Unravelling Cryptic Radiation and High-Altitude Adaptation in a Migratory Bird Species Complex

Inaugural dissertation
of the Faculty of Science,
University of Bern

presented by

Qindong, Tang

from China

Supervisors of the doctoral thesis:
Prof. Gerald Heckel and Dr. Manuel Schweizer

Institute of Ecology and Evolution, University of Bern

Unravelling Cryptic Radiation and High-Altitude Adaptation in a Migratory Bird Species Complex

Inaugural dissertation
of the Faculty of Science,
University of Bern

presented by

Qindong, Tang

from China

Supervisors of the doctoral thesis :
Prof. Gerald Heckel and Dr. Manuel Schweizer
Institute of Ecology and Evolution, University of Bern

Accepted by the Faculty of Science.

Bern, **25.10.2023**

The Dean
Prof. Dr. Marco Herwegh

Public defence

October 25, 2023

University of Bern

Institute of Ecology and Evolution

Baltzerstrasse 6, 3018 Bern

Composition of the jury

Chair: Prof. Dr. Madhav Prakash Thakur

Supervisors: Prof. Gerald Heckel and Dr. Manuel Schweizer

Examiner: Prof. Darren E. Irwin

Table of Contents

Summary

Biodiversity is often accompanied with drastic phenotypic diversity. Morphological differentiation can play a crucial role in the process of speciation, such as during adaptive radiation that ecological and phenotypic diversity evolved within rapidly multiplying lineages (Schluter, 2000). However, there is a considerable proportion of biodiversity that is constituted of cryptic species which are defined as evolutionary lineages with restricted gene flow that do not form diagnostic morphological clusters (Struck et al., 2018). Our understanding of cryptic diversity and mechanisms underlying the evolution of new species without distinguished morphological differentiation remains limited. Other traits could also play an important role in speciation. By studying how speciation happened under morphological stasis and how cryptic diversity can be maintained afterwards could help us to understand other mechanisms that restrict or prevent gene flow with minimal morphological and ecological differentiation.

This thesis mainly focused on the cryptic diversification of two migratory bird species complexes, pale sand martin *Riparia diluta* and collared sand martin *Riparia riparia*, with the aim of understanding the processes and mechanisms that lead to morphologically cryptic divergence and maintain genetic partitioning. I have conducted comprehensive geographical sampling for the two taxa in East Asia including the Qinghai-Tibetan Plateau and areas surrounding the Taklamakan desert.

In Chapter 1, by using genome-wide, morphology and phenology data, I revealed cryptic radiation in the pale sand martin in Central and East Asia. My results indicated that allochrony caused by different migration/breeding time and migratory divide might play an important role to restrict gene flow and maintain evolutionary diversity in birds under morphological stasis. In Chapter 2, I conducted phylogenomic reconstruction and demographic inferences to study the biogeography of pale sand martin and collared sand martin. Contrasting diversification pattern were found between the two sister taxa: mitochondrial-nuclear discordance in Holarctic collared sand martin versus strong genomic divergence in the pale sand martin supported the existence of multiple morphologically cryptic evolutionary with comparably limited

distribution range in Asia. In Chapter 3, I investigated potential genomic adaptation to high altitudes in pale sand martin. By comparing the genomic differentiation between different populations breeding on different altitudes with large elevation change, I have identified EPAS1 as potential gene together with HBAD facilitating migratory birds' adaptation to high altitudes. Furthermore, my results indicated EPAS1 haplotypes from lowland subspecies *R. d. diluta* were likely adaptively introgressed into populations *R. d. tibetana* in central Mongolia that have expanded their range this area from the Qinghai-Tibetan plateau.

In this thesis, analyses of genome-wide data revealed cryptic radiation within a migratory bird species complex and revealed their genomic adaptation for high altitudes. My work indicated potential mechanism to restrict gene flow under cryptic diversification that in bird. We hypothesize that pronounced dispersal propensity in the strongly migratory nominate form of collared sand martin has hindered lineage divergence across its vast distribution range, while differential seasonal migratory behaviour contributes to maintain evolutionary diversity in the pale sand martin species complex. Furthermore my work indicated adaptive introgression can facilitate migratory bird to colonize new environment.

General Introduction

General Introduction

Cryptic diversification

With the advance of DNA sequencing technology, unexpected evolutionary diversity across the tree of life has been discovered through genetic methods during the past two decades. There is a considerable proportion of biodiversity that is constituted of cryptic species which are defined as evolutionary lineages with restricted gene flow that do not form diagnostic morphological clusters (Struck et al., 2018). This challenged the traditional view that speciation is always accompanied by morphological differentiation (Bickford et al., 2007; Fišer et al., 2018). Four hypotheses have been invoked that might underline cryptic diversity to explain low level of morphological differentiation: recent divergence, parallelism, convergence, and morphological stasis (Bickford et al., 2007; Fišer et al., 2018; Struck et al., 2018).

One of the most common assumption is that cryptic species have recently diverged with differentiation in sexual signaling, physiology or phenology and no substantial morphological differences has yet been accumulated. However, ancient divergence was found among cryptic species in different organisms such as bonefish (Colborn et al., 2001), amphipods (Lefébure et al., 2006) and copepods (Rocha-Olivares et al., 2001) and only a small fraction of known cryptic species of amphipods showed recent divergence younger than 1 million year (Fišer et al., 2018). In parallelism, cryptic species evolve from morphological similar ancestors while morphological similarity could also happen through convergence by independent evolutions of morphologically dissimilar ancestors (Struck et al., 2018). In both cases, extrinsic factors such as deterministic environmental pressure might play an important role in cryptic diversification (Struck et al., 2018). Morphological stasis assumes that morphological differentiation of descendant species is constrained by biological mechanisms (Bickford et al., 2007). High degree of morphological similarity between sister taxa can be maintained over extended period due to low standing genetic variation and/or developmental constraints on the morphospace, or the

ecology of the cryptic taxa has remained relatively constant through time and similar morphology was retained by strong stabilizing selection (Struck et al., 2018).

However, 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 cryptic diversity is maintained in populations in secondary contact. Other traits in sexual signalling, physiology or phenology might have played important role to restrict gene flow among morphologically cryptic lineages. Diverged social signals (Tobias et al., 2020), contrasting habitat preferences precluding secondary sympatry or different timing of reproduction facilitating co-existence through temporal segregation (allochrony) can act as prezygotic/premating isolation (Leys et al., 2017; Taylor & Friesen, 2017). Furthermore, recent studies on cryptic sister species of birds in Amazonia found strong postzygotic reproductive isolation with little evidence for premating isolation in relatively young pairs (Cronemberger et al., 2020; Pulido-Santacruz et al., 2018). However, whether premating or postzygotic reproductive isolation would play a major role in maintaining cryptic diversity remains to under debates. 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 (Beysard et al., 2015; Beysard & Heckel, 2014).

Role of introgression in cryptic diversification

Allopatric isolation has been shown to be the major driver in bird speciation (Phillimore et al., 2008) that reproductive isolating mechanisms might arise in allopatry as a by-product of populations adapting to different environments, sexual selection regimes, etc (Mayr, 1963; Price, 2008). However, increasing number of genetic studies showed gene flow happens frequently between divergent lineages that 16.4% of avian species have hybridized with at least one other species in the wild (Ottenburghs et al., 2015) and introgression has been suggested fairly common across

the suboscine bird radiation (Singhal et al., 2021). The role of introgressive hybridization in evolution is controversial. On the one hand hybridization might hinder speciation by slowing or reversing differentiation due to gene flow whereas introgression of a few loci may promote adaptive divergence facilitating speciation (Abbott et al., 2013). Recruitment of ancestral variations through hybridization could be more efficient to induce adaptation in a short period of time than de novo mutation. Increasing studies show that reassembly of standing genetic variation into new combinations through introgression played a key role in adaptive radiations and rapid radiation (Marques et al., 2019; Meier et al., 2017; Rubin et al., 2022). Under cryptic diversification, gene flow might happen more frequently especially between recently divergent lineages when they come into secondary contact (Dufresnes et al., 2019). Thus, studying cryptic diversity might provide us opportunities to evaluate the role of gene flow during the process of speciation.

Diversification under Pleistocene climate fluctuations

Current biodiversity patterns have been strongly influenced by past Pleistocene climate fluctuations, which promoted speciation and population diversification and altered distributions (Head & Gibbard, 2005; Hewitt, 2004; Lovette, 2005). Extensive ice sheets were built up during the periods of global cooling, followed by ice sheets retreat during brief interglacial periods. Habitat shifts and fragmentation caused by global temperature fluctuations and glacial advances during the Pleistocene. Rapid diversification could happen under the such impacts of Pleistocene climate cycling (Weir & Schluter, 2004) and induce cryptic diversity (Weir et al., 2016). There are 16-17 genetically distinct lineages identifed within the currently recognized five kiwi species and most their diversification dates to the seven major glacial advances that directly fragmented New Zealand into a series of glacial refugia (Weir et al., 2016).

The climatic changes, particularly during the Pleistocene glacial cycles, have disparate impacts on different geographic regions and have led to discernible differences in spatio-temporal patterns of diversification (Weir and Schluter 2004; Lovette 2005; Qu et al. 2014).

Contrasting with temperate and boreal taxa, diversification patterns within taxa from the Sino-Himalayan and adjacent regions might exhibit significant differences. Recent evidence suggests that some bird populations in Europe experience post-Last Glacial Maximum (LGM) expansion, while those in eastern Asia remained relatively stable during the same period (Song et al., 2016). Prolonged population isolation, coupled with complex environmental dynamics, could contribute to more regional patterns in East Asia, potentially accounting for the remarkably high diversity observed in central and southern China (Liu et al., 2012; Myers et al., 2000; Qu et al., 2014; Song et al., 2009, 2016; Wu et al., 2012). The comparative investigation of species complexes with a wide distribution in different biogeographic realms is a particularly promising approach to understand divers of diversification in a regional context.

Previous studies of diversification within pale sand martin and collared sand martin

The collared sand martin *Riparia diluta* and the pale martin *Riparia diluta* are two migratory sister species within the genus *Riparia* in the swallow family Hirundinidae. In total five extant species are included in *Riparia* (Sheldon et al., 2005) and they are quite similar in plumage and all nest in tunnels in natural sand banks or earth mounds. Collared sand martin and pale sand martin have been considered as conspecifics until recently, however they were shown to have wide zone of overlap in Central Asia and showed genetic, vocal and slight morphological and morphometric differences (Goroshko, 1993; Pavlova et al., 2008). The collared sand martin is widespread across the whole Holarctic including the Middle East and the Nile Valley in Egypt (Turner, 2004), with shallow phylogeographic structure in mitochondrial DNA (mtDNA) across its entire breeding range indicating that its large range might due to a relatively recent expansion (Pavlova et al. 2008; Schweizer et al. 2018). Three subspecies are broadly accepted (Shirihai & Svensson, 2018): small-ranged *R. r. shelleyi* of Egypt and possibly the Levant, *R. r. ijimae* with an intermediate-sized range in Central Mongolia, south-eastern Siberia, north-eastern China and Japan, and nominate *R. r. riparia* with a vast breeding range comprising Asia, Europe and North America. The pale sand martin however has a relatively restricted distribution range in Central and Eastern Asia with usually four accepted subspecies corresponding to

four distinct mtDNA lineages that diverged during the last million years (Schweizer et al., 2018). Furthermore, these four subspecies are morphologically similar and breed at a broad range of elevations and different habitats/climate zones: *R. d. fohkienensis* breed in the lowland of South China, *R. d. diluta* in Central Asia, *R. d. tibetana* on the Qinghai-Tibetan plateau (above 3500m) and *R. d. indica* around the Indian subcontinent.

The previous studies were only based on few genetic markers including one mtDNA marker, one Z-linked nuclear intron and one autosomal nuclear intron (Pavlova et al., 2008; Schweizer et al., 2018) and I want to check whether the diversification patter is consistent on genomic level. The geographic sampling was limited and lack of samples from parapatric regions between different *R. diluta* subspecies. While *R. d. indica* might be isolated, *R. d. diluta* and *R. d. tibetana* might come into contact along the Kunlun Shan or at the southern margin of the Tarim Basin. Moreover, haplotypes of *R. d. tibetana* found in north-eastern Mongolia (Schweizer et al., 2018) might indicate that the range of *R. d. tibetana* extends much further northeast and a potential contact between the two subspecies. In addition, whether *R. d. fohkienensis* and *R. d. tibetana* are in contact at the eastern edge of the Qinghai-Tibetan Plateau remains unclear. Local adaptation might constrain the breeding range of the two subspecies with the former restricted to the lowland and the latter to high altitudinal areas.

Outlines of the thesis

In summary, pale sand martin provides a promising system to study cryptic diversification in birds. The aim of this thesis is to study the diversification and adaptation within pale sand martin and to understand how the cryptic diversity can be maintained.

In Chapter 1, I mainly focus on studying the diversification in pale sand martin using GBS (genotyping by sequencing) data from multiple populations of different subspecies within pale sand martin in comparison with collared sand martin. I also checked for morphological differentiation among all subspecies of pale sand martin

subspecies and I tested whether there is gene flow between parapatric populations of different subspecies within pale sand martin especially on the north-western and north-eastern edges of Qinghai-Tibetan plateau.

In Chapter 2, I performed phylogeographic and demographic reconstruction for both pale sand martin and collared sand martin based on individual re-sequencing data to study the regional effects of past Pleistocene clime fluctuation on the diversification pattern of the two sister species.

In Chapter 3, I investigated potential genomic adaptation to high altitudes in pale sand martin. Most studies about high-altitude adaptation were focused on resident species inhabiting high elevations. However, whether we can find genetic adaptation to high-altitudes in migratory birds during elevation shifts are still largely unknown.

Reference

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., … Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*(2), 229–246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

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.

Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., Ingram, K. K., & Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, *22*(3), 148–155. https://doi.org/10.1016/j.tree.2006.11.004

Colborn, J., Crabtree, R. E., Shaklee, J. B., Pfeiler, E., & Bowen, B. W. (2001). THE EVOLUTIONARY ENIGMA OF BONEFISHES (ALBULA SPP.): CRYPTIC SPECIES AND ANCIENT SEPARATIONS IN A GLOBALLY DISTRIBUTED SHOREFISH. *Evolution*, *55*(4), 807. https://doi.org/10.1554/0014-3820(2001)055[0807:TEEOBA]2.0.CO;2

Dufresnes, C., Strachinis, I., Suriadna, N., Mykytynets, G., Cogălniceanu, D., Székely, P., Vukov, T., Arntzen, J. W., Wielstra, B., Lymberakis, P., Geffen, E., Gafny, S., Kumlutaş, Y., Ilgaz, Ç., Candan, K., Mizsei, E., Szabolcs, M., Kolenda, K., Smirnov, N., … Denoël, M. (2019). Phylogeography of a cryptic speciation continuum in Eurasian spadefoot toads (*Pelobates*). *Molecular Ecology*, *28*(13), 3257–3270. https://doi.org/10.1111/mec.15133

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. https://doi.org/10.1111/mec.14486

Goroshko, O. A. (1993). Taxonomic status of the pale (sand?) martin *Riparia* (*riparia*?) *diluta* (Sharpe & Wyatt, 1893). *Russ. Ornit. Zhurnal*, *2*(3), 303–323.

Head, M. J., & Gibbard, P. L. (2005). Early-Middle Pleistocene transitions: An overview and recommendation for the defining boundary. *Geological Society, London, Special Publications*, *247*(1), 1–18. https://doi.org/10.1144/GSL.SP.2005.247.01.01

Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *359*(1442), 183–195. https://doi.org/10.1098/rstb.2003.1388

Lefébure, T., Douady, C. J., Gouy, M., Trontelj, P., Briolay, J., & Gibert, J. (2006). Phylogeography of a subterranean amphipod reveals cryptic diversity and dynamic evolution in extreme environments: CRYPTIC AND DYNAMIC EVOLUTION IN SUBSURFACE. *Molecular Ecology*, *15*(7), 1797–1806. https://doi.org/10.1111/j.1365-294X.2006.02888.x

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. https://doi.org/10.1111/mec.14026

Liu, H., Wang, W., Song, G., Qu, Y., Li, S.-H., Fjeldså, J., & Lei, F. (2012). Interpreting the Process behind Endemism in China by Integrating the Phylogeography and Ecological Niche Models of the Stachyridopsis ruficeps. *PLoS ONE*, *7*(10), e46761. https://doi.org/10.1371/journal.pone.0046761

Lovette, I. (2005). Glacial cycles and the tempo of avian speciation. *Trends in Ecology & Evolution*, *20*(2), 57–59. https://doi.org/10.1016/j.tree.2004.11.011

Marques, D. A., Meier, J. I., & Seehausen, O. (2019). A Combinatorial View on Speciation and Adaptive Radiation. *Trends in Ecology & Evolution*, *34*(6), 531–544. https://doi.org/10.1016/j.tree.2019.02.008

Mayr, E. (1963). *Animal Species and Evolution*. Belknap Press.

Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications*, *8*(1), 14363. https://doi.org/10.1038/ncomms14363

Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, *403*(6772), 853–858. https://doi.org/10.1038/35002501

Ottenburghs, J., Ydenberg, R. C., Van Hooft, P., Van Wieren, S. E., & Prins, H. H. T. (2015). The Avian Hybrids Project: Gathering the scientific literature on avian hybridization. *Ibis*, *157*(4), 892–894. https://doi.org/10.1111/ibi.12285

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. https://doi.org/10.1016/j.ympev.2008.03.030

Phillimore, A. B., Orme, C. D. L., Thomas, G. H., Blackburn, T. M., Bennett, P. M., Gaston, K. J., & Owens, I. P. F. (2008). Sympatric Speciation in Birds Is Rare: Insights from Range Data and Simulations. *The American Naturalist*, *171*(5), 646–657. https://doi.org/10.1086/587074

Price, T. (2008). *Speciation in birds*. Roberts and Co.

Price, T. D. (2008). *Speciation in Birds*. Roberts and Company.

Qu, Y., Ericson, P. G. P., Quan, Q., Song, G., Zhang, R., Gao, B., & Lei, F. (2014). Long-term isolation and stability explain high genetic diversity in the Eastern Himalaya. *Molecular Ecology*, *23*(3), 705–720. https://doi.org/10.1111/mec.12619

Rocha-Olivares, A., Fleeger, J. W., & Foltz, D. W. (2001). Decoupling of Molecular and Morphological Evolution in Deep Lineages of a Meiobenthic Harpacticoid Copepod. *Molecular Biology and Evolution*, *18*(6), 1088–1102. https://doi.org/10.1093/oxfordjournals.molbev.a003880

Rubin, C.-J., Enbody, E. D., Dobreva, M. P., Abzhanov, A., Davis, B. W., Lamichhaney, S., Pettersson, M., Sendell-Price, A. T., Sprehn, C. G., Valle, C. A., Vasco, K., Wallerman, O., Grant, B. R., Grant, P. R., & Andersson, L. (2022). Rapid adaptive radiation of Darwin's finches depends on ancestral genetic modules. *Science Advances*, *8*(27), eabm5982. https://doi.org/10.1126/sciadv.abm5982

Schluter, D. (2000). *The ecology of adaptive radiation*. Oxford University Press.

