution of ecological and phenotypic diversity within rapidly multiplying lineages - often adaptive radiations - have been instrumental for our understanding on how adaptive processes trigger speciation (McGee et al., 2020; Nosil, 2012; Schluter, 2000; Yoder et al., 2010). However, divergent adaptation is not the only path towards speciation and evolutionary diversification does not necessarily result in remarkable morphological and ecological differentiation. In so-called non-adaptive radiations (aRundell & Price, 2009), lineages diversify within similar environments in allopatry or parapatry, and thus under similar regimes of natural selection. Such radiations can therefore be accompanied by minimal morphological and ecological differentiation (e.g. Fink et al., 2010). As a result, this may lead to genetically differentiated, yet morphologically cryptic lineages but differences in other aspects such as sexual signalling, physiology or phenology might have evolved (Braune et al., 2008; Feckler et al., 2014; Taylor & Friesen, 2017).

Over the last two decades, an increasing number of such morphologically cryptic lineages has been discovered through genetic methods, thus uncovering an unexpected evolutionary diversity across the tree of life (e. g. Kozak et al., 2006; Leys et al., 2017; Slavenko et al., 2020; Weir et al., 2016). Indeed, a considerable proportion of biodiversity may be constituted of 'cryptic species', i.e. evolutionary lineages with restricted gene flow that "do not form diagnostic morphological clusters" (Struck et al., 2018). Investigating their evolution is important to determine the processes under which biodiversity evolves in general and to increase our still limited understanding of diversification under morphological stasis in particular (Fišer et al., 2018). The latter may contrast with the processes and mechanisms governing adaptive radiations, e.g. in intensively studied systems such as Darwin's finches (Grant & Grant, 2011) or African lake cichlids (McGee et al., 2020; Seehausen, 2006).

Unlike the seemingly rare cases of explosive diversification in adaptive radiations, speciation is generally considered a slow process with spatial isolation usually required to initiate lineage divergence (Price, 2008; Tobias et al., 2020). After an initial phase of allopatry during the speciation process, the critical question is whether or not the differentiated lineages withstand gene flow at secondary contact. In the context of cryptic radiations, investigating evolutionary

processes that might restrict gene flow among lineages lacking obvious morphological differences is crucial for our understanding on how diversity can be maintained in such cases (e.g. Beysard et al., 2015; Beysard & Heckel, 2014).

The break-down of lineage integrity – often referred to as ‘speciation reversal’ (Kearns et al., 2018; Seehausen et al., 2008) – is the expected outcome of secondary contact in taxa with insufficient or ephemeral reproductive isolation. Extensive hybridization between morphologically cryptic lineages and the potentially resulting fusion of lineages, however, might remain difficult to detect, unless species complexes are sampled comprehensively across their ranges and analysed with genome-wide data (Slager et al., 2020). However, morphologically cryptic lineages might have different adaptations in life-history traits not reflected in morphology. These might include contrasting habitat preferences precluding secondary sympatry or different timing of reproduction facilitating co-existence through temporal segregation (allochrony) (Leys et al., 2017; Taylor & Friesen, 2017). Additionally, morphologically cryptic lineages might have diverged in social signals (Tobias et al., 2020) and thus show premating isolation at secondary contact. The resulting assortative mating by itself might not be sufficient to prevent eventual lineage fusion, unless hybrid fitness is reduced (Irwin, 2020). Such postzygotic isolation might not only be caused by hybrids showing intermediate or ‘transgressive’ signals, but also by genetic incompatibilities. The evolution of genetic incompatibilities, however, is usually considered to be too slow to generally play an important role at the onset of the speciation process (Price, 2008; Price & Bouvier, 2002), although the relationship between divergence time in allopatry and different stages along the speciation continuum remains poorly understood (Beysard & Heckel, 2014; Dufresnes et al., 2019). Interestingly, in recent studies on cryptic sister species of birds in Amazonia, strong postzygotic reproductive isolation was found with little evidence for premating isolation even in relatively young pairs (Cronemberger et al., 2020; Pulido-Santacruz et al., 2018). Whether genetic incompatibilities among cryptic species might play a more important role than previously thought and actually accumulate faster than premating isolation under certain circumstances (Cronemberger et al., 2020) is under debate, and remains to be tested in additional geographic contexts and systems.

The processes and mechanisms that lead to morphologically cryptic divergence and maintain genetic partitioning thereafter are best amenable to research in systems where multiple cryptic

lineages are found in geographic contact. The pale sand martin (*Riparia diluta*) of Central and East Asia provides such a promising natural system to study diversification under morphological stasis. Its four recognized subspecies overlap in morphometrics and have no diagnostic differences in plumage features (Schweizer et al., 2018). The subtle differences in plumage comprise color shade, prominence of ear-covert coloration as well as extension of breast-band, but the identification of single individuals without context is usually not possible (Schweizer et al., 2018; Shirihai & Svensson, 2018). This crypsis in morphology contrasts with deep phylogeographic structure in mtDNA among subspecies (Schweizer et al., 2018). For a long time, these birds were considered conspecific with the collared sand martin *Riparia riparia* that has a Holarctic distribution. The ranges of *R. diluta* and *R. riparia* widely overlap in East Asia without apparent interbreeding, and they show subtle but consistent differences in plumage features, vocalizations and genetics (Gavrilov & Savchenko, 1991; Goroshko, 1993; Pavlova et al., 2008; Schweizer & Aye, 2007). The morphologically cryptic phylogeographic lineages within *R. diluta* occur in different climate zones and widely differing altitudinal ranges. Nominate *R. d. diluta* is found in the steppes of Central Asia, *R. d. fohkienensis* in subtropical south China, *R. d. indica* in the north-western part of the Indian subcontinent, and *R. d. tibetana* breeds on the Qinghai-Tibetan plateau (Figure 1). While *R. d. indica* is geographically isolated, the breeding areas of the remaining subspecies are thought to be largely contiguous, but geographic sampling in previous studies was limited, especially in areas of potential secondary contact (Schweizer et al., 2018). Here, analyzing variation at >7100 single nucleotide polymorphisms (SNPs) based on a comprehensive geographic sampling of pale sand martin populations around the Qinghai-Tibetan plateau, we aimed to test 1) if the deep phylogeographic structure indicated by mtDNA is also reflected in genome-wide variation and 2) if the phylogeographic lineages indeed overlap in morphospace and can thus be considered cryptic. Finally, we 3) investigated, if gene flow between phylogeographic lineages is actually reduced in areas of potential contact.

2 Material and Methods

2.1 Sampling and DNA extraction

Blood was collected from 149 *Riparia* individuals (119 *R. diluta*; 30 *R. riparia*) on their breeding grounds (Fig. 1, Table S1.). We sampled 14 breeding colonies of *R. diluta* covering the breeding

Accepted Article

areas of three subspecies and including potential areas of contact among them at the north-western and eastern edges of the Qinghai-Tibetan plateau, and additionally three colonies of *R. riparia*. This included 37 individuals from four populations of *R. d. diluta* from north-western China, 29 individuals from three populations of *R. d. tibetana* from the Qinghai-Tibetan plateau, 29 individuals from three populations of *R. d. fohkienensis* from south-eastern China, and 24 individuals of *R. d. diluta* from Mongolia. We grouped three pairs of breeding colonies from the same geographic regions in one population each resulting in a total of 14 populations (11 for *R. diluta* and 3 for *R. riparia*) for further analyses (Table S1, Figure 1). All samples were preserved in 99% ethanol and later stored at -20 °C. DNA was extracted with a modified salt extraction protocol (Pavlova et al., 2008). Birds were captured and sampled under permits and approvals from the relevant authorities in China and Mongolia.

2.2 Sequencing and analysis of mtDNA

To investigate whether the deep phylogeographic structure among subspecies in mtDNA can be recovered with a more extensive geographic sampling, we amplified a fragment (~850 bp) of the mitochondrial NADH dehydrogenase subunit 2 (ND2) gene of all 149 individuals using the same protocol as Schweizer et al. (2018). Sanger sequencing was performed in both directions by LGC Genomics GmbH (Germany). Complementary strands were aligned using the package SeqMan in DNASTar (Burland, 2000). Unique ND2 haplotypes were detected using DnaSP V6.12 (Rozas et al., 2017). Phylogenetic reconstruction based on unique haplotypes was done using maximum likelihood (ML) and Bayesian inference (BI). Barn swallow *Hirundo rustica* (GeneBank accession number DQ176515) was used as outgroup and one sequence of *R. d. indica* (Schweizer et al., 2018; GeneBank accession number MG881167) was additionally added. Sequence alignment was done using ClustalW in MEGA 7 (Kumar et al., 2016). TN93+I was selected as the best-fitting model of nucleotide substitution with phymtest in PhyML (Guindon & Gascuel, 2003) with R 3.6.2 (R core Team, 2013). ML tree search was employed in PhyML with 1,000 bootstrap replicates. BI was performed with MrBayes 3.2.7 (Ronquist et al., 2012) with four independent runs of Metropolis-coupled Markov chain Monte Carlo analyses. Each run comprised one cold chain and three heated chains at a default temperature of 0.1. The chains were run for 20 million generations and sampled every 100 generations. Tracer 1.7 (Rambaut et al., 2018) was used to

assess the length of burn-in and to confirm adequate effective sample sizes ($ESS > 200$) of the posterior distribution. Clades were considered as supported by our analyses when bootstrap values were $> 70\%$ (Hillis and Bull, 1993) and clade credibility values for the BI > 0.95 (Huelsenbeck and Ronquist, 2001). The final ML and BI phylogenetic trees were edited for display in FigTree v1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>).