Schweizer, M., Liu, Y., Olsson, U., Shirihai, 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

Sheldon, F. H., Whittingham, L. A., Moyle, R. G., Slikas, B., & Winkler, D. W. (2005). Phylogeny of swallows (Aves: Hirundinidae) estimated from nuclear and mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, *35*(1), 254–270. https://doi.org/10.1016/j.ympev.2004.11.008

Shirihai, H., & Svensson, L. (2018). *Handbook of Western Palearctic Birds*. A. & C. Black.

Singhal, S., Derryberry, G. E., Bravo, G. A., Derryberry, E. P., Brumfield, R. T., & Harvey, M. G. (2021). The dynamics of introgression across an avian radiation. *Evolution Letters*, *5*(6), 568–581. https://doi.org/10.1002/evl3.256

Song, G., Qu, Y., Yin, Z., Li, S., Liu, N., & Lei, F. (2009). Phylogeography of the Alcippe morrisonia(Aves: Timaliidae): long population history beyond late Pleistocene glaciations. *BMC Evolutionary Biology*, *9*(1), 143. https://doi.org/10.1186/1471-2148-9-143

Song, G., Zhang, R., DuBay, S. G., Qu, Y., Dong, L., Wang, W., Zhang, Y., Lambert, D. M., & Lei, F. (2016). East Asian allopatry and north Eurasian sympatry in Long-tailed Tit lineages despite similar population dynamics during the late

Pleistocene. *Zoologica Scripta*, *45*(2), 115–126. https://doi.org/10.1111/zsc.12148

Struck, T. H., Feder, J. L., Bendiksby, M., Birkeland, S., Cerca, J., Gusarov, V. I., Kistenich, S., Larsson, K.-H., Liow, L. H., Nowak, M. D., Stedje, B., Bachmann, L., & Dimitrov, D. (2018). Finding Evolutionary Processes Hidden in Cryptic Species. *Trends in Ecology & Evolution*, *33*(3), 153–163. https://doi.org/10.1016/j.tree.2017.11.007

Taylor, R. S., & Friesen, V. L. (2017). The role of allochrony in speciation. *Molecular Ecology*, *26*(13), 3330–3342.

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

Turner, A. K. (2004). Cotingas to Pipits. In *J. del Hoyo, A. Elliot, and D. A. Christie, eds. Handbook of the Birds of the World.* Lynx Edicions.

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). https://doi.org/10.1073/pnas.1603795113

Weir, J. T., & Schluter, D. (2004). Ice sheets promote speciation in boreal birds. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *271*(1551), 1881–1887. https://doi.org/10.1098/rspb.2004.2803

Wu, Y., Huang, J., Zhang, M., Luo, S., Zhang, Y., Lei, F., Sheldon, F. H., & Zou, F. (2012). Genetic divergence and population demography of the Hainan endemic Black-throated Laughingthrush (Aves: Timaliidae, *Garrulax chinensis monachus*) and adjacent mainland subspecies. *Molecular Phylogenetics and Evolution*, *65*(2), 482–489. https://doi.org/10.1016/j.ympev.2012.07.005

Chapter 1

Received: 23 January 2021 | Revised: 7 October 2021 | Accepted: 12 October 2021

DOI: 10.1111/mec.16241

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

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

Qindong Tang[1,2] | Reto Burri[3] | Yang Liu[4] | Alexander Suh[5,6] | Gombobaatar Sundev[7] | Gerald Heckel[1] | Manuel Schweizer[1,2]

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

[2]Natural History Museum, Bern, Switzerland

[3]Schweizerische Vogelwarte, Sempach, Switzerland

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

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

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

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

Correspondence
Manuel Schweizer, Natural History Museum, Bern, Switzerland; Institute of Ecology and Evolution, University of Bern, Switzerland.
Email: manuel.schweizer@nmbe.ch

Funding information
Oversea Study Program of Guangzhou Elite Project, Grant/Award Number: JY201726; Basler Stiftung für biologische Forschung; Knut och Alice Wallenbergs Stiftelse

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 nonoptimal 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.

KEYWORDS
allochrony, cryptic diversification, phylogeography, population genomics, *Riparia diluta*, Sand martin

Gerald Heckel and Manuel Schweizer are joint senior authors.

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 nonadaptive radiations (Rundell & 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", that is, 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, for example, 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 & Heckel, 2014; Beysard et al., 2015).

The breakdown 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, analysing 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 (a) if the deep phylogeographic structure indicated by mtDNA is also reflected in genome-wide variation and (b) if the phylogeographic lineages indeed overlap in morphospace and can thus be considered cryptic. Finally, we (c) investigated, if gene flow between phylogeographic lineages is actually reduced in areas of potential contact.

2 | MATERIALS 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 (Figure 1, Table S1). We sampled 14 breeding colonies of *R. diluta* covering the breeding 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

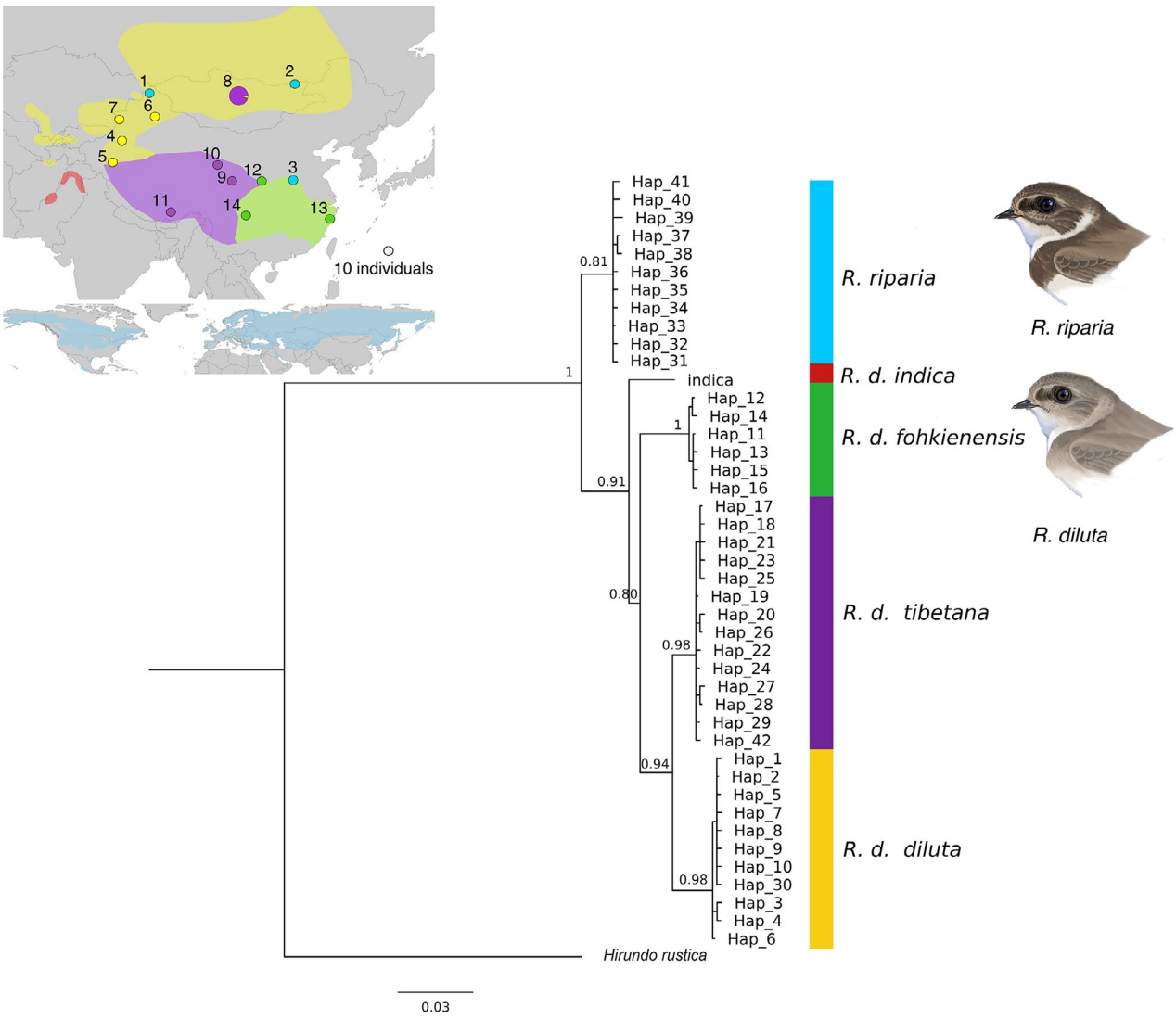


FIGURE 1 Bayesian majority rule consensus tree based on haplotypes of the mtDNA gene NADH dehydrogenase subunit II (ND2) of *R. diluta* and *R. riparia*. Bayesian posterior probabilities are given for major nodes. Colour shades indicate the potential breeding ranges of different subspecies of *R. diluta* (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 colours 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

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 (Aljanabi & Martinez, 1997). 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 phyml-test in PhyML (Guindon & Gascuel, 2003) with R 3.6.2 (R core Team, 2013). ML tree search was employed in PhyML with 1000 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 & Bull, 1993) and clade credibility values for the BI > 0.95 (Huelsenbeck & 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/).

2.3 | 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 10× 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).

2.4 | 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 $p < .001$ 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 <10% missing data (-uniqueOnly 1 -skipTriallelic 1 -minInd 135).

2.5 | 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 or *R. riparia*, the folded SFS was used.

2.6 | 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 data set 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.

2.7 | 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 fieldwork 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, that is, before the 10th, between the 10th and the 20th, and after the 20th. The three geographic regions were defined as follows: (a) 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*, (b) Qinghai-Tibetan plateau above 3000 m asl corresponding to the supposed breeding range of *R. d. tibetana*, (c) 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 and 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 7640 autosomal SNPs in the data set 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 = .92; $p = 5.02 \times 10^{-14}$) between the number of SNPs called per chromosome and chromosome size (Figures S2 and 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 (26.66% 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 and 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 6857 SNPs supported the closer affinity of Mongolian birds to *R. d. tibetana*. Overall differentiation between populations was very high with $F_{ST} = 0.225$ ($p < .0001$). 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 $p < .0001$). 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 < .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 $p < .05$; Table S4) while all pairwise comparisons between populations from different subspecies ranged between $F_{ST} = 0.1$ and 0.36 (all $p < .05$; 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 3.09×10^{-3} to 4.49×10^{-3}; Figure 4). All populations of *R. d. fohkienensis* showed comparatively lower levels of nucleotide diversity (π ranging from 1.99×10^{-3} to 2.39×10^{-3}). The highest values of nucleotide diversity were found in the three analysed populations of *R. riparia* (π ranging from 6.69×10^{-3} to 7.49×10^{-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

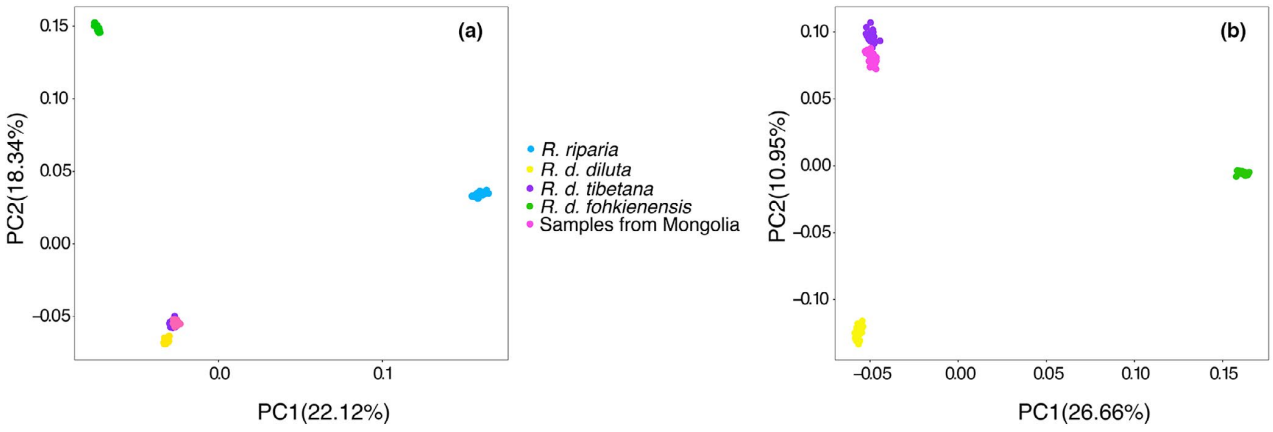


FIGURE 2 Principal component analysis (PCA) on nuclear SNPs of *R. diluta* and *R. riparia* (a, based on 7,640 SNPs) and only of *R. diluta* (b, based on 7118 SNPs). Colours correspond to different subspecies except pink dots indicate the samples collected from Mongolia in the potential breeding range of *R. d. diluta*

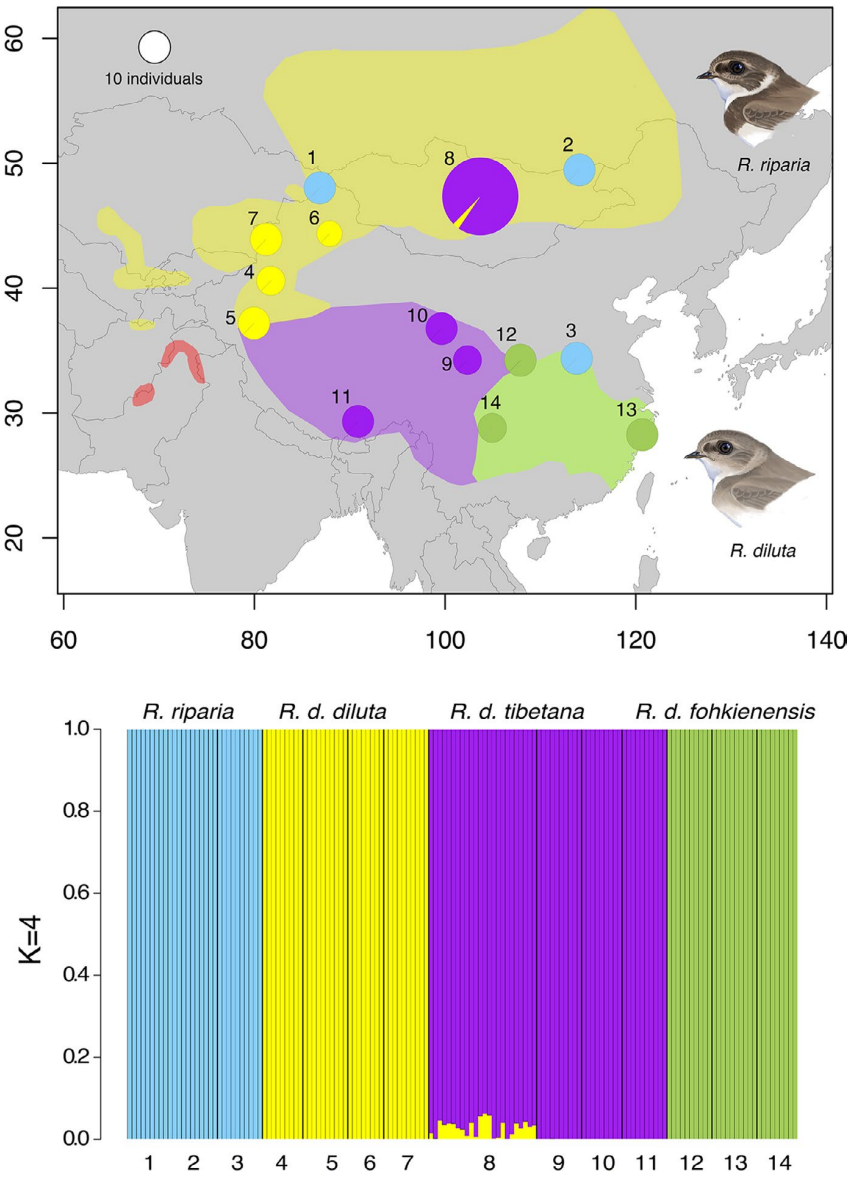


FIGURE 3 Top: Map with pie charts representing genome-wide ancestry assignment in *Riparia* populations with $K = 4$ using NgsAdmix. Limited admixture between *R. d. diluta* and *R. d. tibetana* was found in the population from Mongolia (8) with $K = 4$. Colour shades indicate the potential breeding 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

between populations ($r = .5486$; $p = .001$; 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 ($r = -.304$; $p = .87$), 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.