2.4 *De novo* reference genome sequencing and assembly

A draft genome was sequenced and assembled *de novo* by National Genomic Infrastructure Stockholm, Sweden, from a male *R. riparia* collected in Zhengzhou, Henan province, China in 2017 (SYSb6505, Table S2). DNA extraction and linked-read sequencing were performed as in Lutgen et al. (2020). In brief, DNA was extracted using the Qiagen MagAttract HMW DNA kit following the manufacturer's instructions except for using the double volume and prolonging the digestion time. A single linked-read sequencing library was then prepared using the 10X Genomics Chromium Genome library kit, and sequenced on half an S4 lane on a NovaSeq 6000 instrument at NGI Stockholm. A *de novo* reference genome was assembled with Supernova assembler version 2.1.0 (Weisenfeld et al., 2017). We obtained a pseudohaploid draft reference with a total assembly length of 1.18 Gb, effective read coverage of 40.9X and scaffold N50 of 12.6 Mb. All scaffolds were then mapped to the well annotated genome of zebra finch *Taeniopygia guttata* version bTaeGut1_v1.p (Korlach et al., 2017) using minimap2 (Li, 2018). Only scaffolds that were larger than 1000 bp and uniquely mapped to the zebra finch reference genome were retained with the largest scaffold being 32.8 Mb in length. As the reference genome was from a male (homomorphic sex ZZ in birds), scaffolds mapping to the Z chromosome were excluded in further analyses to avoid underestimation of heterozygosity in females (heteromorphic sex ZW). Moreover, scaffolds of the mitochondrial genome were also excluded and thus only autosomal SNPs retained. The total length of the final assembly was 980 Mb. Genome heterozygosity as the proportion of heterozygous sites of the sand martin reference genome was estimated based on k-mer count distribution of kmer length of 21 ($m=21$) using Jellyfish (Marçais & Kingsford, 2011) combined with GenomeScope (<http://qb.cshl.edu/genomescope/>) (Vurture et al., 2017).

2.5 Genotyping by sequencing

Genotyping by sequencing (GBS) (Elshire et al., 2011) was conducted by ecogenics GmbH (Switzerland). Individually MID-tagged reduced representation libraries were generated using the standard enzyme combination EcoRI/MseI and single-end reads of 75 bp were sequenced on an Illumina NextSeq instrument. After checking the quality of the raw reads with FASTQC version 0.10.1 (Andrew, 2019), leading and trailing low quality reads were removed using TRIMMOMATIC version 0.39 (Bolger et al., 2014). Trimmed reads were aligned to the collared sand martin draft genome using BWA MEM version 0.7.17 (Li, 2013). Single-nucleotide polymorphisms (SNPs) were called and genotyped with angsd (Korneliussen et al., 2014) based on GATK genotype likelihoods (McKenna et al., 2010) retaining sites with $p < 0.001$ for being variable, a minimum mapping quality of 20, a minimum base quality score of 20, a minimum total read depth of 300, a minimum individual read depth of five and minor allele frequency of 0.05 (-GL 2 -SNP_pval 1e-3 -minQ 20 -minMapQ 20 -setMinDepth 300 -geno_minDepth 5 -minMaf 0.05). We only retained uniquely mapped reads and biallelic SNPs with less than 10% missing data (-uniqueOnly 1 -skipTriallelic 1 -minInd 135).

2.6 Analyses of population genomic structure

To examine genetic structure among populations, we first conducted a principal component analysis (PCA) based on individual genotype likelihoods using PCAngsd (Meisner & Albrechtsen, 2018). This was done for *R. diluta* and *R. riparia* together and for *R. diluta* separately. The eigenvectors from the covariance matrix were computed using the function “eigen” in R 3.6.2. To check for potential gene flow between populations, particularly in putative contact zones, we also performed admixture analysis in NGSadmix (Skotte et al., 2013). NGSadmix was also based on genotype-likelihood, thus accounting for the uncertainty of called genotypes. It was run ten times on all individuals (*R. diluta* and *R. riparia* together) for each K and the number of ancestral populations K set from 2 to 10. The optimal K was evaluated using CLUMPAK (Kopelman et al., 2015).

We additionally did two Analyses of Molecular Variance (AMOVA) and computed pairwise F_{ST} between populations of *R. diluta* in Arlequin version 3.5.2.2 (Excoffier & Lischer, 2010) using SNPs with less than 5% missing data (default setting) and 1000 permutations. For AMOVA, the

three subspecies were defined as groups with the Mongolian population either included in *R. d. tibetana* or in *R. d. diluta*. The distance matrix was computed using pairwise distances. The input file for Arlequin was generated using PGDSpider version 2.1.1.5 (Lischer & Excoffier, 2012). To investigate the influence of geographical distance on population structure, we tested for isolation by distance between different populations of *R. diluta*. We applied Mantel tests (Mantel, 1967) to half matrices of genetic ($F_{ST}/(1-F_{ST})$) and logarithmic (ln) Euclidean geographic distances between populations using the R package ade4 (Dray & Dufour, 2007) with 999 Monte-Carlo permutations.

Nucleotide diversity (π) of each population was computed using angsd (Korneliussen et al., 2014). Filters for SNPs were again set to a minimum mapping quality of 20, a minimum base quality score of 20, minimum read depth of five for each individual, and less than 10% missing data. We only kept uniquely mapped reads for the estimation of posterior probabilities of sample allele frequency (SAF) for each population, and then computed the site frequency spectrum (SFS) using realSFS. Pairwise nucleotide diversity for each site was computed using thetaStat based on the SFS and the average was used as nucleotide diversity for each population. The unfolded SFS was estimated using the reference genome of *R. riparia* for the characterization of ancestral states in populations of *R. diluta*. For populations of *R. riparia*, the folded SFS was used.

2.7 Morphological analysis

To investigate the extent of morphological differences between subspecies of *R. diluta*, we collected mensural data of 190 individuals. Eighty-four of these were also included in the GBS analyses (complete data could not be obtained for the remaining 35 individuals used for GBS), subspecific identity of the others was based on breeding colony origin. Eight traits were measured following Eck et al. (2012): length of bill tip to feathering, bill depth, bill width, wing length, length of P8 (third outermost primary), tail length, length of tail fork and tarsus length (Table S3). In an additional analysis, we combined our dataset with the one of Schweizer et al. (2018) from which three morphometric traits (wing length, tail length and length of tail fork) for 120 individuals of *R. diluta* (32 of *R. d. diluta*, 19 of *R. d. tibetana*, 36 of *R. d. indica*, 33 of *R. d. fohkienensis*) were available stemming mainly from museum specimens. After log-transformation of measurements, morphometric differences among subspecies were explored using a principal

component analysis (PCA) on the correlation matrix using the function `prcomp` of the R package 'stats'.

2.8 Seasonal occurrence patterns

As *R. diluta* occurs in different climate zones and across a broad altitudinal range, we assessed differences in seasonal occurrence patterns (phenology) between different geographic regions corresponding to the supposed distribution of the three subspecies. To this end, we used data from our own field work and compiled records of *R. diluta* from the two citizen science databases ebird (<https://ebird.org/home>) and BirdReport of China (<http://www.birdrecord.cn/>). Each record was allocated to three periods in each month, i.e. before the 10th, between the 10th and the 20th, and after the 20th. The three geographic regions were defined as follows: 1) south and central China east of the Qinghai-Tibetan plateau below 3000 m above sea level (asl) corresponding to the supposed breeding distribution of *R. d. fohkienensis*, 2) Qinghai-Tibetan plateau above 3000 m asl corresponding to the supposed breeding range of *R. d. tibetana*, 3) the region of China north-west of the Qinghai-Tibetan plateau corresponding to the supposed breeding distribution of *R. d. diluta*.

3 Results

3.1 MtDNA Phylogeny

The final ND2 alignment was 851 bp in length with 32 haplotypes in *R. diluta* and 11 haplotypes in *R. riparia*. Within *R. diluta*, BI and ML phylogenetic reconstructions recovered well supported clades mostly consistent with the distribution ranges of the morphologically-defined subspecies (Figure 1, Figure S1): One clade consisted of all samples of *R. d. fohkienensis* from lowland south-eastern China, one of *R. d. tibetana* from the Qinghai-Tibetan Plateau and one of *R. d. diluta* from north-western China (Figure 1). Samples collected from Mongolia in the potential breeding range of *R. d. diluta*, however, clustered and shared haplotypes with samples of *R. d. tibetana* from the Qinghai-Tibetan plateau. Only one individual from Mongolia with a unique haplotype (Hap 30) did not cluster with the remaining samples from this region and *R. d. tibetana*, and was instead

found in the *R. d. diluta* clade. The positions of these different clades as well as that of *R. d. indica* were not robustly supported throughout. The monophyly of the haplotypes found in *R. riparia* was only supported in ML analyses, but not with BI.