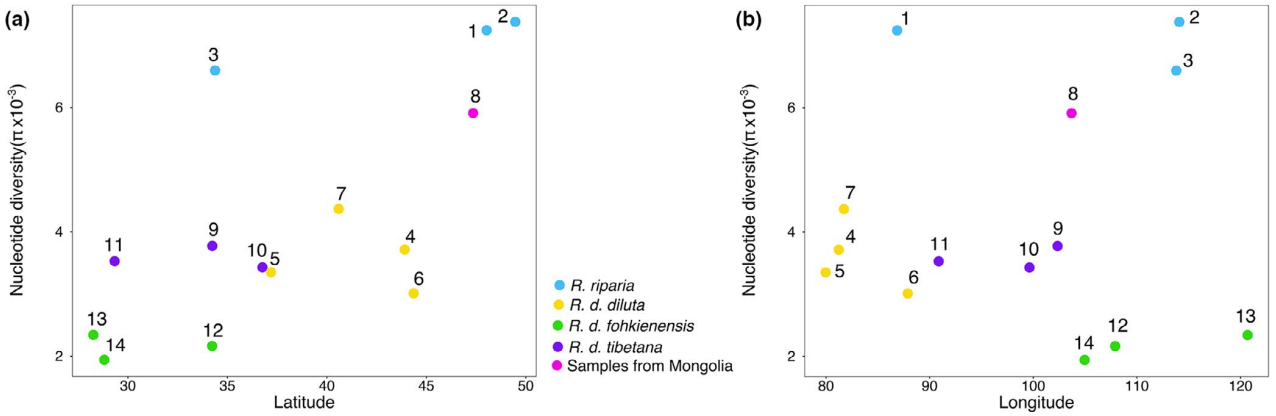


FIGURE 4 Nucleotide diversity (π) of *R. diluta* and *R. riparia* populations plotted against latitude (a) and longitude (b). Numbers correspond to the different populations shown in Figures 1 and 3

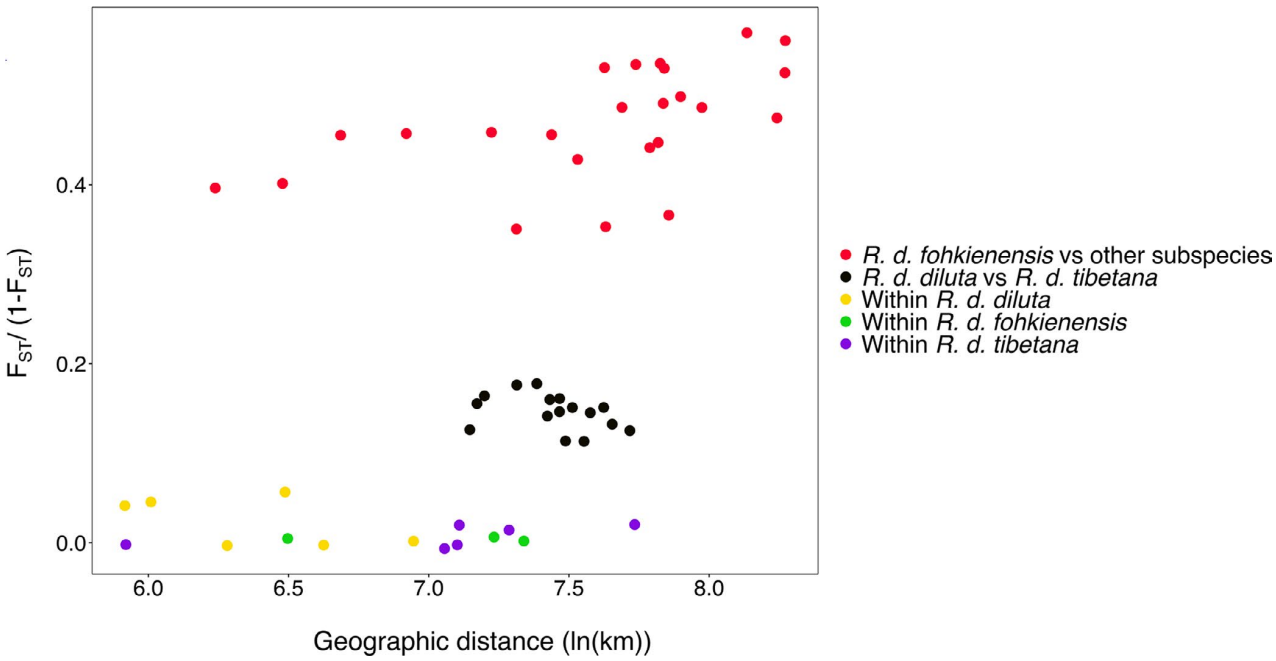


FIGURE 5 Genetic differentiation between populations of *R. diluta* belonging to three subspecies relative to geographic distance. Genetic differentiation between populations of different subspecies was considerably larger than comparisons within subspecies

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

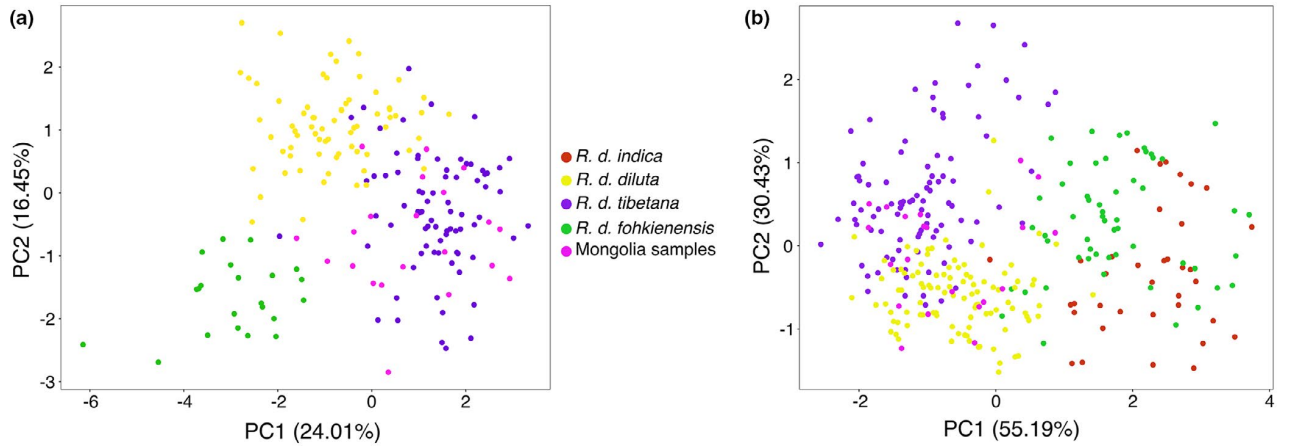


FIGURE 6 Principal component analysis (PCA) of subspecies within *R. diluta* based on eight (a), and on three morphometric traits (b) with additional samples from a previous analysis (Schweizer et al., 2018) and including *R. d. indica*. Colours correspond to the different subspecies, except pink dots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

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, 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 - probably 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, that is,

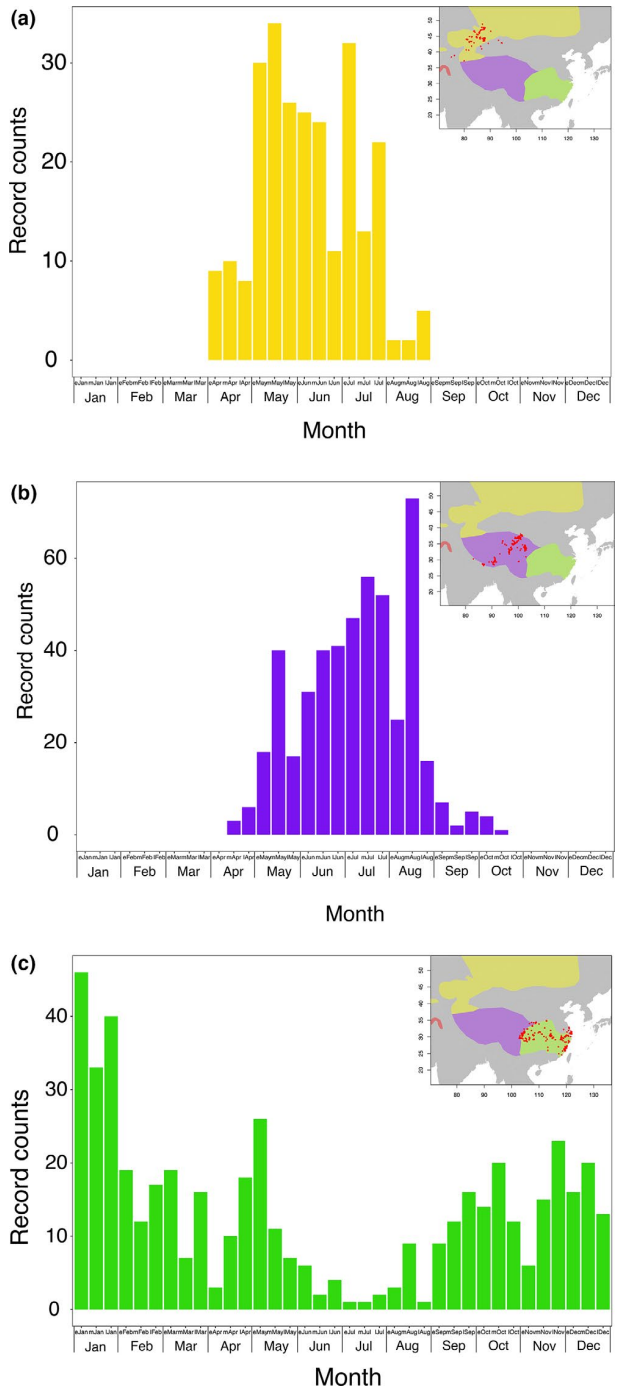


FIGURE 7 Phenology of 1253 records of *R. diluta* in China between 1985 and 2021. (a) Records in the breeding range of *R. d. diluta* to the north-west of the Qinghai-Tibetan plateau. (b) Records in the breeding range of *R. d. tibetana* on the Qinghai-Tibetan plateau above 3,000 m above sea level (asl). (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. Colour 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

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; Gómez-Bahamón et al., 2020; Kimmitt et al., 2019; Sirkiä et al., 2018; Taylor et al., 2018, 2019; Taylor & Friesen, 2017). However, other potential prezygotic isolation mechanisms such as differences in mating behaviour between *R. d. fohkienensis* and *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 (c. 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 probably 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 nonoptimal intermediate migration routes and might be selected against (Berthold et al., 1992; Delmore, Hübner, et al., 2015; Delmore, Kenyon, 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 crypsis 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, Hübner, et al., 2015; Delmore, Kenyon, 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 Li, 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 Programme of Guangzhou Elite Project (no. JY201726) and of Basler Stiftung für biologische Forschung. We moreover thank the Cornell Laboratory 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. Open Access Funding provided by Universitat Bern.

AUTHOR CONTRIBUTIONS

Manuel Schweizer and Gerald Heckel conceived the study; Qindong Tang, Manuel Schweizer and Gombobaatar Sundev collected samples with assistance from Yang Liu. Qindong Tang conducted laboratory work and analysed the data together with Gerald Heckel and Manuel Schweizer and assistance from Reto Burri; Alexander Suh and Reto Burri contributed materials; Qindong Tang, Manuel Schweizer and Gerald Heckel wrote the manuscript with support from all authors.

DATA AVAILABILITY STATEMENT

The data supporting this study have been made 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 nos. SRR15558367–SRR15558515 and BioSample numbers SAMN20856160–SAMN20856308 for raw individual GBS sequences, and accession no. JAIXNV000000000 and BioSample no. SAMN20845799 and for the genome assembly.

ORCID

Reto Burri https://orcid.org/0000-0002-1813-0079
Yang Liu https://orcid.org/0000-0003-4580-5518
Alexander Suh https://orcid.org/0000-0002-8979-9992
Gerald Heckel https://orcid.org/0000-0002-0162-323X
Manuel Schweizer https://orcid.org/0000-0002-7555-8450

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. https://doi.org/10.1093/nar/25.22.4692

Andrew, S. (2019). *FastQC: A quality control tool for high throughput sequence data 2010*. Retrieved from: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, *16*(1), 113–148. https://doi.org/10.1146/annurev.es.16.110185.000553

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. https://doi.org/10.1186/s12862-015-0455-5

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. https://doi.org/10.1093/bioinformatics/btu170

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 S. Misener & S. A. Krawetz (Eds.), *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*, *74*(11):2512–2525. https://doi.org/10.1111/evo.14103

del Hoyo, J., & Collar, N. J. (2016). *HBW and BirdLife International illustrated checklist of the birds of the world: Vol. 2: Passerines*. Lynx Edicions.

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. https://doi.org/10.1111/mec.13150

Delmore, K. E., & Irwin, D. E. (2014). Hybrid songbirds employ intermediate routes in a migratory divide. *Ecology Letters*, *17*(10), 1211–1218. https://doi.org/10.1111/ele.12326

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. https://doi.org/10.1098/rspb.2015.1921

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*. Wilhelmshaven: Deutsche Ornithologen-Gesellschaft.

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), e19379. https://doi.org/10.1371/journal.pone.0019379

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. https://doi.org/10.1093/sysbio/syq042

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. https://doi.org/10.1111/mec.14486

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 of the United States of America*, *104*(47), 18589–18594. https://doi.org/10.1073/pnas.0700446104

Gavrilov, E. I., & Savchenko, A. P. (1991). On species validity of the Pale Sand Martin (*Riparia diluta* Sharpe et Wyatt, 1893). *Bjulleten' Moskovskogo Obscestva Ispytatelej Prirody*, *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). *Russkiĭ ornitologicheskiĭ 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. https://doi.org/10.1080/10635150390235520

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. https://doi.org/10.1007/BF00172133

Helbig, A. (1996). Genetic basis, mode of inheritance and evolutionary changes of migratory directions in palaearctic warblers (Aves: Sylviidae). *Journal of Experimental Biology*, *199*(1), 49–55. https://doi.org/10.1242/jeb.199.1.49

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. https://doi.org/10.1093/sysbio/42.2.182

Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, *17*, 754–755. https://doi.org/10.1093/bioinformatics/17.8.754

Irwin, D. E. (2020). Assortative mating in hybrid zones is remarkably ineffective in promoting speciation. *The American Naturalist*, *195*(6), E150–E167. https://doi.org/10.1086/708529

Irwin, D. E., & Irwin, J. H. (2005). Siberian migratory divides: The role of seasonal migration in speciation. In R. Greenberg & P. P. Marra

(Eds.), *Birds of two worlds: The ecology and evolution of migration* (pp. 27–40). Johns Hopkins University Press.

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. https://doi.org/10.1038/s41467-018-03294-w

Kimmitt, 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), 1–16. 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. https://doi.org/10.1186/s12859-014-0356-4

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. https://doi.org/10.1093/molbev/msw054

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. https://doi.org/10.1111/ele.12346

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. https://doi.org/10.1111/mec.14026

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. https://doi.org/10.1093/bioinformatics/bty191

Lischer, H. E., & Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, *28*(2), 298–299. https://doi.org/10.1093/bioinformatics/btr642

Liu, Y., Hu, J. H., Li, S. H., Duchen, P., Wegmann, D., & Schweizer, M. (2016). Sino-Himalayan mountains act as cradles of diversity and immigration centres in the diversification of parrotbills (Paradoxornithidae). *Journal of Biogeography*, *43*, 1488–1501. https://doi.org/10.1111/jbi.12738

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*, *47*, 2714–2727.

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. https://doi.org/10.1002/evl3.15

Lutgen, D., Ritter, R., Olsen, R.-A., Schielzeth, H., Gruselius, J., Ewels, P., García, J. T., Shirihai, H., Schweizer, M., Suh, A., & Burri, R. (2020). Linked-read sequencing enables haplotype-resolved resequencing at population scale. *Molecular Ecology Resources*, *20*(5), 1311–1322. https://doi.org/10.1111/1755-0998.13192

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*, *586*, 75–79.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Mead, C. J. (1979). Colony fidelity and interchange in the sand martin. *Bird Study*, *26*(2), 99–106. https://doi.org/10.1080/00063657909476625

Meisner, J., & Albrechtsen, A. (2018). Inferring population structure and admixture proportions in low-depth NGS data. *Genetics*, *210*(2), 719–731. https://doi.org/10.1534/genetics.118.301336

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. https://doi.org/10.1016/j.ympev.2008.03.030

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. https://doi.org/10.1098/rspb.2017.2081

Qu, Y. H., Ericson, P. G. P., Quan, Q., Song, G., Zhang, R. Y., Gao, B., & Lei, F. M. (2014). Long-term isolation and stability explain high genetic diversity in the Eastern Himalaya. *Molecular Ecology*, *23*, 705–720. https://doi.org/10.1111/mec.12619

R Core Team. (2013). *R: A language and environment for statistical computing*. R Development Core Team.

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. https://doi.org/10.1093/sysbio/syy032

Rasmussen, P. C., Anderton, J. C., & Edicions, L. (2005). *Birds of south Asia: The Ripley guide*. Barcelona: Lynx Edicions.

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. https://doi.org/10.1093/molbev/msx248

Rundell, R. J., & Price, T. D. (2009). Adaptive radiation, nonadaptive radiation, ecological speciation and nonecological speciation. *Trends*

in Ecology & Evolution, *24*(7), 394–399. https://doi.org/10.1016/j.tree.2009.02.007

Schluter, D. (2000). *The ecology of adaptive radiation*. OUP Oxford.

Schweizer, M., & Aye, R. (2007). Identification of the pale sand martin *Riparia diluta* in Central Asia. *Alula*, *4*, 152–158.

Schweizer, M., Liu, Y., Olsson, U., Shirihai, 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

Scordato, E. S. C., Smith, C. C. R., Semenov, G. A., Liu, Y. U., Wilkins, M. R., Liang, W., Rubtsov, A., Sundev, G., Koyama, K., Turbek, S. P., Wunder, M. B., Stricker, C. A., & Safran, R. J. (2020). Migratory divides coincide with reproductive barriers across replicated avian hybrid zones above the Tibetan Plateau. *Ecology Letters*, *23*(2), 231–241. https://doi.org/10.1111/ele.13420

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. https://doi.org/10.1111/j.1365-294X.2007.03529.x

Shirihai, H., & Svensson, L. (2018). *Handbook of western palearctic birds*. A. & C. Black.