3.2 Population genomics of nuclear variation

The heterozygosity of the collared sand martin reference genome was estimated to be 0.1%. In total, we obtained 7,640 autosomal SNPs in the dataset including *R. diluta* and *R. riparia* after aligning trimmed GBS reads to the collared sand martin draft genome. These SNPs were distributed across most autosomes except for the two microchromosomes 16 (1.22 Mb) and 29 (4.21 Mb), and there was a highly significant positive correlation ($r = 0.92$; $P = 5.02 \times 10^{-14}$) between the number of SNPs called per chromosome and chromosome size (Figure S2, S3). In a PCA on both species, *R. diluta* was clearly separated from *R. riparia* in PC1 (22.12 % of the variance), while PC2 (18.34 % of the variance) separated *R. d. fohkienensis* and the remaining samples of *R. diluta* (Figure 2A). It is noteworthy that all *R. diluta* samples from Mongolia clustered with *R. d. tibetana*. In a separate PCA with 7,118 autosomal SNPs called for the 119 *R. diluta* samples only (Figure 2B), *R. d. fohkienensis* was clearly differentiated from the remaining samples along PC1 (26.65% of variance). PC2 (10.95% of variance) separated north-western Chinese *R. d. diluta* from a cluster containing *R. tibetana* from the Qinghai-Tibetan plateau and all samples from Mongolia.

Admixture analyses resulted in $K=4$ as the best-fitting number of ancestral populations with a somewhat lower likelihood for $K=3$ (Figure S4). However, *R. riparia*, *R. d. fohkienensis* and north-western Chinese *R. d. diluta* were always resolved as distinct genetic clusters with separate ancestries (Figure 3, Figure S5). For $K=4$, *R. d. tibetana* from the Qinghai-Tibetan plateau and individuals from Mongolia formed one additional genetic cluster, with the latter showing limited evidence of mixed ancestry with north-western Chinese *R. d. diluta* (average 2.9%; range per individual: 0 to 6.2%).

Alternative AMOVAs and F-statistics based on 6,857 SNPs supported the closer affinity of Mongolian birds to *R. d. tibetana*. Overall differentiation between populations was very high with $F_{ST} = 0.225$ ($p < 0.0001$). When the Mongolian population was grouped in an AMOVA with *R. d. tibetana*, the total variation explained by the subspecies reached $F_{CT} = 0.216$ with very little

differentiation within these groups ($F_{SC}=0.011$, both $p < 0.0001$). When the Mongolian population was alternatively grouped with *R. d. diluta*, the proportion of the explained variation dropped to $F_{CT} = 0.186$ and differentiation within the groups of populations increased accordingly ($F_{SC}=0.045$; both p values < 0.0001). We thus consider the population in Mongolia to belong to *R. d. tibetana* outside the assumed distribution range of this subspecies. With this assignment, pairwise F_{ST} between populations within subspecies ranged from 0 to 0.051 (most $p < 0.05$; Table S4) while all pairwise comparisons between populations from different subspecies ranged between $F_{ST} = 0.1$ and 0.36 (all $p < 0.05$; Table S4).

Comparison of molecular diversity showed further that the Mongolian birds featured the highest nucleotide diversity among all analysed *R. diluta* populations ($\pi = 5.9 \times 10^{-3}$). Nucleotide diversity was overall similar across *R. d. diluta* and *R. d. tibetana* populations from the Qinghai-Tibetan plateau with no detectable relation to longitude or latitude (π ranging from 3.0×10^{-3} to 4.4×10^{-3} ; Figure 4). All populations of *R. d. fohkienensis* showed comparatively lower levels of nucleotide diversity (π ranging from 1.9×10^{-3} to 2.3×10^{-3}). The highest values of nucleotide diversity were found in the three analysed populations of *R. riparia* (π ranging from 6.6×10^{-3} to 7.4×10^{-3}).

3.3 Isolation by distance

Given distances of thousands of kilometres between the analysed populations, we tested for the importance of isolation by distance in genetic differentiation between and within subspecies of *R. diluta*. Overall, a Mantel test detected a highly significant relationship between genetic differentiation and spatial distance between populations ($r = 0.5486$; $p = 0.001$; Figure 5). Closer inspection revealed that this relationship was mainly driven by very high pairwise comparisons between subspecies across their parapatric distribution ranges. In contrast, comparisons among large spatial distances within subspecies, especially in *R. d. tibetana*, showed no evidence of elevated genetic differentiation (Figure 5). Accordingly, a Mantel test restricted to pairwise comparisons within subspecies provided no evidence of isolation by distance ($r = -0.304$; $p = 0.87$), indicating considerable dispersal among populations over large distances.

3.4 Morphological differentiation

A PCA on eight morphological traits measured on the novel *R. diluta* samples presented here revealed gradual differences between subspecies with some overlap (Figure 6A). PC1 (24.01% variance) separated largely *R. d. fohkienensis* from *R. d. tibetana* and was mainly influenced by length of P8, wing length and tail length (loading factors of 0.507, 0.498 and 0.452 respectively, Table S5) while PC 2 mainly distinguished *R. d. fohkienensis* from *R. d. diluta* and was dominated by the length of tail fork (loading factor of 0.668, Table S5). Birds from Mongolia clustered among individuals of *R. tibetana* with similarly gradual transitions to the other subspecies (Figure 6A).

The analysis of three morphological traits only, which enabled the inclusion of additional 120 individuals (including *R. d. indica*) from Schweizer et al. (2018), revealed overall less separation between different subspecies of *R. diluta* (Figure 6B). PC1 (55.19% of variance) separated *R. d. fohkienensis* and *R. d. indica* together from *R. d. diluta* and *R. d. tibetana* in PC1), while PC2 (30.43% of variance) tended to differentiate subspecies of these two pairs. However, there was gradual overlap overall.

3.5 Differences in phenology

A total of 1253 records (545 from BirdReport of China and 708 from ebird) of pale sand martin from China were compiled for the years 1985 - 2021. Pale sand martins were found throughout the year but with distinct differences between the regions occupied by the three subspecies. In north-western China in the breeding area of *R. d. diluta*, the first birds appeared in late April, peaks were revealed in May and July, and there were no records from September onwards (Figure 7A). A similar pattern of occurrence was revealed on the Qinghai-Tibetan plateau above 3000 m asl in the breeding range of *R. d. tibetana* (Figure 7B). The first birds were recorded in mid-April with a broad peak of records around mid-July and mid-August and no records after mid-October. In south and central China below 3000 m asl however, records were found throughout the year with most records of pale sand martin stemming from the winter months with a reduction during the summer months (Figure 7C). Although this area corresponds to the traditional breeding range of *R. d. fohkienensis*, migrant and wintering individuals of other subspecies are certainly included in these records. According to our field observations, *R. d. fohkienensis* in the area was breeding already in late April with chicks found in holes until late

May (cf. Table S1) while *R. d. tibetana* on the Qinghai-Tibetan plateau started to build nest-holes only in the beginning of June.

4 Discussion

Here, by using population genomic data, we show that the pale sand martin *Riparia diluta* contains multiple deep evolutionary lineages despite extremely subtle and gradual morphological variation among them and little genetic differences within lineages over large geographic areas. Differentiation patterns in nuclear SNPs are largely consistent with phylogeographic structure in mtDNA (see also Schweizer et al. 2018) and represent different levels. We discuss how the integrity of the evolutionary lineages could be maintained despite morphological stasis. We hypothesize that prezygotic isolation in terms of allochrony and extrinsic postzygotic isolation caused by contrasting migration directions might prevent lineage fusion.

4.1 Phylogeographic structure

Evolutionary lineages within pale sand martin *Riparia diluta* are largely consistent with geographical distribution and at least partly with taxonomic classification of subspecies. *R. d. fohkienensis* has diverged most, and there were no signs of admixture between less divergent *R. d. tibetana* and *R. d. diluta* in the region of presumed contact at the western edge of the Qinghai-Tibetan plateau. Individuals of *R. d. tibetana* in Mongolia however, displayed signs of limited mixed ancestries with *R. d. diluta* from north-western China indicating the absence of full reproductive isolation between these subspecies and potentially recent admixture.

The strong differentiation in genetically distinct but morphologically cryptic lineages within *R. diluta* is in stark contrast with the lack of any phylogeographic structure in its sister species, the collared sand martin *R. riparia*, in our study region. Geographically widespread nuclear homogeneity in *R. riparia* is in agreement with shallow mtDNA diversity over its entire Holarctic breeding range indicating recent demographic expansion (Pavlova et al., 2008; Schweizer et al., 2018). In contrast, extensive genetic structure in *R. diluta* is more similar to other Central and East Asian bird species complexes and has been related to heterogeneous environments and/or a climate being only mildly affected by Pleistocene climate changes that enabled the persistence of

isolated populations in mountainous region at the south-western edge of the Qinghai-Tibetan plateau (e.g. Qu et al. 2014, Liu et al., 2016 Liu et al., 2020). Varying nucleotide diversity among different populations of *R. diluta* indicates contrasting recent demographic histories, which need to be further investigated.

The distribution of *R. d. tibetana* is usually considered to be restricted to the Qinghai-Tibetan plateau (e.g. del Hoyo & Collar 2016), but we showed here with genome-wide data that this evolutionary lineage extends into central Mongolia. This is consistent with cases of shared mtDNA haplotypes between birds from Mongolia and Russia with *R. d. tibetana* from the Qinghai-Tibetan plateau (Schweizer et al. 2018). It remains to be examined how much the range of *R. d. tibetana* extends farther to the north and west from central Mongolia.

4.2 Cryptic diversification with different levels of reproductive isolation

We detected no evidence of gene flow in the potential areas of contact between lowland *R. d. fohkienensis* and *R. d. tibetana* from the Qinghai-Tibetan plateau. Allochrony – differences in the timing of breeding - in combination with ecological divergence may play an important role in the prevention of hybridization between them. Data on seasonal occurrence patterns and our own observations suggest that breeding of *R. d. fohkienensis* in the lowlands of central and south China takes place considerably earlier than that of *R. d. tibetana* on the Qinghai-Tibetan plateau. Hence, their different phenologies - likely connected to differing migration behaviour with *R. d. fohkienensis* probably only conducting short-distance movements unlike *R. d. tibetana* (see below) - might act as prezygotic isolation mechanisms. Speciation through allochrony, i.e. temporal segregation of breeding populations as an important contributor to reproductive isolation, has been invoked in several animal groups, including birds (e.g. Bearhop et al., 2005; Friesen et al., 2007; Taylor & Friesen, 2017; Taylor et al., 2018, 2019; Gómez-Bahamón et al., 2020; Kimmitt et al., 2019; Sirkiä et al., 2018). However, other potential prezygotic isolation mechanisms such as differences in mating behaviour between *R. d. fohkienensis* *R. d. tibetana* remain to be investigated. Moreover, the involvement of additional factors cannot be excluded, especially ecological differences as *R. d. fohkienensis* and *R. d. tibetana* might have adapted to different climate regimes. Comparatively fast rates of climate-niche evolution could be expected in the temporally and spatially heterogeneous climate in the region of the Qinghai-Tibetan plateau and central and south China (cf. Lawson & Weir, 2014).

However, we cannot exclude the existence of an undetected contact zone with hybridization between *R. d. fohkienensis* and *R. d. tibetana* at the edge of the plateau given that the closest sampled populations were 512 km apart. Hybrid zones between bird species may be considerably narrower, but in bird taxa with recent divergence, relatively wide hybrid zones of >100 km have also been found (Price, 2008). Hybrid zone width is, among other factors, strongly influenced by dispersal distance (Barton & Hewitt, 1985; McEntee et al., 2020). This has not been studied in *R. diluta*, rendering it difficult to make predictions about expected potential hybrid zone width. In *R. riparia*, its sister species, however, 7% of juveniles were found >199 km away from their natal colonies in consecutive years in Britain (Mead, 1979) indicating comparatively high colonisation potential (Tittler et al., 2009). Considerable dispersal between colonies over large distances is also indicated in *R. diluta* by low genetic differentiation and a lack of patterns of isolation by distance within evolutionary lineages. Finer-scale sampling of the potential contact area would be necessary to specifically test for interbreeding. Given the absence of any traces of admixture in the sampled populations (e.g. Figure 3), large-scale gene flow between *R. d. fohkienensis* and *R. d. tibetana* has probably ceased comparatively long ago given that divergence of *R. d. fohkienensis* has been estimated to have occurred about 1.2 million years ago (Schweizer et al., 2018).

Our results clearly show closer relationships between *R. d. diluta* and *R. d. tibetana* and thus the possibility of hybridization between them may not be too unexpected. It is probably more surprising that the signals of limited autosomal introgression were not detected at the north-western edge of the Qinghai Tibetan plateau (cf. Figure 1), but rather into Mongolian *R. d. tibetana*. The comparatively high levels of nucleotide diversity in the Mongolian population of *R. d. tibetana* might also be a consequence of admixture with *R. d. diluta*. Further sampling in western Mongolia would be needed to determine the extent and spatial structure of hybridization between the two.

The mountains of Central Mongolia between north-western Chinese populations of *R. d. diluta* and Mongolian *R. d. tibetana* have been identified as potentially limiting extensive hybridization between two subspecies of barn swallow *Hirundo rustica* which show contrasting migration routes around the Qinghai-Tibetan plateau (Scordato et al., 2020). The area where the lineages of *R. d. diluta* and *R. d. tibetana* might be in contact in Mongolia is indeed located in a migratory

divide in different bird species complexes (Irwin & Irwin, 2005; Scordato et al., 2020). The Qinghai-Tibetan plateau has been proposed as a major barrier to bird migration and a majority of migrant Siberian species use just one migratory route – east or west – around it or show different routes in different subspecies (Irwin & Irwin, 2005). *R. d. diluta* has been recorded to winter in the north-western part of the Indian Subcontinent and rarely on the Arabian peninsula, and likely takes a western route around the Qinghai-Tibetan plateau with migration documented in north-western South Asia (Rasmussen et al., 2005, Shirihai & Svensson 2018). By contrast, *R. d. tibetana* might winter in Southeast Asia and circumnavigate the Qinghai-Tibetan plateau in the east or follow river valleys at its south-eastern edge and winter in the northern, central and north-eastern Indian Subcontinent (own data; Rasmussen et al., 2005). As shown for several songbird species, migration direction may have a genetic basis and thus hybrids may have non-optimal intermediate migration routes and might be selected against (Berthold et al., 1992; Delmore et al., 2015; Delmore & Irwin, 2014; Helbig, 1991, 1996; Lundberg et al., 2017). In combination with geographical barriers, such a mechanism might prevent lineage fusion between *R. d. diluta* and *R. d. tibetana* in Mongolia. Secondary contact of populations with different migration directions could trigger the evolution of prezygotic isolation mechanisms (Scordato et al., 2020). Given the morphological cypsis of the lineages in *R. diluta*, we hypothesize that the secondary contact might be too recent and/or admixture not extensive enough for this process to play a role. However, mating behaviour of the evolutionary lineages remains to be thoroughly investigated to reveal potentially hitherto undocumented differences.

4.3 Conclusion

Genome-wide differentiation between the morphologically cryptic lineages of the pale sand martin *Riparia diluta* represents different levels, and we hypothesize that they might differ in their strength of reproductive isolation. While genomic differentiation of *R. d. fohkienensis* suggests effective reproductive isolation for a comparatively long time period, footprints of introgression were found between *R. d. diluta* and *R. d. tibetana*. In the absence of obvious sexually selected traits, seasonal migration behaviour might be an essential factor in maintaining genetic integrity of these morphologically cryptic evolutionary lineages.

Seasonal migration behaviour has for a long time been considered as playing an important role in generating and maintaining evolutionary divergence (reviewed in Turbek et al., 2018). It has been hypothesized that fusion of different evolutionary lineages might be prevented by contrasting migration behavior even with comparatively little differentiation in other traits (Delmore et al., 2015). Here we add an additional aspect to this: seasonal migration behaviour might be an essential mechanism to maintain evolutionary diversity under morphological stasis.

Acknowledgements

We thank Qin Huang, Xinyuan Pan, Dan Liang, Yun Livin Lin, Xia Zhan, Wenjie Cheng, Paul Walser Schwyzer and Chentao Wei who assisted with fieldwork in China. Sarangua Bayrgerel (National University of Mongolia), Turmunbaatar Damba, Tuvshin Unenbat (Mongolian Ornithological Society), Paul Walser Schwyzer and Silvia Zumbach were of invaluable help with fieldwork in Mongolia and Susanne Tellenbach for laboratory work. We acknowledge support from the National Genomics Infrastructure in Stockholm funded by the Science for Life Laboratory, the Knut and Alice Wallenberg Foundation and the Swedish Research Council, specifically, Fanny Taborsak-Lines for the preparation of linked-read sequencing libraries. We acknowledge the financial support of Oversea Study Program of Guangzhou Elite Project (No. JY201726) and of Basler Stiftung für biologische Forschung. We moreover thank the Cornell Lab of Ornithology and BirdReport of China for their support with data on records of *R. diluta*. The comments and suggestions by Darren Irwin, three anonymous reviewers as well as the editors considerably improved the manuscript.

References

- Aljanabi, S. M., & Martinez, I. (1997). Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*, 25(22), 4692–4693.
- Andrew, S. (2019). *FastQC: A quality control tool for high throughput sequence data 2010* [Available from: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>]. Accessed.
- Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, 16(1), 113–148.