Sirkiä, P. M., McFarlane, S. E., Jones, W., Wheatcroft, D., Ålund, 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. https://doi.org/10.1111/evo.13404

Skotte, L., Korneliussen, T. S., & Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. *Genetics*, *195*(3), 693–702. https://doi.org/10.1534/genetics.113.154138

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. https://doi.org/10.1111/mec.15377

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. https://doi.org/10.1016/j.ympev.2020.106749

Struck, T. H., Feder, J. L., Bendiksby, M., Birkeland, S., Cerca, J., Gusarov, V. I., Kistenich, S., Larsson, K.-H., Liow, L. H., Nowak, M. D., Stedje, B., Bachmann, L., & Dimitrov, D. (2018). Finding evolutionary processes hidden in cryptic species. *Trends in Ecology & Evolution*, *33*(3), 153–163. https://doi.org/10.1016/j.tree.2017.11.007

Taylor, R. 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, R. S., Bolton, M., Beard, A., Birt, T., Deane-Coe, P., Raine, A. F., González-Solís, J., Lougheed, 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, 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

Tittler, R., Villard, M.-A., & Fahrig, L. (2009). How far do songbirds disperse? *Ecography*, *32*(6), 1051–1061. https://doi.org/10.1111/j.1600-0587.2009.05680.x

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. https://doi.org/10.1016/j.tree.2017.11.008

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 of the United States of America*, *113*(38), E5580–E5587. https://doi.org/10.1073/pnas.1603795113

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. https://doi.org/10.1101/gr.214874.116

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., Sarver, B. A. J., Schenk, J. J., Spear, S. F., & Harmon, L. J. (2010). Ecological opportunity and the origin of adaptive radiations. *Journal of Evolutionary Biology*, *23*(8), 1581–1596. https://doi.org/10.1111/j.1420-9101.2010.02029.x

SUPPORTING INFORMATION

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

How to cite this article: Tang, Q., Burri, R., Liu, Y., Suh, A., Sundev, G., Heckel, G., & Schweizer, M. (2022). Seasonal migration patterns and the maintenance of evolutionary diversity in a cryptic bird radiation. *Molecular Ecology*, 31, 632–645. https://doi.org/10.1111/mec.16241

Supplements

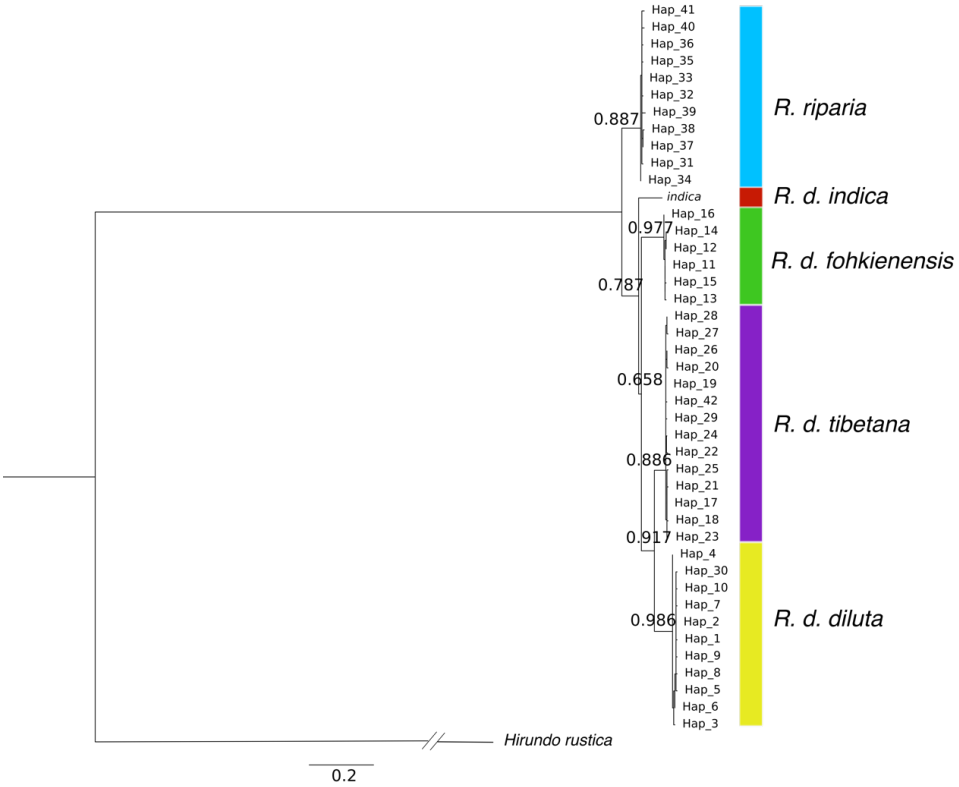


Figure S1. Maximum likelihood phylogeny based on haplotypes of the mtDNA gene NADH dehydrogenase subunit II (ND2) of *R. diluta* and *R. riparia*. Bootstrap values are given for major nodes.

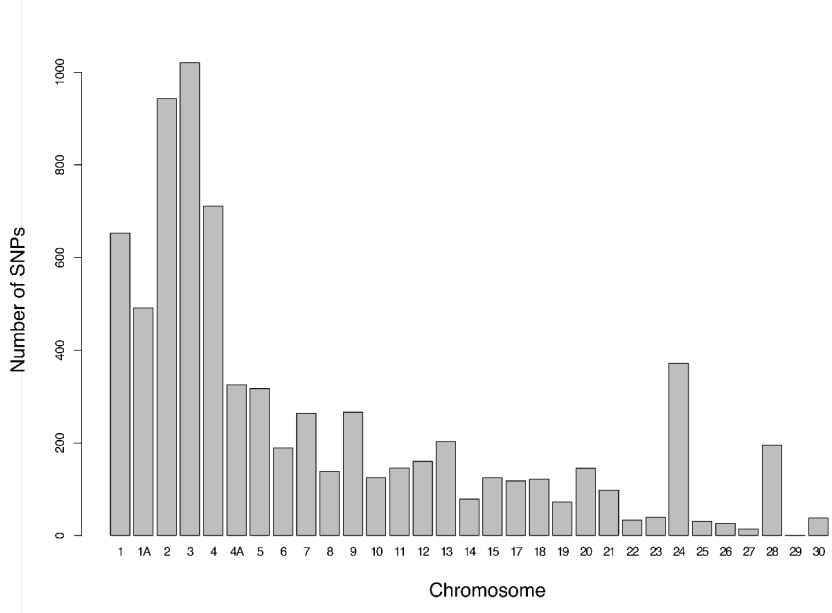


Figure S2. Number of SNPs genotyped in *Riparia diluta* and *R. riparia* populations per autosome.

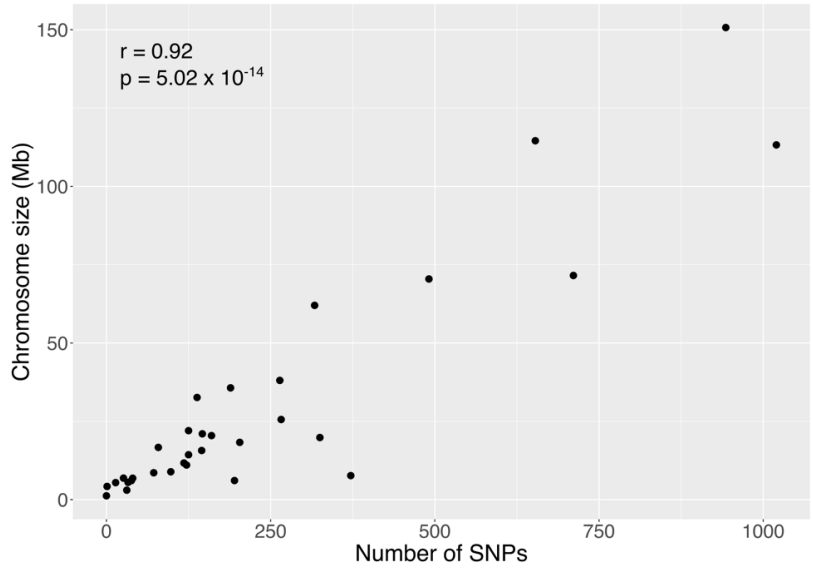


Figure S3. Relationship between the number of SNPs genotyped per autosome and chromosome size in megabases (Mb).

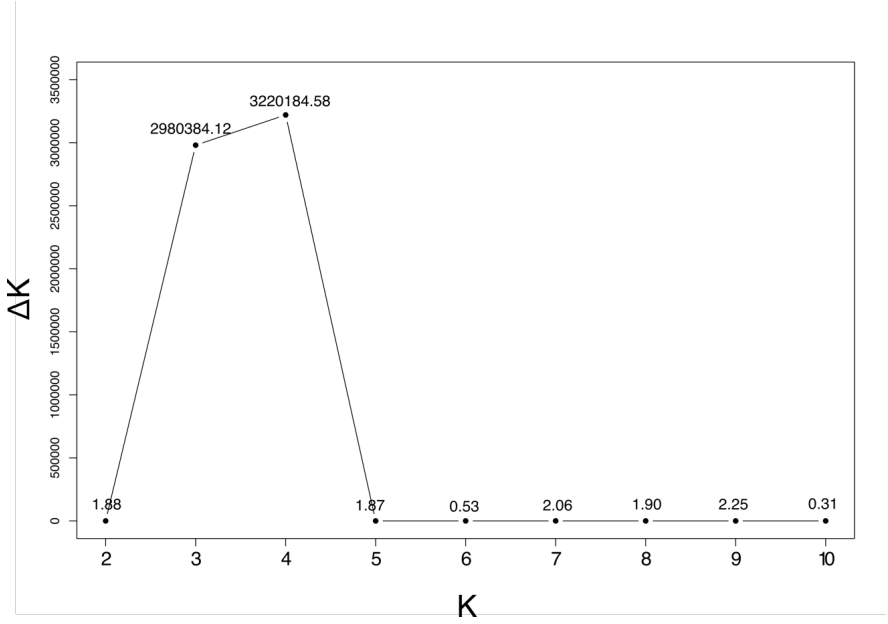


Figure S4. ΔK statistics for admixture analysis from K=2 to K=10.

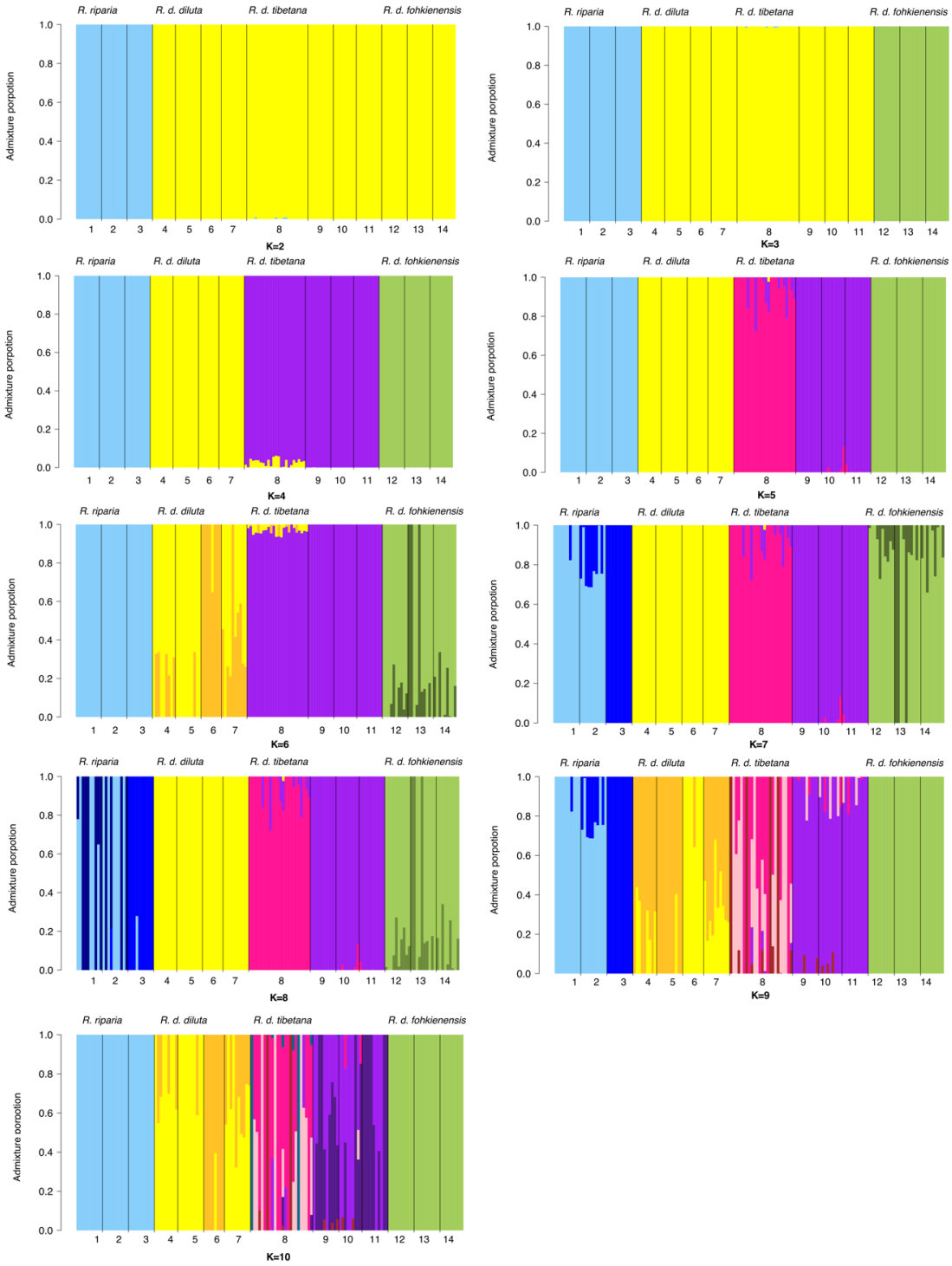


Figure S5. Individual ancestry assignment for admixture analysis from from K=2 to K=10. Numbers correspond to population IDs shown in Figure 3. Taxon labels for different populations were based on mitochondrial lineages.

Table S1 Population information. We grouped breeding colonies 8a and 8b, 9a and 9b, 10a and 10b as three populations (8, 9, 10 in Figure 3) separately since they were breeding in the same geographic region and we selected the coordinates of the first colony for analyses. Subspecies delineation is based on genetic data.

Population ID	Species	Sample size	Date of sampling	Locality	Coordinates
1	*R. riparia*	10	2019/5/29	Baisha Mountain, Buerjin,Altay, Xinjiang Uygur Autonomous Region, PR China	48°1'58''N, 86°51'41''E
2	*R. riparia*	10	2018/6/25	Dzharakhain Urto (Jaraakhain Urtuu), Mongolia	49°28.4560'N, 114°05.9451'E
3	*R. riparia*	10	2017/5/20	Xingzheng District, Zhengzhou, Henan Province, PR China	34°23'N, 113°48'E
4	*R. d. diluta*	9	2019/5/10	Alar, Xinjiang Uygur Autonomous Region, PR China	40°35.433'N, 81°43.39'E
5	*R. d. diluta*	10	2019/5/12	Khotan, Xinjiang Uygur Autonomous Region, PR China	37°11'39.51"N, 79°57'41.05''E
6	*R. d. diluta*	8	2017/5/27	Wujiaqu, Changji, Xinjiang Uygur Autonomous Region, PR China	44°22'N, 87°53'E
7	*R. d. diluta*	10	2019/5/22	Yining, Xinjiang Uygur Autonomous Region, PR China	43°54'43.92"N, 81°14'8.52"E
8a	*R. d. tibetana*	18	2018/6/21	Elsen Tasarkhai, Mongolia	47°21.5352'N, 103°41.9692'E
8b	*R. d. tibetana*	6	2018/6/18	Argalant, Mongolia	47°52.4902'N, 105°47.7577'E
9a	*R. d. tibetana*	5	2018/7/5	Zoige, Sichuan Province, PR China	33°27'12”N, 102°28'55”E
9b	*R. d. tibetana*	5	2018/6/22	Gahai Lake, Gansu Province, PR China	34°14'50‘’N, 102°20'32‘’E
10a	*R. d. tibetana*	5	2018/6/29	Qinghai Lake, Qinghai Province, PR China	36°34'15‘’N, 100°44'39‘’E
10b	*R. d. tibetana*	4	2018/6/27	Rubber Mountain, Qinghai Province, PR China	36°45'55''N, 99°38'20''E
11	*R. d. tibetana*	10	2018/6/6	Qushui County, Lhasa, Tibet autonomous region, PR China	29°20'13"N, 90°52'36''E
12	*R. d. fohkienensis*	10	2017/5/21	Meixian county, Shaanxi Province, PR China	34°14'N, 107°54'E
13	*R. d. fohkienensis*	10	2016/4/26	Yongjia Academy, Yongjia County, Wenzhou, Zhejiang Province, PR China	28°16'9.66"N, 120°40'54.89"E
14	*R. d. fohkienensis*	9	2018/4/29	Lanxi County, Yibin, Sichuan, PR China	28°48'45.90"N, 104°57'1.65"E

Table S2 Specimens sampled, museum number or sample name, collection locality, GenBank accession numbers for the of ND2 haplotype sequences and accessions for NCBI sequence read archive (SRA) under BioProject PRJNA755835 of each individual.

Species	Museum number	Locality	ND2	SRA
R. d. diluta	SYSb006513	Alar, Xinjiang, PR China	MZ747658	SAMN20856160
R. d. diluta	SYSb006515	Alar, Xinjiang, PR China	MZ747659	SAMN20856161
R. d. diluta	SYSb006517	Alar, Xinjiang, PR China	MZ747657	SAMN20856162
R. d. diluta	SYSb006518	Alar, Xinjiang, PR China	MZ747657	SAMN20856163
R. d. diluta	SYSb006521	Alar, Xinjiang, PR China	MZ747656	SAMN20856164
R. d. diluta	SYSb006522	Alar, Xinjiang, PR China	MZ747657	SAMN20856165
R. d. diluta	SYSb006755	Alar, Xinjiang, PR China	MZ747658	SAMN20856166
R. d. diluta	SYSb006756	Alar, Xinjiang, PR China	MZ747657	SAMN20856167
R. d. diluta	SYSb007331	Alar, Xinjiang, PR China	MZ747657	SAMN20856168
R. d. diluta	SYSb007336	Khotan, Xinjiang, PR China	MZ747657	SAMN20856169
R. d. diluta	SYSb007338	Khotan, Xinjiang, PR China	MZ747660	SAMN20856170
R. d. diluta	SYSb007342	Khotan, Xinjiang, PR China	MZ747657	SAMN20856171
R. d. diluta	SYSb007343	Khotan, Xinjiang, PR China	MZ747657	SAMN20856172
R. d. diluta	SYSb007345	Khotan, Xinjiang, PR China	MZ747657	SAMN20856173
R. d. diluta	SYSb007348	Khotan, Xinjiang, PR China	MZ747657	SAMN20856174
R. d. diluta	SYSb007349	Khotan, Xinjiang, PR China	MZ747657	SAMN20856175
R. d. diluta	SYSb007351	Khotan, Xinjiang, PR China	MZ747657	SAMN20856176
R. d. diluta	SYSb007352	Khotan, Xinjiang, PR China	MZ747657	SAMN20856177
R. d. diluta	SYSb007962	Khotan, Xinjiang, PR China	MZ747661	SAMN20856178
R. d. diluta	SYSb007964	Yining, Xinjiang, PR China	MZ747664	SAMN20856179
R. d. diluta	SYSb007965	Yining, Xinjiang, PR China	MZ747665	SAMN20856180
R. d. diluta	SYSb007967	Yining, Xinjiang, PR China	MZ747662	SAMN20856181

R. d. diluta	SYSb008009	Yining, Xinjiang, PR China	MZ747662	SAMN20856182
R. d. diluta	SYSb008334	Yining, Xinjiang, PR China	MZ747657	SAMN20856183
R. d. diluta	SYSb008336	Yining, Xinjiang, PR China	MZ747657	SAMN20856184
R. d. diluta	SYSb008339	Yining, Xinjiang, PR China	MZ747657	SAMN20856185
R. d. diluta	SYSb008340	Yining, Xinjiang, PR China	MZ747657	SAMN20856186
R. d. diluta	SYSb008343	Yining, Xinjiang, PR China	MZ747657	SAMN20856187
R. d. diluta	SYSb008344	Yining, Xinjiang, PR China	MZ747660	SAMN20856188
R. d. fohkienensis	SYSb003791	Yongjia, Wenzhou, Zhejiang, PR China	MZ747666	SAMN20856189
R. d. fohkienensis	SYSb003792	Yongjia, Wenzhou, Zhejiang, PR China	MZ747666	SAMN20856190
R. d. fohkienensis	SYSb003793	Yongjia, Wenzhou, Zhejiang, PR China	MZ747666	SAMN20856191
R. d. fohkienensis	SYSb003794	Yongjia, Wenzhou, Zhejiang, PR China	MZ747666	SAMN20856192
R. d. fohkienensis	SYSb003795	Yongjia, Wenzhou, Zhejiang, PR China	MZ747668	SAMN20856193
R. d. fohkienensis	SYSb003796	Yongjia, Wenzhou, Zhejiang, PR China	MZ747666	SAMN20856194
R. d. fohkienensis	SYSb003797	Yongjia, Wenzhou, Zhejiang, PR China	MZ747666	SAMN20856195
R. d. fohkienensis	SYSb003798	Yongjia, Wenzhou, Zhejiang, PR China	MZ747669	SAMN20856196
R. d. fohkienensis	SYSb003799	Yongjia, Wenzhou, Zhejiang, PR China	MZ747666	SAMN20856197
R. d. fohkienensis	SYSb003800	Yongjia, Wenzhou, Zhejiang, PR China	MZ747666	SAMN20856198
R. d. fohkienensis	SYSb006526	Meixian, Shaanxi, PR China	MZ747666	SAMN20856199
R. d. fohkienensis	SYSb006528	Meixian, Shaanxi, PR China	MZ747666	SAMN20856200
R. d. fohkienensis	SYSb006529	Meixian, Shaanxi, PR China	MZ747666	SAMN20856201
R. d. fohkienensis	SYSb006531	Meixian, Shaanxi, PR China	MZ747667	SAMN20856202
R. d. fohkienensis	SYSb006532	Meixian, Shaanxi, PR China	MZ747666	SAMN20856203
R. d. fohkienensis	SYSb006533	Meixian, Shaanxi, PR China	MZ747666	SAMN20856204
R. d. fohkienensis	SYSb006534	Meixian, Shaanxi, PR China	MZ747666	SAMN20856205
R. d. fohkienensis	SYSb006535	Meixian, Shaanxi, PR China	MZ747667	SAMN20856206

R. d. fohkienensis	SYSb006536	Meixian, Shaanxi, PR China	MZ747666	SAMN20856207
R. d. fohkienensis	SYSb006537	Meixian, Shaanxi, PR China	MZ747666	SAMN20856208
R. d. diluta	SYSb006540	Wujiaqu, Xinjiang, PR China	MZ747662	SAMN20856209
R. d. diluta	SYSb006541	Wujiaqu, Xinjiang, PR China	MZ747663	SAMN20856210
R. d. diluta	SYSb006542	Wujiaqu, Xinjiang, PR China	MZ747663	SAMN20856211
R. d. diluta	SYSb006543	Wujiaqu, Xinjiang, PR China	MZ747657	SAMN20856212
R. d. diluta	SYSb006544	Wujiaqu, Xinjiang, PR China	MZ747661	SAMN20856213
R. d. diluta	SYSb006545	Wujiaqu, Xinjiang, PR China	MZ747657	SAMN20856214
R. d. diluta	SYSb006546	Wujiaqu, Xinjiang, PR China	MZ747657	SAMN20856215
R. d. diluta	SYSb006547	Wujiaqu, Xinjiang, PR China	MZ747657	SAMN20856216
R. d. fohkienensis	SYSb008373	Lanxi, Yibin, Sichuan, PR China	MZ747671	SAMN20856217
R. d. fohkienensis	SYSb008375	Lanxi, Yibin, Sichuan, PR China	MZ747666	SAMN20856218
R. d. fohkienensis	SYSb008380	Lanxi, Yibin, Sichuan, PR China	MZ747666	SAMN20856219
R. d. fohkienensis	SYSb008381	Lanxi, Yibin, Sichuan, PR China	MZ747666	SAMN20856220
R. d. fohkienensis	SYSb008382	Lanxi, Yibin, Sichuan, PR China	MZ747666	SAMN20856221
R. d. fohkienensis	SYSb008384	Lanxi, Yibin, Sichuan, PR China	MZ747666	SAMN20856222
R. d. fohkienensis	SYSb008385	Lanxi, Yibin, Sichuan, PR China	MZ747666	SAMN20856223
R. d. fohkienensis	SYSb008388	Lanxi, Yibin, Sichuan, PR China	MZ747666	SAMN20856224
R. d. fohkienensis	SYSb008391	Lanxi, Yibin, Sichuan, PR China	MZ747670	SAMN20856225
R. d. tibetana	SYSb003782	East Huanhu Road, Qinghai, PR China	MZ747672	SAMN20856226
R. d. tibetana	SYSb003783	East Huanhu Road, Qinghai, PR China	MZ747673	SAMN20856227
R. d. tibetana	SYSb003941	Lhasa, Tibet, PR China	MZ747674	SAMN20856228
R. d. tibetana	SYSb003942	Lhasa, Tibet, PR China	MZ747679	SAMN20856229
R. d. tibetana	SYSb003943	Lhasa, Tibet, PR China	MZ747674	SAMN20856230
R. d. tibetana	SYSb003945	Lhasa, Tibet, PR China	MZ747672	SAMN20856231

R. d. tibetana	SYSb003968	Lhasa, Tibet, PR China	MZ747672	SAMN20856232
R. d. tibetana	SYSb004145	Lhasa, Tibet, PR China	MZ747677	SAMN20856233
R. d. tibetana	SYSb004151	Lhasa, Tibet, PR China	MZ747674	SAMN20856234
R. d. tibetana	SYSb004153	Lhasa, Tibet, PR China	MZ747678	SAMN20856235
R. d. tibetana	SYSb004154	Lhasa, Tibet, PR China	MZ747672	SAMN20856236
R. d. tibetana	SYSb004156	Lhasa, Tibet, PR China	MZ747672	SAMN20856237
R. d. tibetana	SYSb004163	Gahai Lake, Gansu, PR China	MZ747675	SAMN20856238
R. d. tibetana	SYSb004164	Gahai Lake, Gansu, PR China	MZ747672	SAMN20856239
R. d. tibetana	SYSb004169	Gahai Lake, Gansu, PR China	MZ747672	SAMN20856240
R. d. tibetana	SYSb004187	Gahai Lake, Gansu, PR China	MZ747676	SAMN20856241
R. d. tibetana	SYSb004192	Gahai Lake, Gansu, PR China	MZ747672	SAMN20856242
R. d. tibetana	SYSb004194	Rubber Mountain, Qinghai, PR China	MZ747674	SAMN20856243
R. d. tibetana	SYSb004197	Rubber Mountain, Qinghai, PR China	MZ747674	SAMN20856244
R. d. tibetana	SYSb004198	Rubber Mountain, Qinghai, PR China	MZ747672	SAMN20856245
R. d. tibetana	SYSb004200	Rubber Mountain, Qinghai, PR China	MZ747672	SAMN20856246
R. d. tibetana	SYSb005922	East Huanhu Road, Qinghai, PR China	MZ747674	SAMN20856247
R. d. tibetana	SYSb005923	East Huanhu Road, Qinghai, PR China	MZ747672	SAMN20856248
R. d. tibetana	SYSb005924	East Huanhu Road, Qinghai, PR China	MZ747672	SAMN20856249
R. d. tibetana	SYSb005934	Zoige, Sichuan, PR China	MZ747672	SAMN20856250
R. d. tibetana	SYSb005939	Zoige, Sichuan, PR China	MZ747680	SAMN20856251
R. d. tibetana	SYSb003656	Zoige, Sichuan PR China	MZ747674	SAMN20856252
R. d. tibetana	SYSb006084	Zoige, Sichuan, PR China	MZ747672	SAMN20856253
R. d. tibetana	SYSb006087	Zoige, Sichuan, PR China	MZ747672	SAMN20856254
R. d. tibetana	NMBE1078288	Elsen Tasarkhai, Mongolia	MZ747681	SAMN20856255
R. d. tibetana	NMBE1078289	Argalant, Mongolia	MZ747697	SAMN20856256

R. d. tibetana	NMBE1078290	Argalant, Mongolia	MZ747674	SAMN20856257
R. d. tibetana	NMBE1078291	Argalant, Mongolia	MZ747681	SAMN20856258
R. d. tibetana	NMBE1078292	Argalant, Mongolia	MZ747683	SAMN20856259
R. d. tibetana	NMBE1078293	Argalant, Mongolia	MZ747674	SAMN20856260
R. d. tibetana	NMBE1078294	Argalant, Mongolia	MZ747681	SAMN20856261
R. d. tibetana	NMBE1078295	Elsen Tasarkhai, Mongolia	MZ747674	SAMN20856262
R. d. tibetana	NMBE1078296	Elsen Tasarkhai, Mongolia	MZ747685	SAMN20856263
R. d. tibetana	NMBE1078297	Elsen Tasarkhai, Mongolia	MZ747682	SAMN20856264
R. d. tibetana	NMBE1078298	Elsen Tasarkhai, Mongolia	MZ747674	SAMN20856265
R. d. tibetana	NMBE1078299	Elsen Tasarkhai, Mongolia	MZ747683	SAMN20856266
R. d. tibetana	NMBE1078300	Elsen Tasarkhai, Mongolia	MZ747674	SAMN20856267
R. d. tibetana	NMBE1078301	Elsen Tasarkhai, Mongolia	MZ747681	SAMN20856268
R. d. tibetana	NMBE1078302	Elsen Tasarkhai, Mongolia	MZ747683	SAMN20856269
R. d. tibetana	NMBE1078303	Elsen Tasarkhai, Mongolia	MZ747683	SAMN20856270
R. d. tibetana	NMBE1078304	Elsen Tasarkhai, Mongolia	MZ747683	SAMN20856271
R. d. tibetana	NMBE1078305	Elsen Tasarkhai, Mongolia	MZ747684	SAMN20856272
R. d. tibetana	NMBE1078306	Elsen Tasarkhai, Mongolia	MZ747683	SAMN20856273
R. d. tibetana	NMBE1078307	Elsen Tasarkhai, Mongolia	MZ747674	SAMN20856274
R. d. tibetana	NMBE1078287	Elsen Tasarkhai, Mongolia	MZ747683	SAMN20856275
R. d. tibetana	NMBE1078308	Elsen Tasarkhai, Mongolia	MZ747674	SAMN20856276
R. d. tibetana	NMBE1078309	Elsen Tasarkhai, Mongolia	MZ747681	SAMN20856277
R. d. tibetana	NMBE1078310	Elsen Tasarkhai, Mongolia	MZ747674	SAMN20856278
R. riparia	SYSb008345	Baisha Mountain, Altay, Xiniiang, PR China	MZ747686	SAMN20856279
R. riparia	SYSb008346	Baisha Mountain, Altay, Xiniiang, PR China	MZ747687	SAMN20856280
R. riparia	SYSb008349	Baisha Mountain, Altay, Xiniiang, PR China	MZ747688	SAMN20856281

R. riparia	SYSb008351	Baisha Mountain, Altay, Xiniiang, PR China	MZ747688	SAMN20856282
R. riparia	SYSb008352	Baisha Mountain, Altay, Xiniiang, PR China	MZ747688	SAMN20856283
R. riparia	SYSb008353	Baisha Mountain, Altay, Xiniiang, PR China	MZ747689	SAMN20856284
R. riparia	SYSb008355	Baisha Mountain, Altay, Xiniiang, PR China	MZ747690	SAMN20856285
R. riparia	SYSb008356	Baisha Mountain, Altay, Xiniiang, PR China	MZ747688	SAMN20856286
R. riparia	SYSb008357	Baisha Mountain, Altay, Xiniiang, PR China	MZ747688	SAMN20856287
R. riparia	SYSb008358	Baisha Mountain, Altay, Xiniiang, PR China	MZ747688	SAMN20856288
R. riparia	NMBE1078312	Dzharakhain Urto (Jaraakhain Urtuu), Mongolia	MZ747688	SAMN20856289
R. riparia	NMBE1078313	Dzharakhain Urto (Jaraakhain Urtuu), Mongolia	MZ747691	SAMN20856290
R. riparia	NMBE1078314	Dzharakhain Urto (Jaraakhain Urtuu), Mongolia	MZ747692	SAMN20856291
R. riparia	NMBE1078315	Dzharakhain Urto (Jaraakhain Urtuu), Mongolia	MZ747688	SAMN20856292
R. riparia	NMBE1078316	Dzharakhain Urto (Jaraakhain Urtuu), Mongolia	MZ747693	SAMN20856293
R. riparia	NMBE1078318	Dzharakhain Urto (Jaraakhain Urtuu), Mongolia	MZ747688	SAMN20856294
R. riparia	NMBE1078319	Dzharakhain Urto (Jaraakhain Urtuu), Mongolia	MZ747693	SAMN20856295
R. riparia	NMBE1078322	Dzharakhain Urto (Jaraakhain Urtuu), Mongolia	MZ747688	SAMN20856296
R. riparia	NMBE1078323	Dzharakhain Urto (Jaraakhain Urtuu), Mongolia	MZ747693	SAMN20856297
R. riparia	NMBE1078327	Dzharakhain Urto (Jaraakhain Urtuu), Mongolia	MZ747688	SAMN20856298
R. riparia	SYSb006501	Zhengzhou, Henan, PR China	MZ747694	SAMN20856299
R. riparia	SYSb006502	Zhengzhou, Henan, PR China	MZ747695	SAMN20856300
R. riparia	SYSb006503	Zhengzhou, Henan, PR China	MZ747693	SAMN20856301
R. riparia	SYSb006504	Zhengzhou, Henan, PR China	MZ747693	SAMN20856302
R. riparia	SYSb006505	Zhengzhou, Henan, PR China	MZ747693	SAMN20856303
R. riparia	SYSb006506	Zhengzhou, Henan, PR China	MZ747696	SAMN20856304
R. riparia	SYSb006507	Zhengzhou, Henan, PR China	MZ747688	SAMN20856305
R. riparia	SYSb006508	Zhengzhou, Henan, PR China	MZ747693	SAMN20856306

R. riparia	SYSb006509	Zhengzhou, Henan, PR China	MZ747696	SAMN20856307
R. riparia	SYSb006510	Zhengzhou, Henan, PR China	MZ747693	SAMN20856308
R. d. indica	NHMUK1949.W.1.5879	Jhelum, Punjab, Pakistan	MG881167	
Hirundo rustica			DQ176515	

Table S3 Morphological data. All the measurements are shown in mm. The measurements below are bill to feathering (BF), bill depth (BP), bill width (BW), wing length (WL), length of P8 (P8), tail length (Ltail), length of tail fork (LF) and length of tarsus (LTar).

Sample ID	Species	BF	BP	BW	WL	P8	Ltail	LF	LTar	Coordinates
SYSb003656	*R. d. tibetana*	5.7	1.9	3.6	108	84	53	6.8	11.2	33°27'12N, 102°28'55E
SYSb003657	*R. d. tibetana*	6	2.2	3.3	110	85	54	7.7	10.8	29°41.273'N, 91°18.513'E
SYSb003941	*R. d. tibetana*	6.8	2.2	4.3	106	89	47	2.2	11.6	29°41.273'N, 91°18.513'E
SYSb003942	*R. d. tibetana*	5.7	2	3.5	109	84	56	4	11	29°41.273'N, 91°18.513'E
SYSb003943	*R. d. tibetana*	5.3	2.1	3.4	107	83	51	5.5	11.3	29°41.273'N, 91°18.513'E
SYSb003944	*R. d. tibetana*	5.6	2.2	3.8	107	84	53	7	11.1	29°41.273'N, 91°18.513'E
SYSb003945	*R. d. tibetana*	5.9	2.3	3.7	110	80	53	5.2	11.6	29°20'12.8”N, 90°52'36.8"E
SYSb003968	*R. d. tibetana*	5.2	2.1	3.4	106	84	52	7.7	11.1	29°20'12.8”N, 90°52'36.8"E
SYSb004104	*R. d. tibetana*	6.3	2.2	3.9	111	86	52	5.4	11	29°20'12.8”N, 90°52'36.8"E
SYSb004145	*R. d. tibetana*	5.8	2.2	3.9	110	84	48	3.1	10	29°20'12.8”N, 90°52'36.8"E
SYSb004146	*R. d. tibetana*	5.8	2.3	4.5	109	83	49	1.8	10.5	29°20'12.8”N, 90°52'36.8"E
SYSb004147	*R. d. tibetana*	5.8	2.3	4.1	108	83	50	2.2	10.8	29°20'12.8”N, 90°52'36.8"E
SYSb004148	*R. d. tibetana*	4.1	2.4	3.7	102	81	49	3	9.8	29°20'12.8”N, 90°52'36.8"E
SYSb004149	*R. d. tibetana*	5.9	2.2	2.8	107	83	49	7.8	10.8	29°20'12.8”N, 90°52'36.8"E
SYSb004150	*R. d. tibetana*	5.3	2.3	4.1	107	82	49	6.1	11.4	29°20'12.8”N, 90°52'36.8"E
SYSb004151	*R. d. tibetana*	6.1	2.5	4.2	109	83	52	4.2	10.8	29°20'12.8”N, 90°52'36.8"E
SYSb004152	*R. d. tibetana*	6	2.3	4	104	82	52	7.2	11.8	29°20'12.8”N, 90°52'36.8"E
SYSb004153	*R. d. tibetana*	6.4	2.3	4.4	112	84	58	3.8	11.7	29°20'12.8”N, 90°52'36.8"E
SYSb004154	*R. d. tibetana*	5.8	2.1	4	107	81	49	2.3	11.8	29°20'12.8”N, 90°52'36.8"E
SYSb004155	*R. d. tibetana*	6.1	2.3	4.1	110	85	50	5.9	11.8	29°20'12.8”N, 90°52'36.8"E
SYSb004156	*R. d. tibetana*	6.1	2.3	4.1	111	84	51	1.8	11.8	29°20'12.8”N, 90°52'36.8"E
SYSb004157	*R. d. tibetana*	6.1	2.2	4.5	110	83.5	52	4.1	11.6	34°14'49.6"N, 102°20'31.5"E