- Bearhop, S., Fiedler, W., Furness, R. W., Votier, S. C., Waldron, S., Newton, J., Bowen, G. J., Berthold, P., & Farnsworth, K. (2005). Assortative mating as a mechanism for rapid evolution of a migratory divide. *Science*, *310*(5747), 502–504.
- Berthold, P., Helbig, A. J., Mohr, G., & Querner, U. (1992). Rapid microevolution of migratory behaviour in a wild bird species. *Nature*, *360*(6405), 668–670.
- Beysard, M., & Heckel, G. (2014). Structure and dynamics of hybrid zones at different stages of speciation in the common vole (*Microtus arvalis*). *Molecular Ecology*, *23*(3), 673–687.
- Beysard, M., Krebs-Wheaton, R., & Heckel, G. (2015). Tracing reinforcement through asymmetrical partner preference in the European common vole *Microtus arvalis*. *BMC Evolutionary Biology*, *15*(1), 1–8.
- BirdLife International & Handbook of the Birds of the World. (2016). Bird species distribution maps of the world (version 6.0). Retrieved from <http://datazone.birdlife.org/>.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.
- Braune, P., Schmidt, S., & Zimmermann, E. (2008). Acoustic divergence in the communication of cryptic species of nocturnal primates (*Microcebus ssp.*). *BMC Biology*, *6*(1), 19. <https://doi.org/10.1186/1741-7007-6-19>
- Burland, T. G. (2000). DNASTAR's Lasergene sequence analysis software. In *Bioinformatics methods and protocols* (pp. 71–91). Springer.
- Cronemberger, Á. A., Aleixo, A., Mikkelsen, E. K., & Weir, J. T. (2020). Postzygotic isolation drives genomic speciation between highly cryptic *Hypocnemis* antbirds from Amazonia. *Evolution*, evo.14103. <https://doi.org/10.1111/evo.14103>
- del Hoyo J, Collar NJ. (2016). *HBW and BirdLife International illustrated checklist of the birds of the world: Vol. vol 2: Passerines*. Lynx Edicions, Barcelona.
- Delmore, K. E., Hübner, S., Kane, N. C., Schuster, R., Andrew, R. L., Câmara, F., Guigó, R., & Irwin, D. E. (2015). Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. *Molecular Ecology*, *24*(8), 1873–1888.
- Delmore, K. E., & Irwin, D. E. (2014). Hybrid songbirds employ intermediate routes in a migratory divide. *Ecology Letters*, *17*(10), 1211–1218.

- Delmore, K. E., Kenyon, H. L., Germain, R. R., & Irwin, D. E. (2015). Phenotypic divergence during speciation is inversely associated with differences in seasonal migration. *Proceedings of the Royal Society B: Biological Sciences*, 282(1819), 20151921.
- Dray, S., & Dufour, A.-B. (2007). The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software*, 22(4), 1–20.
- Dufresnes, C., Strachinis, I., Suriadna, N., Mykytynets, G., Cogălniceanu, D., Székely, P., Vukov, T., Arntzen, J. W., Wielstra, B., & Lymberakis, P. (2019). Phylogeography of a cryptic speciation continuum in Eurasian spadefoot toads (*Pelobates*). *Molecular Ecology*, 28(13), 3257–3270.
- Eck, S., Fiebig, J., Fiedler, W., Heynen, I., Nicolai, B., Töpfer, T., van den Elzen, R., Winkler, R., & Woog, F. (2012). *Measuring Birds—Vögel vermessen*.
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One*, 6(5).
- Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10(3), 564–567.
- Feckler, A., Zubrod, J. P., Thielsch, A., Schwenk, K., Schulz, R., & Bundschuh, M. (2014). Cryptic species diversity: An overlooked factor in environmental management? *Journal of Applied Ecology*, 51(4), 958–967. <https://doi.org/10.1111/1365-2664.12246>
- Fink, S., Fischer, M. C., Excoffier, L., & Heckel, G. (2010). Genomic scans support repetitive continental colonization events during the rapid radiation of voles (Rodentia: *Microtus*): the utility of AFLPs versus mitochondrial and nuclear sequence markers. *Systematic Biology*, 59(5), 548–572.
- Fišer, C., Robinson, C. T., & Malard, F. (2018). Cryptic species as a window into the paradigm shift of the species concept. *Molecular Ecology*, 27(3), 613–635.
- Friesen, V. L., Smith, A. L., Gomez-Diaz, E., Bolton, M., Furness, R. W., González-Solís, J., & Monteiro, L. R. (2007). Sympatric speciation by allochrony in a seabird. *Proceedings of the National Academy of Sciences*, 104(47), 18589–18594.

- Gavrilov, E. I., & Savchenko, A. P. (1991). On species validity of the Pale Sand Martin (*Riparia diluta* Sharpe et Wyatt, 1893). *Byul. Mosk. O-va Isp. Prir., Otd. Biol*, 96(4), 34–44.
- Gómez-Bahamón, V., Márquez, R., Jahn, A. E., Miyaki, C. Y., Tuero, D. T., Laverde-R, O., Restrepo, S., & Cadena, C. D. (2020). Speciation associated with shifts in migratory behavior in an avian radiation. *Current Biology*, 30(7), 1312-1321.e6. <https://doi.org/10.1016/j.cub.2020.01.064>
- Goroshko, O. A. (1993). Taxonomic status of the pale (sand?) martin *Riparia (riparia?) diluta* (Sharpe & Wyatt, 1893). *Russ. Ornith. Zhurnal*, 2(3), 303–323.
- Grant, P. R., & Grant, B. R. (2011). *How and why species multiply: The radiation of Darwin's finches*. Princeton University Press.
- Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, 52(5), 696–704.
- Helbig, A. (1996). Genetic basis, mode of inheritance and evolutionary changes of migratory directions in palaeartic warblers (Aves: Sylviidae). *Journal of Experimental Biology*, 199(1), 49–55.
- Helbig, A. J. (1991). Inheritance of migratory direction in a bird species: A cross-breeding experiment with SE-and SW-migrating blackcaps (*Sylvia atricapilla*). *Behavioral Ecology and Sociobiology*, 28(1), 9–12.
- Hillis, D. M., & Bull, J.J. 1993. An empirical-test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology* 42:182–192.
- Huelsenbeck, J. P., & Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755.
- Irwin, D. E. (2020). Assortative mating in hybrid zones is remarkably ineffective in promoting speciation. *The American Naturalist*, 195(6), E150–E167.
- Irwin, D. E., & Irwin, J. H. (2005). Siberian migratory divides: The role of seasonal migration in speciation. *Birds of Two Worlds: The Ecology and Evolution of Migration*, 27–40.
- Kearns, A. M., Restani, M., Szabo, I., Schrøder-Nielsen, A., Kim, J. A., Richardson, H. M., Marzluff, J. M., Fleischer, R. C., Johnsen, A., & Omland, K. E. (2018). Genomic evidence of speciation reversal in ravens. *Nature Communications*, 9(1), 1–13.

- Kimmit, A. A., Hardman, J. W., Stricker, C. A., & Ketterson, E. D. (2019). Migratory strategy explains differences in timing of female reproductive development in seasonally sympatric songbirds. *Functional Ecology*, *33*(9), 1651–1662. <https://doi.org/10.1111/1365-2435.13386>
- Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, *15*(5), 1179–1191.
- Korlach, J., Gedman, G., Kingan, S. B., Chin, C.-S., Howard, J. T., Audet, J.-N., Cantin, L., & Jarvis, E. D. (2017). De novo PacBio long-read and phased avian genome assemblies correct and add to reference genes generated with intermediate and short reads. *GigaScience*, *6*(10). <https://doi.org/10.1093/gigascience/gix085>
- Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics*, *15*(1), 356.
- Kozak, K. H., Weisrock, D. W., & Larson, A. (2006). Rapid lineage accumulation in a non-adaptive radiation: Phylogenetic analysis of diversification rates in eastern North American woodland salamanders (Plethodontidae: *Plethodon*). *Proceedings of the Royal Society B: Biological Sciences*, *273*(1586), 539–546.
- Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, *33*(7), 1870–1874.
- Lawson, A. M., & Weir, J. T. (2014). Latitudinal gradients in climatic-niche evolution accelerate trait evolution at high latitudes. *Ecology Letters*, *17*(11), 1427–1436.
- Leys, M., Keller, I., Robinson, C. T., & Räsänen, K. (2017). Cryptic lineages of a common alpine mayfly show strong life-history divergence. *Molecular Ecology*, *26*(6), 1670–1686.
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv Preprint ArXiv:1303.3997*.
- Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094–3100.
- Lischer, H. E., & Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, *28*(2), 298–299.

- Liu, S., Liu, Y., Jelen, E., Alibadian, M., Yao, C.-T., Li, X., Kayvanfar, N., Wang, Y., Vahidi, F. S., Han, J.-L., Sundev, G., Zhang, Z. & Schweizer, M. (2020). Regional drivers of diversification in the late Quaternary in a widely distributed generalist species, the common pheasant *Phasianus colchicus*. *Journal of Biogeography*.
- Lundberg, M., Liedvogel, M., Larson, K., Sigeman, H., Grahn, M., Wright, A., Åkesson, S., & Bensch, S. (2017). Genetic differences between willow warbler migratory phenotypes are few and cluster in large haplotype blocks. *Evolution Letters*, 1(3), 155–168.
- Lutgen, D., Ritter, R., Olsen, R.-A., Schielzeth, H., Gruselius, J., Ewels, P., García, J. T., Shirihi, H., Schweizer, M., & Suh, A. (2020). Linked-read sequencing enables haplotype-resolved resequencing at population scale. *Molecular Ecology Resources*.
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27(2 Part 1), 209–220.
- Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, 27(6), 764–770. <https://doi.org/10.1093/bioinformatics/btr011>
- McEntee, J. P., Burleigh, J. G., & Singhal, S. (2020). Dispersal predicts hybrid zone widths across animal diversity: Implications for species borders under incomplete reproductive isolation. *The American Naturalist*, 196(1), 9–28. <https://doi.org/10.1086/709109>
- McGee, M. D., Borstein, S. R., Meier, J. I., Marques, D. A., Mwaiko, S., Taabu, A., Kishe, M. A., O'Meara, B., Bruggmann, R., & Excoffier, L. (2020). The ecological and genomic basis of explosive adaptive radiation. *Nature*, 1–5.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., & Daly, M. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297–1303.
- Mead, C. J. (1979). Colony fidelity and interchange in the sand martin. *Bird Study*, 26(2), 99–106.
- Meisner, J., & Albrechtsen, A. (2018). Inferring population structure and admixture proportions in low-depth NGS data. *Genetics*, 210(2), 719–731.
- Nosil, P. (2012). *Ecological speciation*. Oxford University Press.