SYSb004158	*R. d. tibetana*	6.1	2.3	3.8	109	85	53	5	11.9	34°14'49.6"N, 102°20'31.5"E
SYSb004159	*R. d. tibetana*	5.4	2	3.6	108	84	51	6.2	11.7	34°14'49.6"N, 102°20'31.5"E
SYSb004160	*R. d. tibetana*	5.8	2.1	3.8	116	87	53	5.5	11.7	34°14'49.6"N, 102°20'31.5"E
SYSb004161	*R. d. tibetana*	5.9	2.2	4.3	111	84	53	3.1	11.7	34°14'49.6"N, 102°20'31.5"E
SYSb004162	*R. d. tibetana*	6.2	2.3	4.5	114	87	55	4.9	11.8	34°14'49.6"N, 102°20'31.5"E
SYSb004163	*R. d. tibetana*	6.1	2.2	4.3	103	83	51	4.4	11.8	34°14'49.6"N, 102°20'31.5"E
SYSb004164	*R. d. tibetana*	6.1	2.2	4.1	111	83	51	4.9	11.8	34°14'49.6"N, 102°20'31.5"E
SYSb004165	*R. d. tibetana*	5.8	2	4.5	109	86	53	7.6	11.7	34°14'49.6"N, 102°20'31.5"E
SYSb004166	*R. d. tibetana*	5.7	2.3	4.2	111	83	53	5.4	11.2	34°14'49.6"N, 102°20'31.5"E
SYSb004167	*R. d. tibetana*	6.2	2.1	4	107	85	51	4.5	9.9	34°14'49.6"N, 102°20'31.5"E
SYSb004168	*R. d. tibetana*	5.2	2	4.6	109	84	51	3.7	11.7	34°14'49.6"N, 102°20'31.5"E
SYSb004169	*R. d. tibetana*	6.2	2.2	4.5	108	83	51	7.2	11.8	34°14'49.6"N, 102°20'31.5"E
SYSb004170	*R. d. tibetana*	6.3	2.1	3.7	112	84	51	3.2	11.8	34°14'49.6"N, 102°20'31.5"E
SYSb004187	*R. d. tibetana*	6.1	2.2	3.7	113	88	55	6	11.7	34°14'49.6"N, 102°20'31.5"E
SYSb004188	*R. d. tibetana*	5.4	2.2	4	112	85	50	3.9	11.7	34°14'49.6"N, 102°20'31.5"E
SYSb004189	*R. d. tibetana*	5.1	2.5	4.1	114	87	52	2	11.7	34°14'49.6"N, 102°20'31.5"E
SYSb004190	*R. d. tibetana*	6.1	2.2	4	111	83	50	5.2	11.7	34°14'49.6"N, 102°20'31.5"E
SYSb004191	*R. d. tibetana*	6.1	2	3.4	109	86	52	4.2	11.8	34°14'49.6"N, 102°20'31.5"E
SYSb004192	*R. d. tibetana*	6.2	2.2	3.2	110	85	50	5.8	11.8	34°14'49.6"N, 102°20'31.5"E
SYSb004193	*R. d. tibetana*	5.9	2.2	3.5	113	87	52	3	11.8	34°14'49.6"N, 102°20'31.5"E
SYSb004194	*R. d. tibetana*	6.5	2.2	4	107	82	51	5.3	11.7	36°45'54.9"N, 99°38'20.1"E
SYSb004195	*R. d. tibetana*	5.2	2.3	3.7	107	82	52	6.1	10.9	36°45'54.9"N, 99°38'20.1"E
SYSb004196	*R. d. tibetana*	5.5	2.1	3.3	114	85	53	6.7	10	36°45'54.9"N, 99°38'20.1"E
SYSb004197	*R. d. tibetana*	6.1	2.1	3	114	88	55	6.8	11.7	36°45'54.9"N, 99°38'20.1"E
SYSb004198	*R. d. tibetana*	6.1	2.1	3.1	108	84	52	4.2	11.8	36°45'54.9"N, 99°38'20.1"E

SYSb004199	*R. d. tibetana*	5.8	2.3	3.8	112	85.5	55	6.4	11.9	36°45'54.9"N, 99°38'20.1"E
SYSb004200	*R. d. tibetana*	6.7	2.3	4	108	85	54	8.4	11.7	36°45'54.9"N, 99°38'20.1"E
SYSb004365	*R. d. tibetana*	6.1	2.2	3.7	111	86.5	52	5.2	10.9	36°45'54.9"N, 99°38'20.1"E
SYSb005920	*R. d. tibetana*	6.2	2.1	3.4	111	86	52	3.1	11.3	36°45'54.9"N, 99°38'20.1"E
SYSb005921	*R. d. tibetana*	6.1	2.2	3.7	113.5	88	52	2.8	11.6	36°45'54.9"N, 99°38'20.1"E
SYSb005922	*R. d. tibetana*	6.2	2.1	3.7	107	82	50	6.6	9.8	36° 34'15.3"N, 100°44'38.7"E
SYSb005923	*R. d. tibetana*	5.9	2.3	3.7	111	84	49	5.5	11.7	36° 34'15.3"N, 100°44'38.7"E
SYSb005925	*R. d. tibetana*	6.1	2.2	3.1	106	82	50	5.5	11.2	36°50'10.1"N, 99°43'14.2"E
SYSb005926	*R. d. tibetana*	6	2.3	4	113	87	54	6.1	11.7	36°50'10.1"N, 99°43'14.2"E
SYSb005927	*R. d. tibetana*	6	2.3	3.3	116	83	52	5.5	11	36°50'10.1"N, 99°43'14.2"E
SYSb005928	*R. d. tibetana*	6	2.2	3.5	110	84.5	51	4.9	11.2	36°50'10.1"N, 99°43'14.2"E
SYSb005929	*R. d. tibetana*	6	2.2	4.1	109.5	83.5	52	2.5	11.7	36°50'10.1"N, 99°43'14.2"E
SYSb005930	*R. d. tibetana*	6.2	2.2	4.3	111	85	52	5.3	11.8	33°34'1.1"N, 102°28'35.4"E
SYSb005931	*R. d. tibetana*	5.9	2	4.8	112	83.5	50	2.3	11.8	33°34'1.1"N, 102°28'35.4"E
SYSb005932	*R. d. tibetana*	5.7	2.3	3.5	113	85	54	3	11.2	33°34'1.1"N, 102°28'35.4"E
SYSb005933	*R. d. tibetana*	6.3	2.5	3.8	109	83	52	4.5	11.8	33°34'1.1"N, 102°28'35.4"E
SYSb005934	*R. d. tibetana*	5.9	2.4	3.8	106	78	50	4.7	11.2	33°34'1.1"N, 102°28'35.4"E
SYSb005935	*R. d. tibetana*	5.9	2.1	4.1	112	85	53	6.5	11.3	33°34'1.1"N, 102°28'35.4"E
SYSb005936	*R. d. tibetana*	6.7	2.2	3.7	112	86	52	6.4	11.7	33°34'1.1"N, 102°28'35.4"E
SYSb005937	*R. d. tibetana*	6	2.2	3.8	111	80	54	5.7	11.7	33°34'1.1"N, 102°28'35.4"E
SYSb005938	*R. d. tibetana*	6.1	2.2	3.2	119	82	51	5.6	11.8	33°34'1.1"N, 102°28'35.4"E
SYSb005939	*R. d. tibetana*	6.4	2.3	3.4	114	86	55	5.3	11.5	33°34'1.1"N, 102°28'35.4"E
SYSb006050	*R. d. tibetana*	6.8	2.4	3.9	108	88	51	2.8	11.7	33°34'1.1"N, 102°28'35.4"E
SYSb006051	*R. d. tibetana*	6.1	2	3.5	116	90	56	4.4	11.7	33°34'1.1"N, 102°28'35.4"E
SYSb006053	*R. d. tibetana*	6.2	2.1	4.2	109	82	51	4.1	11.7	33°34'1.1"N, 102°28'35.4"E

SYSb006084	*R. d. tibetana*	6.2	2.4	3.8	114	83	52	5.1	11.4	33°34'1.1"N, 102°28'35.4"E
SYSb006085	*R. d. tibetana*	6.4	2.2	3.6	118	83	49	4.2	11.7	33°34'1.1"N, 102°28'35.4"E
SYSb006086	*R. d. tibetana*	6.2	2.1	3.8	110	83	50	5.2	11.7	33°34'1.1"N, 102°28'35.4"E
SYSb006087	*R. d. tibetana*	6.7	2.3	4	111	85	52	5.6	11.7	33°34'1.1"N, 102°28'35.4"E
SYSb006088	*R. d. tibetana*	6.3	2	4.1	110	85	55	5.9	11.6	33°34'1.1"N, 102°28'35.4"E
SYSb006089	*R. d. diluta*	4.89	1.8	3.39	107	80.5	51	5.9	11.78	44°21'49.85' N, 87°53'11.77''E
SYSb006090	*R. d. diluta*	6.7	2.2	3.7	106	80.05	47	6.08	10.61	44°21'49.85' N, 87°53'11.77''E
SYSb006091	*R. d. diluta*	6.8	2.2	3.9	105.5	81	51	8.8	10.1	44°21'49.85' N, 87°53'11.77''E
SYSb006092	*R. d. diluta*	5.7	1.9	3.9	102	79	53	7.7	10.65	44°21'49.85' N, 87°53'11.77''E
SYSb006093	*R. d. diluta*	5.9	2.5	3.59	109.5	82	52.5	7.1	10.9	44°21'49.85' N, 87°53'11.77''E
SYSb006094	*R. d. diluta*	5.8	2.02	3.73	105	80.5	52.5	5.9	10.3	44°21'49.85' N, 87°53'11.77''E
SYSb006095	*R. d. diluta*	5.7	2	3.95	109.5	83	52	7.8	11.35	44°21'49.85' N, 87°53'11.77''E
SYSb006096	*R. d. diluta*	6.15	1.75	3.6	106	80.5	51.5	5.75	10.5	44°21'49.85' N, 87°53'11.77''E
SYSb006098	*R. d. diluta*	6	1.9	3.9	104.5	79.5	51.5	6.8	10.7	44°21'49.85' N, 87°53'11.77''E
SYSb006099	*R. d. diluta*	5.1	1.9	3	105	78.5	49.5	6.3	8.8	44°21'49.85' N, 87°53'11.77''E
SYSb006100	*R. d. diluta*	5.95	2	3.75	105	81	51.5	7	92.5	44°21'49.85' N, 87°53'11.77''E
SYSb006514	*R. d. diluta*	5.7	1.9	3.45	108	82.5	52.5	7.15	10.7	40°35.433' N, 81°43.39'E
SYSb006515	*R. d. diluta*	5.2	1.85	3.1	105.5	82	52	8.85	10.6	40°35.433' N, 81°43.39'E
SYSb006516	*R. d. diluta*	5.8	2.1	3.5	104	79.5	51.5	7.95	10.4	40°35.433' N, 81°43.39'E
SYSb006517	*R. d. diluta*	6.1	2.25	4.05	109.5	83.5	54	9.15	10.7	40°35.433' N, 81°43.39'E
SYSb006518	*R. d. diluta*	5.8	2	3.6	105	80	49.5	3.8	10.3	40°35.433' N, 81°43.39'E
SYSb006520	*R. d. diluta*	5.9	2.05	3.25	105	80	47.5	6.75	10	40°35.433' N, 81°43.39'E
SYSb006521	*R. d. diluta*	6	1.95	4.45	104	80	42	5.85	9.65	40°35.433' N, 81°43.39'E
SYSb006522	*R. d. diluta*	6	1.9	3.45	100.5	76.5	48	7.8	9.7	40°35.433' N, 81°43.39'E
SYSb006523	*R. d. diluta*	5.7	1.6	3.3	104	78	45.5	6.75	10	40°35.433' N, 81°43.39'E

SYSb006755	*R. d. diluta*	5.3	1.95	4	107	82.5	51	9.6	9.8	40°35.433' N, 81°43.39'E
SYSb006756	*R. d. diluta*	5.7	1.9	3.5	103	73.5	49	7.75	9.7	40°35.433' N, 81°43.39'E
SYSb006757	*R. d. diluta*	6	1.75	3.5	102.5	79.5	47	6.7	10.7	40°35.433' N, 81°43.39'E
SYSb006758	*R. d. diluta*	5.7	1.9	3.85	104	79	49	6.15	10.15	40°35.433' N, 81°43.39'E
SYSb006759	*R. d. diluta*	5.2	1.75	3.65	105.5	81	44.5	5.2	10.1	40°35.433' N, 81°43.39'E
SYSb007331	*R. d. diluta*	6	1.9	2.95	101	78	47.5	6.3	10.4	40°35.433' N, 81°43.39'E
SYSb007332	*R. d. diluta*	5.4	1.8	3.75	104.5	77.5	50	7.7	10.7	40°35.433' N, 81°43.39'E
SYSb007333	*R. d. diluta*	5.7	1.75	3.4	100	79.5	49	5.1	10.85	40°35.433' N, 81°43.39'E
SYSb007334	*R. d. diluta*	5.8	1.6	3.2	106	81	49	5.8	9.5	40°35.433' N, 81°43.39'E
SYSb007335	*R. d. diluta*	5.6	1.8	3.8	105	81.5	46	7.7	10.5	40°35.433' N, 81°43.39'E
SYSb007336	*R. d. diluta*	5.7	2	3.4	103	78.5	48	5.8	10.35	37°11'39.51" N, 79°57'41.05''E
SYSb007337	*R. d. diluta*	6.2	1.9	3.2	105	83	52	8.2	9.9	37°11'39.51" N, 79°57'41.05''E
SYSb007338	*R. d. diluta*	5.2	2.1	4	99.5	75.5	45	5.8	10.7	37°11'39.51" N, 79°57'41.05''E
SYSb007340	*R. d. diluta*	6	1.9	4.15	107	81	52	7.7	10.8	37°11'39.51" N, 79°57'41.05''E
SYSb007341	*R. d. diluta*	5.4	2	3.35	106.5	82	48	6	10.3	37°11'39.51" N, 79°57'41.05''E
SYSb007342	*R. d. diluta*	5.1	2.1	3.4	104	79	48	6.1	10.8	37°11'39.51" N, 79°57'41.05''E
SYSb007343	*R. d. diluta*	6.2	2	4	103.5	77.5	47	7.9	11.1	37°11'39.51" N, 79°57'41.05''E
SYSb007344	*R. d. diluta*	6.9	2.3	3.8	109	82.5	51	8.7	10.2	37°11'39.51" N, 79°57'41.05''E
SYSb007345	*R. d. diluta*	6.4	2.1	3.7	107	82	53	6.1	10	37°11'39.51" N, 79°57'41.05''E
SYSb007346	*R. d. diluta*	6.2	2.1	4.2	105.5	79	49	6.9	10.7	37°11'39.51" N, 79°57'41.05''E
SYSb007347	*R. d. diluta*	6.3	2.1	3.7	104	78	50	7.7	10.2	37°11'39.51" N, 79°57'41.05''E
SYSb007348	*R. d. diluta*	5.9	1.9	3.9	101	78	45	4	10	37°11'39.51" N, 79°57'41.05''E
SYSb007349	*R. d. diluta*	6.2	1.95	3.9	102	78	49	7.8	10.2	37°11'39.51" N, 79°57'41.05''E
SYSb007350	*R. d. diluta*	5.7	2	3.4	104	80.5	49	5.9	10.3	37°11'39.51" N, 79°57'41.05''E
SYSb007351	*R. d. diluta*	6.8	1.9	4.1	105	81.5	53.5	8.5	10.3	37°11'39.51" N, 79°57'41.05''E

SYSb007352	*R. d. diluta*	5.6	2	3.5	102.5	77.8	48	7.8	9.9	37°11'39.51" N, 79°57'41.05''E
SYSb007353	*R. d. diluta*	6.7	2	3.8	104	79	47	6.8	10.2	37°11'39.51" N, 79°57'41.05''E
SYSb007354	*R. d. diluta*	5.7	1.9	3.8	104	81	51	6.7	10.8	37°11'39.51" N, 79°57'41.05''E
SYSb007962	*R. d. diluta*	6.1	2	4.2	106	79.5	53	6.9	10.2	37°11'39.51" N, 79°57'41.05''E
SYSb007963	*R. d. diluta*	6.2	2	4.15	104	78	47	6.7	10.2	37°11'39.51" N, 79°57'41.05''E
SYSb007964	*R. d. diluta*	6.2	2.1	3.5	110	82	56	6.8	10.6	43°54'43.9"N, 81°14'8.5"E
SYSb007965	*R. d. diluta*	5.8	2.2	3.55	104	79	44	8	10.2	43°54'43.9"N, 81°14'8.5"E
SYSb007966	*R. d. diluta*	5.9	2.1	3.7	108	77	50.5	7.05	10.8	43°54'43.9"N, 81°14'8.5"E
SYSb007967	*R. d. diluta*	6.2	2.05	3.55	104	81	47	7.2	10.8	43°54'43.9"N, 81°14'8.5"E
SYSb007969	*R. d. diluta*	6	2.2	4.4	102	76.5	46	5.3	10.2	43°54'43.9"N, 81°14'8.5"E
SYSb008009	*R. d. diluta*	6.8	2.1	3.4	107	82	51	6.45	10.5	43°54'43.9"N, 81°14'8.5"E
SYSb008010	*R. d. diluta*	5.5	2	3.5	108	83	49	5.5	10.7	43°54'43.9"N, 81°14'8.5"E
SYSb008011	*R. d. diluta*	5.6	1.9	3.8	100	75	47	5.95	10.5	43°54'43.9"N, 81°14'8.5"E
SYSb008012	*R. d. diluta*	5.7	1.9	3.45	106	81.5	52	8.95	10.7	43°54'43.9"N, 81°14'8.5"E
SYSb008334	*R. d. diluta*	5.7	2	4.2	106	79	51	7	10.9	43°54'43.9"N, 81°14'8.5"E
SYSb008335	*R. d. diluta*	6.4	2.05	3.3	102	77.5	49	6.1	10.1	43°54'43.9"N, 81°14'8.5"E
SYSb008336	*R. d. diluta*	6.2	2.05	3.4	104.5	81	46.5	6.7	10.3	43°54'43.9"N, 81°14'8.5"E
SYSb008337	*R. d. diluta*	5.2	1.9	3.3	104	79.5	52.5	9.4	10.9	43°54'43.9"N, 81°14'8.5"E
SYSb008338	*R. d. diluta*	6.2	2	3	106	81.5	48	5.8	10.7	43°54'43.9"N, 81°14'8.5"E
SYSb008339	*R. d. diluta*	6.1	1.9	3.7	109	83	51	7.1	10.9	43°54'43.9"N, 81°14'8.5"E
SYSb008340	*R. d. diluta*	6	2.1	3.9	107	84	52	6.1	10.7	43°54'43.9"N, 81°14'8.5"E
SYSb008341	*R. d. diluta*	6.1	2.1	3.5	104	80	51	6.2	11.5	43°54'43.9"N, 81°14'8.5"E
SYSb008342	*R. d. diluta*	5.2	2.1	3.4	104.5	80.5	52.5	9.3	10.8	43°54'43.9"N, 81°14'8.5"E
SYSb008343	*R. d. diluta*	5.5	2	2.9	105	81	47	6.5	10.8	43°54'43.9"N, 81°14'8.5"E
SYSb008344	*R. d. diluta*	3.7	1.9	3.4	107	82	53.5	8.7	10.9	43°54'43.9"N, 81°14'8.5"E