- Pavlova, A., Zink, R. M., Drovetski, S. V., & Rohwer, S. (2008). Pleistocene evolution of closely related sand martins *Riparia riparia* and *R. diluta*. *Molecular Phylogenetics and Evolution*, *48*(1), 61–73.
- Price, T. (2008). *Speciation in birds*. Roberts and Company.
- Price, T. D., & Bouvier, M. M. (2002). The evolution of F1 postzygotic incompatibilities in birds. *Evolution*, *56*(10), 2083–2089.
- Pulido-Santacruz, P., Aleixo, A., & Weir, J. T. (2018). Morphologically cryptic Amazonian bird species pairs exhibit strong postzygotic reproductive isolation. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1874), 20172081.
- Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, *67*(5), 901.
- Rasmussen, P. C., Anderton, J. C., & Edicions, L. (2005). *Birds of south Asia: The Ripley guide*. Washington, DC.
- Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M. A., & Huelsenbeck, J. P. (2012). MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, *61*(3), 539–542.
- Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, *34*(12), 3299–3302.
- Rundell, R. J., & Price, T. D. (2009). Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. *Trends in Ecology & Evolution*, *24*(7), 394–399.
- Schluter, D. (2000). *The ecology of adaptive radiation*. OUP Oxford.
- Schweizer, M., Liu, Y., Olsson, U., Shirihi, H., Huang, Q., Leader, P. J., Copete, J. L., Kirwan, G. M., Chen, G. L., & Svensson, L. (2018). Contrasting patterns of diversification in two sister species of martins (Aves: Hirundinidae): The sand martin *Riparia riparia* and the pale martin *R. diluta*. *Molecular Phylogenetics And Evolution*, *125*, 116–126. <https://doi.org/10.1016/j.ympev.2018.02.026>
- Schweizer, Manuel, & Aye, R. (2007). Identification of the pale sand martin *Riparia diluta* in Central Asia. *Alula*, *4*, 152–158.

- Scordato, E. S., Smith, C. C., Semenov, G. A., Liu, Y., Wilkins, M. R., Liang, W., Rubtsov, A., Sundev, G., Koyama, K., & Turbek, S. P. (2020). Migratory divides coincide with reproductive barriers across replicated avian hybrid zones above the Tibetan Plateau. *Ecology Letters*, *23*(2), 231–241.
- Seehausen, O. (2006). African cichlid fish: A model system in adaptive radiation research. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1597), 1987–1998.
- Seehausen, O. L. E., Takimoto, G., Roy, D., & Jokela, J. (2008). Speciation reversal and biodiversity dynamics with hybridization in changing environments. *Molecular Ecology*, *17*(1), 30–44.
- Shirihai, H., & Svensson, L. (2018). *Handbook of Western Palearctic Birds*. A. & C. Black.
- Sirkiä, P. M., McFarlane, S. E., Jones, W., Wheatcroft, D., Aalund, M., Rybinski, J., & Qvarnström, A. (2018). Climate-driven build-up of temporal isolation within a recently formed avian hybrid zone. *Evolution*, *72*(2), 363–374.
- Skotte, L., Korneliussen, T. S., & Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. *Genetics*, *195*(3), 693–702.
- Slager, D. L., Epperly, K. L., Ha, R. R., Rohwer, S., Wood, C., Van Hemert, C., & Klicka, J. (2020). Cryptic and extensive hybridization between ancient lineages of American crows. *Molecular Ecology*, *29*(5), 956–969.
- Slavenko, A., Tamar, K., Tallowin, O. J., Allison, A., Kraus, F., Carranza, S., & Meiri, S. (2020). Cryptic diversity and non-adaptive radiation of montane New Guinea skinks (*Papuascincus*; Scincidae). *Molecular Phylogenetics and Evolution*, *146*, 106749.
- Struck, T. H., Feder, J. L., Bendiksbj, M., Birkeland, S., Cerca, J., Gusarov, V. I., Kistenich, S., Larsson, K.-H., Liow, L. H., & Nowak, M. D. (2018). Finding evolutionary processes hidden in cryptic species. *Trends in Ecology & Evolution*, *33*(3), 153–163.
- Taylor, R. S., & Friesen, V. L. (2017). The role of allochrony in speciation. *Molecular Ecology*, *26*(13), 3330–3342. <https://doi.org/10.1111/mec.14126>
- Taylor, Rebecca S., Bailie, A., Gulavita, P., Birt, T., Aarvak, T., Anker-Nilssen, T., Barton, D. C., Lindquist, K., Bedolla-Guzmán, Y., Quillfeldt, P., & Friesen, V. L. (2018). Sympatric population divergence within a highly pelagic seabird species complex (*Hydrobates spp.*). *Journal of Avian Biology*, *49*(1). <https://doi.org/10.1111/jav.01515>

- Taylor, Rebecca S., Bolton, M., Beard, A., Birt, T., Deane-Coe, P., Raine, A. F., González-Solís, J., Loughheed, S. C., & Friesen, V. L. (2019). Cryptic species and independent origins of allochronic populations within a seabird species complex (*Hydrobates spp.*). *Molecular Phylogenetics and Evolution*, *139*, 106552. <https://doi.org/10.1016/j.ympev.2019.106552>
- Taylor, Rebecca S., & Friesen, V. L. (2017). The role of allochrony in speciation. *Molecular Ecology*, *26*(13), 3330–3342.
- Team, R. C. (2013). *R: A language and environment for statistical computing*.
- Tittler, R., Villard, M.-A., & Fahrig, L. (2009). How far do songbirds disperse? *Ecography*, *32*(6), 1051–1061.
- Tobias, J. A., Ottenburghs, J., & Pigot, A. L. (2020). Avian diversity: Speciation, macroevolution, and ecological function. *Annual Review of Ecology, Evolution, and Systematics*, *51*(1), 533–560. <https://doi.org/10.1146/annurev-ecolsys-110218-025023>
- Turbek, S. P., Scordato, E. S., & Safran, R. J. (2018). The role of seasonal migration in population divergence and reproductive isolation. *Trends in Ecology & Evolution*, *33*(3), 164–175.
- Vurture, G. W., Sedlazeck, F. J., Nattestad, M., Underwood, C. J., Fang, H., Gurtowski, J., & Schatz, M. C. (2017). GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics*, *33*(14), 2202–2204. <https://doi.org/10.1093/bioinformatics/btx153>
- Weir, J. T., Haddrath, O., Robertson, H. A., Colbourne, R. M., & Baker, A. J. (2016). Explosive ice age diversification of kiwi. *Proceedings of the National Academy of Sciences*, *113*(38), E5580–E5587.
- Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M., & Jaffe, D. B. (2017). Direct determination of diploid genome sequences. *Genome Research*, *27*(5), 757–767.
- Yoder, J. B., Clancey, E., Des Roches, S., Eastman, J. M., Gentry, L., Godsoe, W., Hagey, T. J., Jochimsen, D., Oswald, B. P., & Robertson, J. (2010). Ecological opportunity and the origin of adaptive radiations. *Journal of Evolutionary Biology*, *23*(8), 1581–1596.

Data Accessibility

The data supporting this study are openly available on GenBank under accessions MZ747656–MZ747697 for *ND2* haplotypes, as well as on the NCBI sequence read archive (SRA) under BioProject PRJNA755835 with accession numbers SRR15558367–SRR15558515 and BioSample numbers SAMN20856160–SAMN20856308 for raw individual GBS sequences, and accession number JAIXNV000000000 and BioSample number SAMN20845799 and for the genome assembly.

Author Contribution

M.S. and G.H. conceived the study; Q.T., M.S., G.S. collected samples with assistance from Y.L.. Q.T. conducted laboratory work and analysed the data together with G.H., M.S. and assistance from R.B.; A.S. and R.B. contributed materials; Q.T., M.S. and G.H wrote the manuscript with support from all authors.

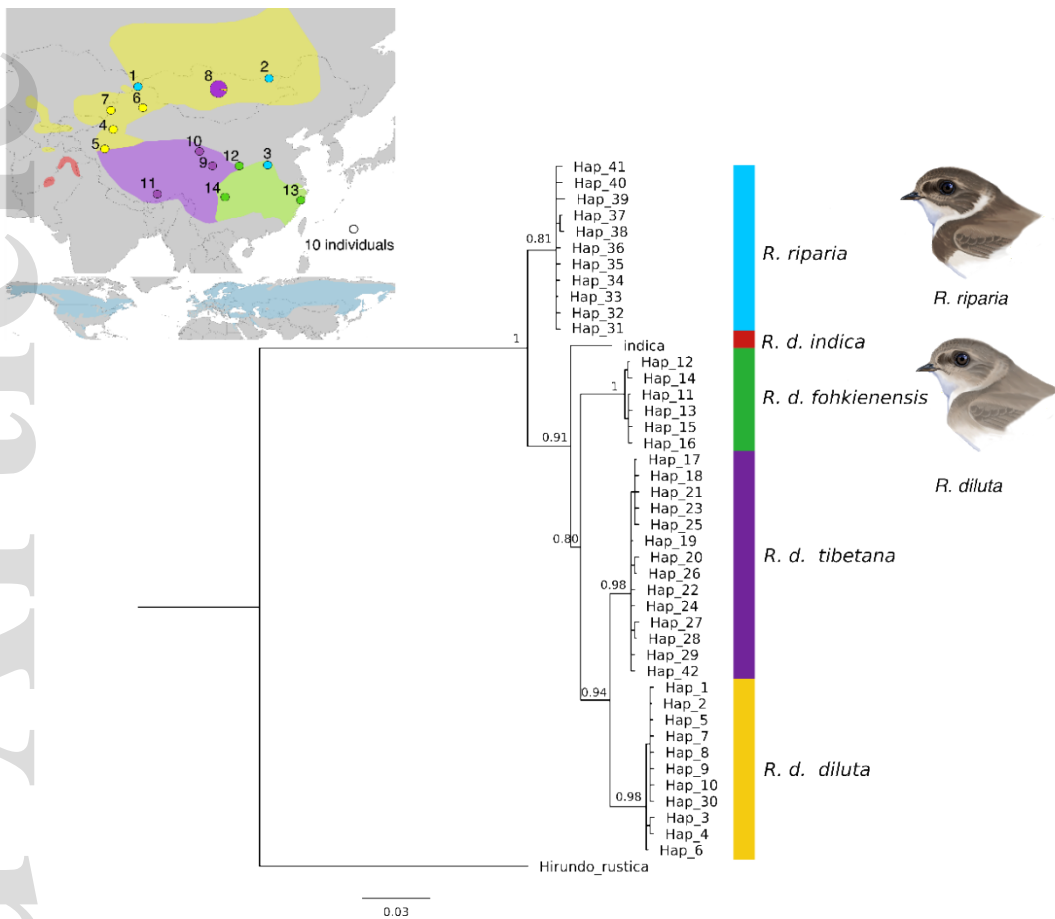


Figure 1. Bayesian majority rule consensus tree based on haplotypes of the mtDNA gene NADH dehydrogenase subunit II (ND2) of *R. diluta* and *R. riparia*. Bayesian posterior probabilities are given for major nodes. Color shades indicate the potential breeding ranges of different subspecies of *R. diluta* (top) and of *R. riparia* (bottom) modified from Bird Life International and Handbook of the Birds of the World (2016). Sampled populations are shown with dots on the maps with colors corresponding to mtDNA clades in the phylogenetic tree. *R. diluta* samples collected in Mongolia clustered with *R. d. tibetana* except one individual that was found in the clade of *R. d. diluta*. Paintings by Manuel Schweizer.

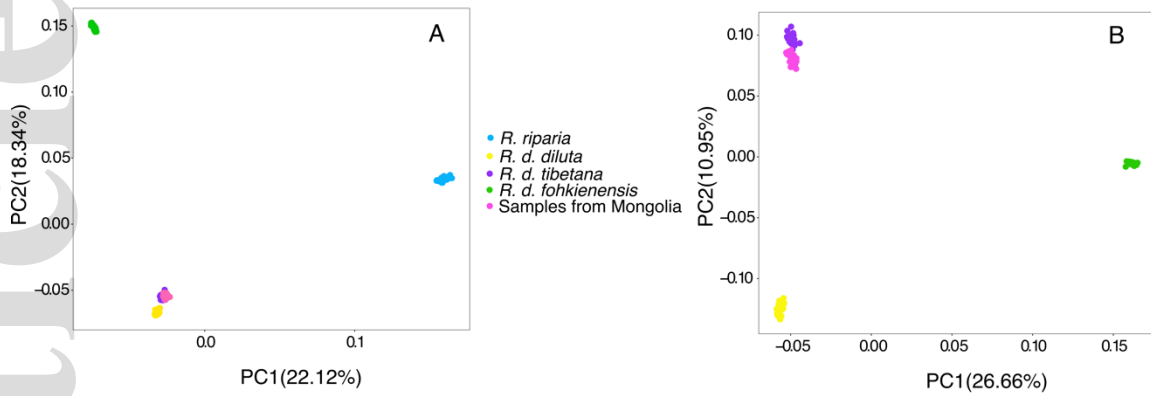


Figure 2. Principal component analysis (PCA) on nuclear SNPs of *R. diluta* and *R. riparia* (A, based on 7640 SNPs) and only of *R. diluta* (B, based on 7118 SNPs). Colors correspond to different subspecies except pink dots indicate the samples collected from Mongolia in the potential breeding range of *R. d. diluta*.

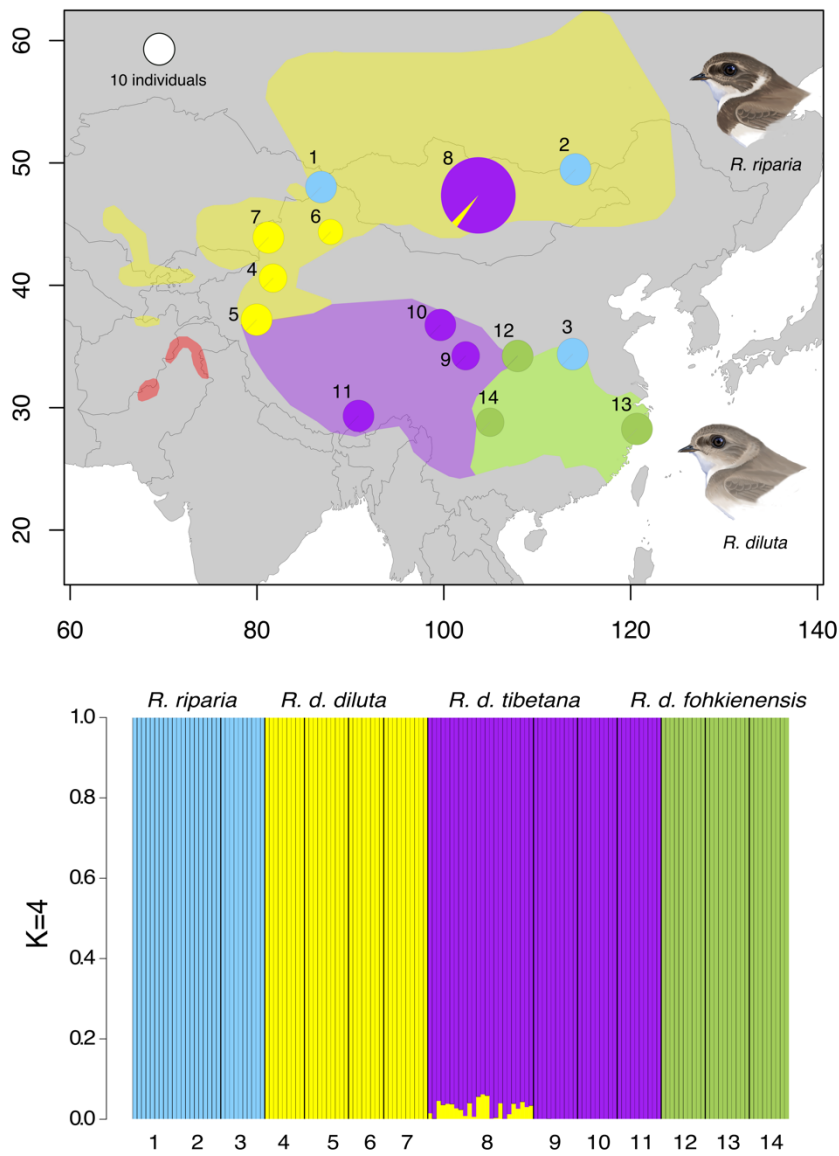


Figure 3. Top: Map with pie charts representing genome-wide ancestry assignment in *Riparia* populations with $K=4$ using NgsAdmix. Limited admixture between *R. d. diluta* and *R. d. tibetana* was found in the population from Mongolia (8) with $K=4$. Color shades indicate the potential breeding ranges of different subspecies of *R. diluta*. Pie sizes are proportional to sample sizes. Bottom: bar plots showing individual ancestry assignments with $K=4$. Numbers correspond to the different populations on the map. No individuals from the range of *R. d. indica* indicated in red on the map could be sampled.