SYSb008370	*R. d. fohkienensis*	5.55	2.23	3.14	99.3	78.5	48.2	2.55	10.79	28°48'45.90"N, 104°57'1.65"E
SYSb008371	*R. d. fohkienensis*	5.58	2.22	3.83	100	75.5	43.5	4.01	11.32	28°48'45.90"N, 104°57'1.65"E
SYSb008372	*R. d. fohkienensis*	5.7	2.17	3.67	99.2	77.8	45.5	2.3	11.29	28°48'45.90"N, 104°57'1.65"E
SYSb008373	*R. d. fohkienensis*	5.46	1.95	3.46	100	76	42.8	3.16	12.5	28°48'45.90"N, 104°57'1.65"E
SYSb008374	*R. d. fohkienensis*	5.83	2.17	3.24	94.5	72.5	46	2.99	10.84	28°48'45.90"N, 104°57'1.65"E
SYSb008375	*R. d. fohkienensis*	5.77	2.17	3.8	96	76	45	2.3	10.71	28°48'45.90"N, 104°57'1.65"E
SYSb008376	*R. d. fohkienensis*	5.88	2.33	3.19	96	74	44	2.34	10.82	28°48'45.90"N, 104°57'1.65"E
SYSb008377	*R. d. fohkienensis*	5.9	1.93	3.5	98	78	43	3.1	10.7	28°48'45.90"N, 104°57'1.65"E
SYSb008378	*R. d. fohkienensis*	5.8	2.06	3.69	100	76.5	47	2.1	11.13	28°48'45.90"N, 104°57'1.65"E
SYSb008379	*R. d. fohkienensis*	5.45	2.05	3.23	98	75	47	3.23	10.59	28°48'45.90"N, 104°57'1.65"E
SYSb008380	*R. d. fohkienensis*	5.36	1.93	3.3	95	76	45	3.49	10.94	28°48'45.90"N, 104°57'1.65"E
SYSb008381	*R. d. fohkienensis*	5.84	2.15	3.5	94	75.5	47	3.48	10.98	28°48'45.90"N, 104°57'1.65"E
SYSb008382	*R. d. fohkienensis*	5.61	2.1	3.49	100	76	47.5	2.9	10.21	28°48'45.90"N, 104°57'1.65"E
SYSb008383	*R. d. fohkienensis*	5.44	2.06	3.24	97	74	43	3.49	11.27	28°48'45.90"N, 104°57'1.65"E
SYSb008384	*R. d. fohkienensis*	5.84	2.13	3.49	100	78	47	3.88	11.49	28°48'45.90"N, 104°57'1.65"E
SYSb008385	*R. d. fohkienensis*	5.6	2.14	3.34	96	76	46	2.53	11.19	28°48'45.90"N, 104°57'1.65"E
SYSb008386	*R. d. fohkienensis*	5.83	2.15	3.26	99.5	79	47	3.93	11.93	28°48'45.90"N, 104°57'1.65"E
SYSb008387	*R. d. fohkienensis*	5.43	1.94	3.1	93	75.5	43	1.39	11.01	28°48'45.90"N, 104°57'1.65"E
SYSb008388	*R. d. fohkienensis*	5.81	2.23	3.01	95	76	43.5	3.13	10.7	28°48'45.90"N, 104°57'1.65"E
SYSb008389	*R. d. fohkienensis*	5.4	2.01	3.21	90	69.5	41	1.95	10.51	28°48'45.90"N, 104°57'1.65"E
SYSb008390	*R. d. fohkienensis*	5.72	2.25	3.36	104	78	47	2.51	11.15	28°48'45.90"N, 104°57'1.65"E
SYSb008391	*R. d. fohkienensis*	5.92	2.03	3.21	96	77	46	3.54	12.35	28°48'45.90"N, 104°57'1.65"E
NMBE1078288	*R. d. tibetana*	5.7	2.8	3.5	110	86	56	5.6	10.3	47°52.4902'N, 105°47.7577'E
NMBE1078289	*R. d. tibetana*	5.2	2.1	3.1	103.5	82	47	3.1	10.9	47°52.4902'N, 105°47.7577'E
NMBE1078290	*R. d. tibetana*	5.8	2.8	3.9	110	86	55	6.3	11.6	47°52.4902'N, 105°47.7577'E

NMBE1078291	*R. d. tibetana*	5.9	2.4	4.7	109	85	55	5.4	11.3	47°52.4902'N, 105°47.7577'E
NMBE1078292	*R. d. tibetana*	6	2.2	3.5	109.5	80	50	5.5	10.2	47°52.4902'N, 105°47.7577'E
NMBE1078293	*R. d. tibetana*	5.7	2.3	4	101.5	79	48	4	10.8	47°52.4902'N, 105°47.7577'E
NMBE1078297	*R. d. tibetana*	5.9	2	3.8	106	83	44	6	11.8	47°21.5352'N, 103°41.9692'E
NMBE1078298	*R. d. tibetana*	6.1	2.5	4.6	103.5	83	55.5	6.2	9.6	47°21.5352'N, 103°41.9692'E
NMBE1078299	*R. d. tibetana*	5.6	2.6	3.7	109	83	45	3.2	10	47°21.5352'N, 103°41.9692'E
NMBE1078300	*R. d. tibetana*	5.7	2.3	4.5	105	82	45	8.2	11	47°21.5352'N, 103°41.9692'E
NMBE1078301	*R. d. tibetana*	5.6	2.5	4.3	104	79.5	46.5	6.5	11.8	47°21.5352'N, 103°41.9692'E
NMBE1078302	*R. d. tibetana*	4.9	2.9	4.1	108	84	51.5	5.4	11.9	47°21.5352'N, 103°41.9692'E
NMBE1078303	*R. d. tibetana*	6.3	2.3	3.9	105	84	51	9.9	12.2	47°21.5352'N, 103°41.9692'E
NMBE1078304	*R. d. tibetana*	6	2.6	3.9	105	82.5	50	7.9	11.3	47°21.5352'N, 103°41.9692'E
NMBE1078305	*R. d. tibetana*	5.5	3	3.8	106	82	48	6.8	11.2	47°21.5352'N, 103°41.9692'E
NMBE1078306	*R. d. tibetana*	5.6	3	4.6	108	84.5	53	7	11.5	47°21.5352'N, 103°41.9692'E
NMBE1078307	*R. d. tibetana*	6.7	3	4	110	86	52	5.5	11.6	47°21.5352'N, 103°41.9692'E
NMBE1078287	*R. d. tibetana*	5.7	3.1	3.4	105	84	53	8.1	10.3	47°21.5352'N, 103°41.9692'E
NMBE1078308	*R. d. tibetana*	6	2.5	4.3	106	81.5	54	5	11.5	47°21.5352'N, 103°41.9692'E
NMBE1078309	*R. d. tibetana*	6.1	2.7	3.8	103	84	44	3.9	12.3	47°21.5352'N, 103°41.9692'E
NMBE1078310	*R. d. tibetana*	5.7	3.1	3.9	100.5	80	49.5	6.4	11.1	47°21.5352'N, 103°41.9692'E

Tale S4 Pairwise F_{ST} (below diagonal) and geographic distances (km, above diagonal) between populations. * indicate significant value $p < 0.05$.

	4	5	6	7	8	9	10	11	12	13	14
4	-	407	657	371	1908	1951	1612	1501	2412	3797	2486
5	0.04369*	-	1039	754	2246	2046	1749	1337	2540	3912	2504
6	0.05371*	0.00185*	-	534	1270	1674	1302	1689	2052	3411	2295
7	0.03987*	-0.00256	-0.0031	-	1786	2110	1748	1830	2531	3906	2692
8	0.10184*	0.11141*	0.11219*	0.10215*	-	1460	1223	2286	1499	2581	2061
9	0.12695*	0.13138*	0.1241*	0.11703*	0.01405	-	372	1214	512	1865	651
10	0.15092*	0.13882*	0.13465*	0.1278*	0.01946*	-0.00199	-	1160	801	2185	1012
11	0.14991*	0.14103*	0.138*	0.13131*	0.01997*	-0.00230	-0.00651	-	1700	2905	1371
12	0.30626*	0.34647*	0.34682*	0.3293*	0.25959*	0.28386*	0.31289*	0.31319*	-	1384	663
13	0.32188*	0.35941*	0.36299*	0.34434*	0.26795*	0.29988*	0.32728*	0.32722*	0.00637	-	1540
14	0.30901*	0.34877*	0.34831*	0.33265*	0.26096*	0.28648*	0.31375*	0.31441*	0.00476*	0.00198*	-

Table S5 Factor loadings of principal component analysis (PCA) based on eight morphometric measurements. Only loadings of the first two components (Comp.1 and Comp.2) are shown.

Measurements	Comp.1	Comp.2
Bill to Feathering	0.215	-0.005
Bill Depth	0.269	-0.413
Bill Width	0.280	0.206
Wing Length	0.498	-0.161
P8 (length of third outermost primary)	0.507	0.057
Tail Length	0.452	0.287
Length of Tail Fork	0.102	0.668
Length of Tarsus	0.288	-0.478

Chapter 2

The Hourglass Runs at Different Speeds in *Riparia* collared sand martins: Phylogenomic History and Diversification of a Cryptic Radiation

Qindong Tang[1,2], Nilofar Alayee[3], Yang Liu[4], Gombobaatar Sundev[5], Hadoram Shirihai[2], Gerald Heckel[1], Manuel Schweizer[1,2*]

[1] *Institute of Ecology and Evolution, University of Bern, Bern, Switzerland*;
[2] *Natural History Museum, Bern, Switzerland*;
[3] *Population Ecology Group, Institute of Ecology and Evolution, Friedrich Schiller University Jena, Dornburger Str. 159, 07743 Jena, Germany*;
[4] *State Key Laboratory of Biocontrol, College of Ecology School of Life Science, Sun Yat-sen University, Guangzhou, China*;
[5] *National University of Mongolia and Mongolian Ornithological Society, Ulaanbaatar, Mongolia*

* *Correspondence to be sent to: Natural History Museum, Bern, Switzerland; Institute of Ecology and Evolution, University of Bern, Bern, Switzerland; E-mail: manuel.schweizer@nmbe.ch*

Gerald Heckel and Manuel Schweizer are joint senior authors

Abstract

Aim: Biogeographic patterns of morphologically cryptic taxa are difficult to explain without detailed knowledge on their diversification history and past demography. Genomic information can provide a robust basis for disentangling the temporal and spatial components of cryptic radiation processes, yet this potential has only been exploited for select organisms. Here, based on genome resequencing data, we reconstructed the diversification history of a cryptic bird radiation.

Location: Eurasia

Taxon: Avian genus *Riparia*

Methods: We analyzed resequencing data of 24 individuals including all Eurasian *Riparia* taxa using multi-pronged phylogenomic and demographic analyses and tested for historical gene flow among lineages during their diversification.

Results: Despite morphological similarity, genomic divergence was dated to more than two million years between accepted species, and distribution range variation within the genus is reflected by different demographic trajectories. The collared sand martin *Riparia riparia* lacks pronounced phylogeographic structure in mtDNA across its Holarctic range but nuclear genomes resolved evolutionary lineages broadly consistent with subspecies designation. Demographic reconstruction showed that its large range resulted from long-lasting population growth during the Late Pleistocene rather than a recent expansion after the Last Glacial Maximum as previously proposed. Strong genomic divergence in the pale sand martin *Riparia diluta* supported the existence of multiple morphologically cryptic evolutionary lineages without gene flow between some and distinct demographic histories consistent with heterogeneous climate in their Asian ranges. The clade comprising the taxa *fohkienensis* and *indica* represents a unique phylogeographic link in birds between subtropical China and the dry northern Indian Subcontinent.

Main conclusion: We hypothesize that pronounced dispersal propensity in the strongly migratory nominate form of collared sand martin has hindered lineage

divergence across its vast distribution range, while differential seasonal migratory behaviour contributes to maintain evolutionary diversity in the pale sand martin complex. The extended levels of genomic divergence within the latter might indicate the existence of two species level taxa.

[Cryptic species; biogeography; demographic history; species tree, collared sand martin, phylogenomics]

1. INTRODUCTION

Interspecific variation in the size of distribution ranges is enormous, with some species inhabiting a virtually cosmopolitan range while others are restricted to a very limited geographic area with particular habitat (Gaston 2003). In general, most species have relatively small ranges, while very large geographic distributions are rather rare (Gaston 1998). Abiotic and biotic factors and their interactions primarily limit species' ranges (e.g. Sexton et al. 2009), but historical processes such as climate-related environmental change and associated range dynamics continue to affect distribution patterns observed today. Understanding the genetic and evolutionary outcomes of past demographic processes is crucial for assessing the potential consequences of the current global climate change on biodiversity (Pauls et al. 2013).

Range dynamics leave long-lasting signatures in the genomic constitution of organisms, and with the advent of genomic sequencing technologies, it is now possible to get a profound picture of the demographic histories of populations (Li and Durbin 2011; Nadachowska-Brzyska et al. 2015). Modelling demographic changes through time in combination with phylogenomic reconstruction provides thus a promising avenue to better understand the impact of past environmental changes on current biogeographic patterns. However, we have only begun to exploit this potential for few species (e.g. Brüniche-Olsen et al., 2021; Chattopadhyay et al., 2019; Pujolar et al., 2022; Taylor et al., 2021a) and our understanding of the impact of past climate change on historical demography and range evolution is still incomplete.

Climate cycles of the last several hundred thousand years comprising the late Pleistocene were accompanied by pronounced environmental alterations that resulted in repeated spatio-temporal changes in the distribution of habitats. Large ranges of extant species are often linked to population expansions after the Last Glacial Maximum (LGM, 0.023 – 0.018 million years ago (Ma)), when suitable habitats became available and were rapidly colonized. However, range dynamics are complex as the strength and nature of species' demographic reactions to climate and associated environment fluctuations are influenced by their ecological specialisation and their ability to track their ecological niche through space by following shifting

climate zones and habitats. The classical pattern of range contraction during glacials and expansions during interglacials might hold mainly for boreal and mesic temperate species, whereas specialists of open habitat or arid areas for example, were shown to have expanded their ranges during cold and dry glacial periods (LGP, 0.11 – 0.012 Ma) (Garcia et al. 2011a, 2011b; Kearns et al. 2014; Alaei Kakhki et al. 2018; Baca et al. 2023).

As a consequence of migratory lifestyles, species could have maintained large populations despite considerable range shifts by tracking ephemeral resources during the Last Glacial Period (LGP, 0.11 – 0.012 Ma) (Thorup et al. 2021). In general, populations of migratory species are expected to be less differentiated than those of residential ones and those of generalists less than specialists (Hung et al. 2017). Moreover, the strength of the effects of climate alterations differs across continents which can lead to regionally idiosyncratic patterns. For example, the last glacial period was relatively mild in east and southeast China compared to temperate regions in Europe, and several east Asian bird species maintained or even increased their population sizes during this time period (Zhao et al. 2012; Dong et al. 2017; Cheng et al. 2021). Accordingly, colonization and demographic histories during the late Pleistocene of co-occurring species or populations in different geographic regions might differ profoundly (Pedreschi et al. 2019; Cheng et al. 2021) and should be studied with a comparative approach. The comparative analysis of closely related evolutionary lineages has the potential to reveal shared and differing processes and factors in their phylogenetic and demographic histories that might be missed in separate analyses of single taxa.

We studied regionally contrasting environmental conditions as drivers of diversification in the radiation of the morphologically cryptic collared sand martins of the genus *Riparia* that unfolded over a broad geographic area containing evolutionary lineages with extensive variation in distribution ranges. Such radiations with low phenotypic diversification over large ranges pose particular challenges, because they might actually be composed of unrecognised phylogeographic units with distinct evolutionary trajectories that are not reflected in currently accepted species level taxonomy (Kindler et al. 2012; Beysard and Heckel 2014; Saxenhofer et al. 2019). One species, the collared sand martin *Riparia riparia,* has a Holarctic

distribution (Garrison and Turner 2020) with only shallow phylogeographic differentiation and low diversity in mitochondrial DNA (mtDNA) across its entire breeding range indicating that its large range is the result of a relatively recent expansion (Pavlova et al. 2008; Schweizer et al. 2018a). Three subspecies are broadly accepted (Shirihai and Svensson 2018): small-ranged *R. r. shelleyi* of Egypt and possibly the Levant (cf. Shirihai and Svensson 2018), *R. r. ijimae* with an intermediate-sized range in Central Mongolia, south-eastern Siberia, north-eastern China and Japan, and nominate *R. r. riparia* with a vast breeding range comprising Asia, Europe and North America. Eurasian breeders of the nominate form are chiefly long-distance migrants and winter in the sub-Saharan Sahel zone and East Africa, while those from North America migrate to winter in South and Central America (Shirihai and Svensson 2018; Garrison and Turner 2020). The wintering range of *R. ijimae* lies in Southeast Asia, and *R. r. shelleyi* is only a short distance migrant to Sudan and northern Ethiopia (Shirihai and Svensson 2018). Although the slight morphological variation that led to the establishment of the three subspecies is not reflected in mtDNA variation (Schweizer et al. 2018a), it might suggest the existence of different evolutionary lineages within the collared sand martin. In contrast, the pale sand martin *Riparia diluta* consists of four phylogeographic units supported by mtDNA and initial genomic data broadly reflecting the distribution area of described subspecies (Schweizer et al. 2018a; Tang et al. 2022). These comprise *R. d. indica* with a very restricted range in the arid part of the north-western Indian Subcontinent, *R. d. fohkienensis* from subtropical south China, *R. d. tibetana* from the Tibetan Plateau and central Mongolia, and *R. d. diluta* with a wide range over the steppe biome of Central Asia. *R. d*. *diluta* winters in the northern and western part of the Indian Subcontinent and rarely on the Arabian Peninsula, while *R. d. tibetana* seems to make only short-distance movements to spend the winter in the northern, central and north-eastern Indian Subcontinent and Indo China (Rasmussen and Anderton 2012; Shirihai and Svensson 2018; own data). The subspecies *R. d. fohkienensis* and *R. d. indica* might be residents or only make short-distance movements in winter (Shirihai and Svensson 2018). Population genetic data revealed no signs of gene flow between lowland *R. d*. *fohkienensis* and high-altitude *R. d*. *tibetana*, while Mongolian *R. d*. *tibetana* showed limited mixed ancestries with

Central Asian *R. d*. *diluta* (Tang et al. 2022). This indicated that some of the phylogeographic units of pale sand martin might actually represent species level taxa and contrasting migration phenology could be an important factor in maintaining evolutionary diversity among these morphologically cryptic lineages (Tang et al. 2022). No genome-wide data has been analysed so far for *R. d. indica* making the picture incomplete.