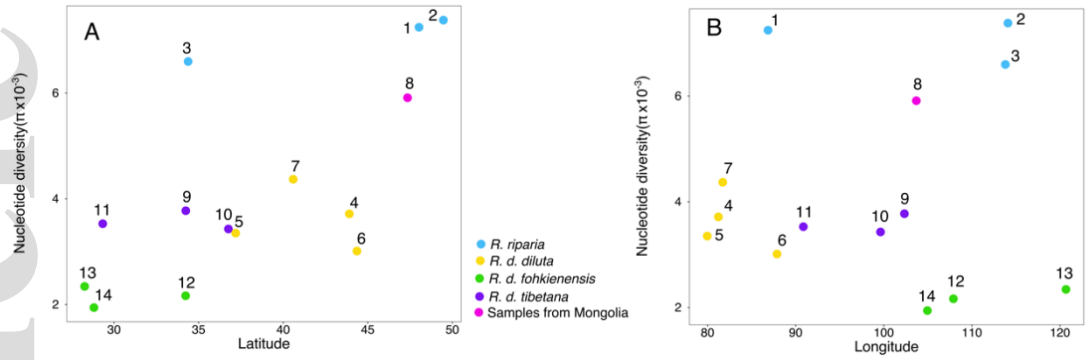


Figure 4. Nucleotide diversity (π) of *R. diluta* and *R. riparia* populations plotted against longitude (A) and latitude (B). Numbers correspond to the different populations shown in Figures 1 and 3.

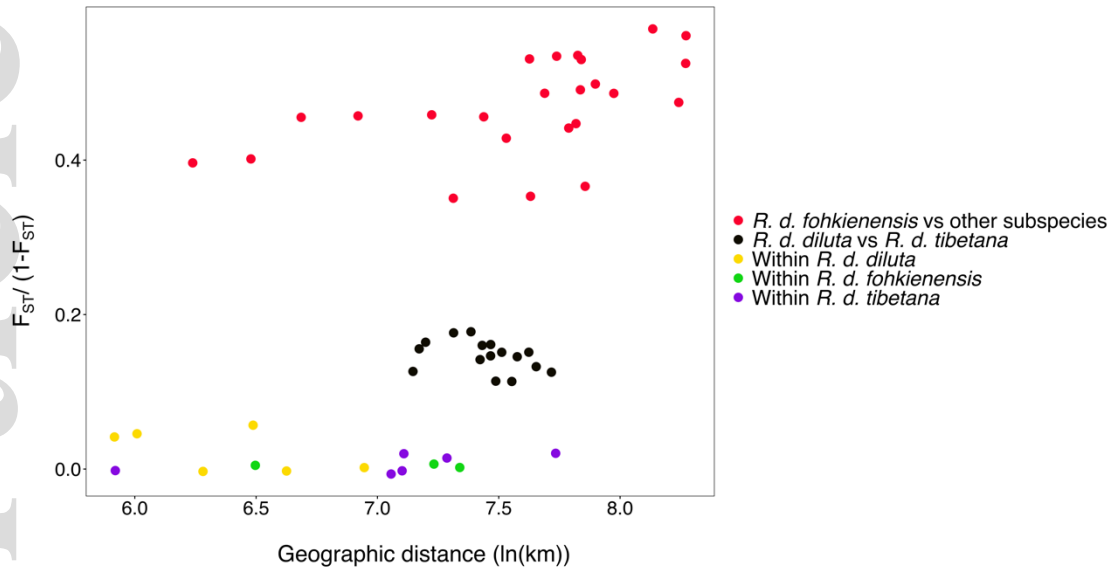


Figure 5. Genetic differentiation between populations of *R. diluta* belonging to three subspecies relative to geographic distance. Genetic differentiation between populations of different subspecies was considerably larger than comparisons within subspecies.

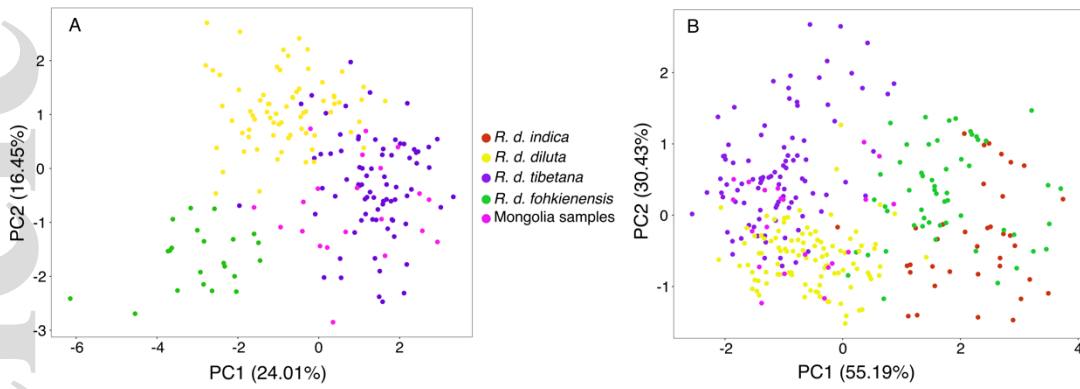


Figure 6. Principal component analysis (PCA) of subspecies within *R. diluta* based on eight (A), and on three morphometric traits (B) with additional samples from a previous analysis (Schweizer et al, 2018) and including *R. d. indica*. Colors correspond to the different subspecies, except pink dots which indicate samples collected in Mongolia.

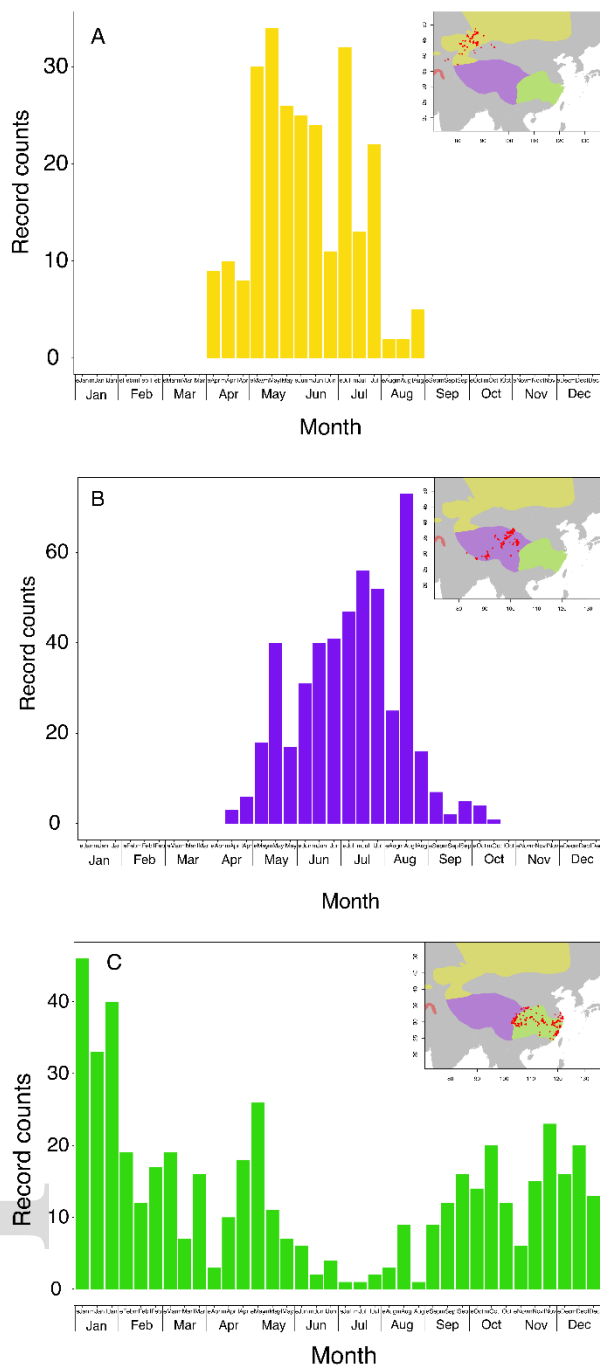


Figure 7. Phenology of 1253 records of *R. diluta* in China between 1985 and 2021. A) Records in the breeding range of *R. d. diluta* to the north-west of the Qinghai-Tibetan plateau. B) Records in the breeding range of *R. d. tibetana* on the Qinghai-Tibetan plateau above 3000 m above sea level (asl). C) Records in the breeding range of *R. d. fohkienensis* in south and central China east of the Qinghai-Tibetan plateau below 3000 m asl. Color shades in the maps indicate the potential

breeding ranges of different subspecies of *R. diluta* (yellow: *R. d. diluta*, purple: *R. d. tibetana*, green: *R. d. fohkienensis*). Each record is shown as a red dot on the maps for the respective geographic areas.

Supplementary Information

Supplementary Figures

Figure S1: Maximum likelihood phylogeny based on ND2 haplotypes.

Figure S2: Number of SNPs genotyped in *Riparia diluta* and *R. riparia* populations per autosome.

Figure S3: Correlation between the number of SNPs genotyped per autosome and chromosome size.

Figure S4: ΔK statistics for admixture analysis from K=2 to K=10.

Figure S5: Individual ancestry assignment for admixture analysis from K=2 to K=10.

Supplementary Tables

Table S1: Population information.

Table S2: Sample list and accession numbers for each individual.

Table S3: Morphological data.

Table S4: Pairwise F_{ST} and geographic distances between populations.

Table S5: Factor loadings of principal component analysis on morphometric measurements.