Here we investigated the diversification history of Sand and pale sand martin including samples from all subspecies using whole genome data as effective means to determine the level of evolutionary divergence between *Riparia* taxa and to infer their demographic histories. In particular, we (i) reconstructed the species tree of the *Riparia* complex and identified distinct evolutionary lineages, (ii) estimated temporal patterns of diversification and (iii) explored the demographic histories of all evolutionary lineages across the breeding range of *R. diluta,* and of samples across the Eurasian breeding range of *R. riparia*.

Based on the results of partial mtDNA in previous studies (Pavlova et al., 2008; Schweizer et al., 2018), we expected shallow genome-wide phylogeographic structure in the collared sand martin and hypothesized that the large range of the nominate subspecies is the result of massive range expansion after the LGP. However, given the availability of suitable environmental conditions for *R. riparia* throughout the LGP according to environmental niche modelling (Ponti et al. 2020) and its mobility, it might also have maintained high populations sizes throughout the Late Pleistocene conflicting with findings from mtDNA variation. In contrast, evolutionary lineages in the pale sand martin might have diverged during the late Pleistocene and evolved under different environmental conditions afterwards reflected in distinct evolutionary trajectories. We hypothesized that Late Pleistocene climate alterations should have only mildly affected the demographic history of subtropical Southern Chinese *R. d. fohkienensis* analogous to other bird species of that region (Zhao et al. 2012; Dong et al. 2017; Cheng et al. 2021). In contrast, a past population decline would be expected for high altitude *R. d. tibetana* similar to other birds of the Qinghai-Tibetan Plateau (Cheng et al. 2021). For *R. d. diluta* and *R. d. indica* of arid and semi-arid areas, the extension of open-habitats during the last

glacial period could have resulted in the expansion of populations (cf. Garcia et al. 2011a, 2011b; Kearns et al. 2014; Alaei Kakhki et al. 2018).

2. MATERIALS AND METHODS

2.1 Sampling and Data Preparation

We analysed in total nine samples of all three subspecies of *R. riparia* (4 *R. r. riparia*, 2 *R. r. ijimae,* 3 *R. r. shelleyi*) and 11 samples of all four subspecies of *R. diluta (*3 *R. d. diluta*, 3 *R. d. fohkienensis*, 2 *R. d. indica*, 3 *R. d. tibetana*). The samples within of *R. diluta* covered the entire ranges of the different subspecies, only samples from the northern part of the range of the nominate form were not available. Moreover, Mongolian samples of *R d. tibetana* were previously shown to be admixed (Tang et al. 2022) and were thus not included in our analyses. The population grouping of *R. riparia* in the mtDNA analyses of Pavlova et al (2008) was based on five predefined geographic clusters: Europe, Western Siberia, Eastern Siberia, Far East, North America. Our study included samples from three of these groups, namely Europe (*R. r. riparia*), Eastern Siberia (*R. r. ijimae* from Mongolia) and from the cluster Far East (*R. r. ijimae* from eastern China). In addition, our sample from north-western China was located between the predefined clusters Western Siberia and Eastern Siberia of Pavlova et al. (2008). Samples from the northern part of the species' range in Asia and North America were not available for our study. Unlike Pavlova et al (2008), we moreover included samples of migrants from the Levant. They phenotypically matched the subspecies *R. r. shelley*, but were either overshooting migrants from the Egyptian population or might stem from potentially unknown populations in the Levant (cf. Shirihai & Svensson 2008). Further, we included three samples of Asian *R. chinensis* and one African *R. paludicola.* In total*,* we obtained new full genome data from 24 samples, stemming from blood and toe pads (Table S1, and see Appendix S1 for further details). Thus, we were able to include all members of the genus *Riparia* in our analyses except *R. congica*, which has a very restricted distribution along the Congo and lower Ubangi rivers in western Africa (Turner 2020). We included one mitochondrial genome of *Hirundo rustica* (NCBI accession: NC_050295.1) (Carter et

al. 2020) and resequencing data from one *H. rustica* individual as outgroup in the analyses (SRA accession: SRS8625592, under BioProject: PRJNA323498 from NCBI).

Details on DNA extraction, sequencing, raw reads processing and data preparation are provided in Appendix S1. In brief, all samples were sequenced on an Illumina NovaSeq 6000 S4 flow cell. After quality control, raw reads were mapped with a *Riparia riparia* reference genome (Tang et al., 2022, GenBank assembly accession GCA_020917445.1) and the paired and mapping rate of reads for each individual is between 0.96-1. We performed SNP calling following the GATK4 best practice workflow (van der Auwera and Connor 2020). After SNP calling, only sites with minimum read depth of 5X for each individual and less than 8% (23 out of 25 individuals) missing data overall were retained. The mean read depth of each individual after filtering ranged from 5.3-14.8 (Appendix Table S1).

2.2 Mitochondrial Phylogenetic Analysis

We assembled eight mitochondrial protein coding genes (ATP6 684 bp, ATP8 168 bp, COX1 1551 bp, COX2 684 bp, COX3 783 bp, CYTB 1143 bp, ND1 978 bp, and ND2 1041) with MITOFINDER (Allio et al, 2020) and aligned them using CLUSTALW (Thompson et al. 1994) in MEGA X (Kumar et al. 2018) including data from *H. rustica* as outgroup. A time-calibrated gene tree was then reconstructed in BEAST 2.6.7 (Bouckaert et al. 2019) using published substitution rates for each mitochondrial gene (see Appendix S1 for details).

2.3 Nuclear Gene Trees and Species Tree Reconstruction

To reconstruct gene trees, we subsampled alignment windows from our nuclear genomic data (Appendix S1) at least 10 kb apart to ensure free inter-locus recombination (Ellegren et al. 2012). Subsampled windows were then used as input for IQ-TREE 2.1.4-beta (Minh et al. 2020) using a unified substitution model GTR+G+I for maximum likelihood tree reconstruction and 1000 ultrafast bootstraps. IQ-Tree utilises a fast and effective stochastic algorithm to infer phylogenetic trees by maximum likelihood (Nguyen et al. 2015). Nodes with bootstrap values below 0.8 in the gene trees were collapsed into polytomies using NEWICK_UTILS (Junier and Zdobnov 2010). Then, we used the resulting gene trees to reconstruct a species tree

using a coalescent-based method in ASTRAL-III 5.7.8 with *H. rustica* as outgroup and estimated local posterior probability (LPP) for each branch (Zhang et al. 2018; Rabiee et al. 2019). Moreover, to quantify the amount of gene tree heterogeneity, relative quartet support for the main topology, first alternative and second alternative topology of the species tree were calculated for each branch respectively.

2.4 Detection of Ancient Introgression

Discordance between gene trees and the species tree can be caused by evolutionary processes such as introgression upon hybridization or incomplete lineage sorting (ILS) (Morales-Briones et al. 2021). To test for potential ancestral introgression between the different linages of *R. diluta* and *R. riparia*, we estimated Patterson's D (ABBA-BABA statistic) and the f4-ratio (admixture fraction f) (Patterson et al. 2012) for all possible trios of lineages using the Dtrios function in DSUITE (Malinsky et al. 2021). We used VCF files generated for phylogenomic analysis (Appendix S1) as input excluding invariable sites and keeping SNPs with a minimal individual read depth of at least five and less than 10 % missing data among all 25 sampled individuals. The species tree inferred with ASTRAL (Fig. 1) was used as input tree. We further calculated the f-branch statistic (Malinsky et al. 2018) using DSUITE. The f-branch statistic was designed to disentangle correlated f4-ratio results and can assign gene flow among different (possibly internal) branches in the population or species tree (Malinsky et al. 2018; Suvorov et al. 2022). Results were plotted using scripts provided in DSUITE (https://github.com/millanek/Dsuite).

2.5 Time Tree Reconstruction

We used the RelTime-ML approach (Tamura et al. 2012) which estimates branch-specific relative rates in MEGA X to generate a time-calibrated tree based on a genomic input tree with branch lengths measured in number of substitutions per site. To this end, we first concatenated all windows to reconstruct a maximum likelihood phylogenetic tree using IQ-TREE with GTR+G+I as substitution model and 1000 bootstraps. The resulting tree was then used as input for the RelTime-ML approach with the split between *R. chinensis* and all the remaining taxa fixed at 5.01 Ma based on the results of the time-calibrated mtDNA gene tree (see below). Due to

computation limits, we only included windows which were at least 10 kb apart and thus unlinked, and with alignment length larger than 6 kb for the time calibration. Local clocks were applied as default setting with GTR+G+I as substitution model and *H. rustica* as outgroup.

2.6 Demographic Inference

To investigate the demographic history of different subspecies of *R. riparia* and *R. diluta,* we used PSMC to estimate change in effective population size (N_e) through time from an unphased diploid genome based on the pairwise sequentially Markovian coalescent model (Li and Durbin 2011). Alternative methods using multiple genomes allow estimation of more recent demographic history and splitting time, however, they require multiple high coverage genomes (>20X) or phased genomes (Schiffels and Wang 2020). Phasing of our data would be error prone due to lack of population level sampling and not high enough coverage (Terhorst et al. 2017; Schiffels and Wang 2020). Only individuals with mean raw read depth larger than 12X were used for further PSMC analyses (Nadachowska-Brzyska et al. 2016; de Greef et al. 2022). For pale sand martin, we thus performed demographic inferences for three individuals each of *R. d. diluta*, *R. d. fohkienensis* and *R. d. tibetana*, as well as one individual of *R. d. indica*. For collared sand martin, two individuals each were analysed for *R. r. riparia* and *R. r. ijimae* and three individuals of *R. r. shelleyi*. See Appendix S1 for further details.

3. **Results**

3.1 Species Tree Reconstruction

Genome-wide data provided a robust phylogenetic hypothesis for the genus *Riparia* based on a species tree approach statistically consistent with multi-species coalescent. We obtained in total 45,810 unlinked genomic windows with a mean alignment length of 5,094 bp (ranging from 471 bp to 9474 bp) including 3,177,915 variable sites covering 35 chromosomes in all 25 individuals (Fig S1, Appendix S1). The species tree revealed Asian *R. chinensis* as the sister species to the remainder of the *Riparia* radiation, with African *R. paludicola* being the sister taxon of the clade

consisting *of R. riparia* and *R. diluta* (Fig. 1). Within *R. riparia*, the samples of the different subspecies formed monophyletic clades each with *R. r. ijimae* as sister lineage to the clade including *R. r. shelleyi* and *R. r. riparia*. However, internal branches in the clade were short except the one leading to *R. r. shellyi*. In contrast, there was deep genetic structure in *R. diluta*, with individuals of each subspecies forming well separated monophyletic groups. One larger clade consisted of the southern taxa *R. d. indicia* and *R. d. fohkienensis,* the other of the northern taxa *R. d. diluta* and *R. d. tibetana*. Local posterior probabilities for branches were one, except for very short branches leading to tip clades, one each within *R. d. fohkienensis* (posterior probability (pp) = 0.66), *R. d. diluta* (pp = 0.97), *R. r. shelleyi* (pp = 0.76) and *R. r. riparia* (pp = 0.67) (marked with * in Fig 1). Gene tree concordance was revealed for most branches in the phylogeny (Fig. 1), with overall very high relative quartet support for the main topology. More gene tree heterogeneity was revealed at short internal branches within *R. chinensis*, *R. d. diluta*, *R. d. fohkienensis* and *R. d. tibetana*, as well as within *R. riparia* except for the relatively long branch leading to *R. r. shelleyi* (relative quartet support for the main topology of 0.87).

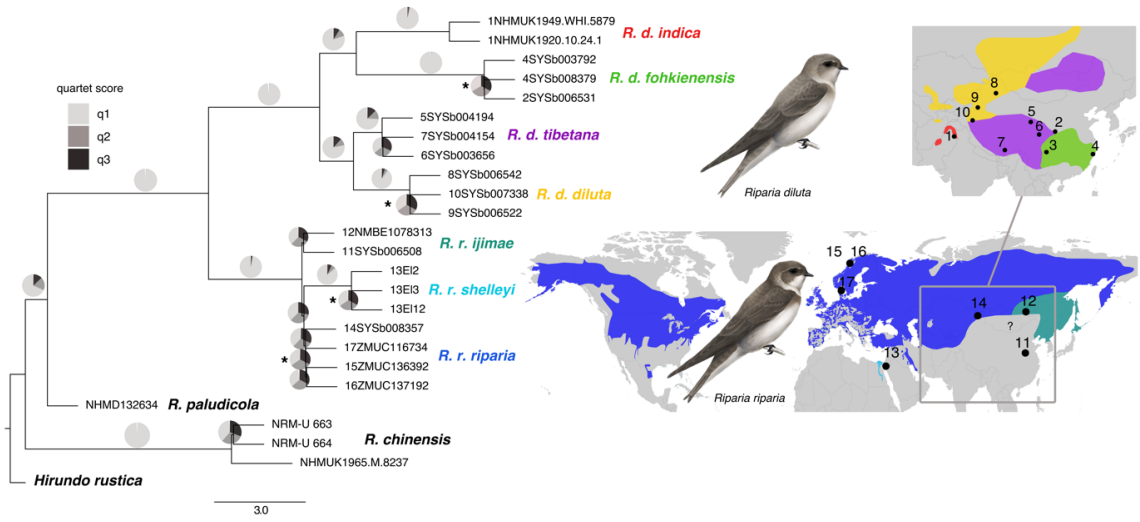


Figure 1. The coalescent-based species tree using whole genome data of *Riparia* martins. Pie charts show the relative quartet support for the main topology (q1), the first alternative (q2), and the second alternative (q3) for each branch. Branches with LPP (local posterior probability) < 1 were marked with *. Breeding distributions of

subspecies within *R. diluta* (top) and *R. riparia* (bottom) were modified from Bird Life International & Handbook of the Birds of the World (2016) and Tang et al. (2022). Numbers on maps represent sample origins (see Table S1) and are also given in individual identifiers in the tree. Subspecies are indicated by different colours. The question mark indicates potential breeding area. Bird illustrations by M. Schweizer.

3.2 Ancestral Introgression

The f-branch statistic suggested an introgression event between the ancestral branch of *R. d. tibetana* and *R. d. diluta* and the three branches within *R. riparia* with an average fraction of excess shared genetic material (f_b) of 0.049 (Fig. 2, Fig. S2, Appendix S1). Furthermore, limited footprints of introgression were also detected between *R. d. fohkienensis* and *R. d. diluta* (f_b = 0.020) and between *R. d. indica* and *R. d. tibetana* (f_b = 0.014). As *R. d. tibetana* and *R. d. diluta* are sister taxa, DSUITE cannot scan for introgression between these but clear signs for it were found in an earlier study (Tang et al. 2022).

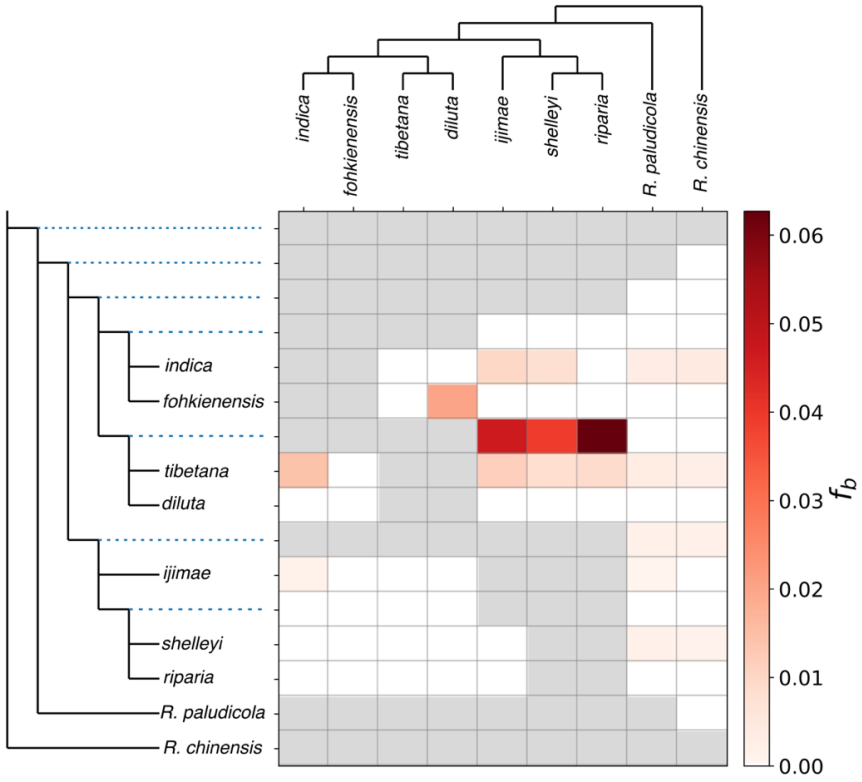


Figure 2. Introgression between *Riparia* taxa inferred by f-branch statistic. f_b values in the matrix represent excess allele sharing between the branch on the expanded

tree on the y axis (relative to its sister branch) and the lineages on the x-axis. Only subspecies names were indicated for *R. diluta* and *R. riparia*.

3.3 Time-calibrated Phylogeny based on Mitochondrial Data

The topology of the mtDNA tree corresponded well with the species tree, although several nodes in the *Riparia* radiation were not robustly supported (Fig. S3, Appendix S1) and there was much less resolution within *R. riparia* compared to the species tree. The split between *R. diluta* and *R. riparia* was estimated at 1.6 Ma, and diversification within *R. diluta* started around 1.2 Ma. The position of *R. d. fohkienensis* and *R. d. indica* was not resolved, while *R. d. diluta* and *R. d. tibetana* formed a robustly supported clade and split at 0.7 Ma. In stark contrast, mitochondrial diversification of *R. riparia* was estimated to have started much more recently, only around 0.1 Ma. The three *R. d. shelleyi* individuals clustered together with robust support, but there was no geographic structure among the samples of *R. r. riparia* and *R. r. ijimae*.

3.4 Time Tree Reconstruction based on Genome-wide Data

The reconstruction of a maximum likelihood tree based on 91,620 concatenated windows yielded a highly supported backbone for the time tree analysis: all nodes had bootstrap values of 100 (Fig. S4, Appendix S1). The topology differed from the species tree only in *R. riparia*: while the samples of *R. r. shelleyi* clustered together, the samples from the other subspecies did not form monophyletic clades. The time-calibrated phylogeny was then computed based on a subsampled dataset of 14,485 windows. The divergence between *R. paludicola* from the clade consisting of *R. diluta* and *R. riparia* was estimated at around 4.3 Ma and the split between *R. diluta* and *R. riparia* at approximately 2.4 Ma (Fig. 3). Divergence within *R. diluta* started around 1.1 Ma, with subspecies splitting at around 0.8 Ma. Similar to *R. diluta*, individuals of *R. riparia* shared a most recent common ancestor at around 1.1 Ma., with a *R. r. shelleyi* at around 0.7 Ma.

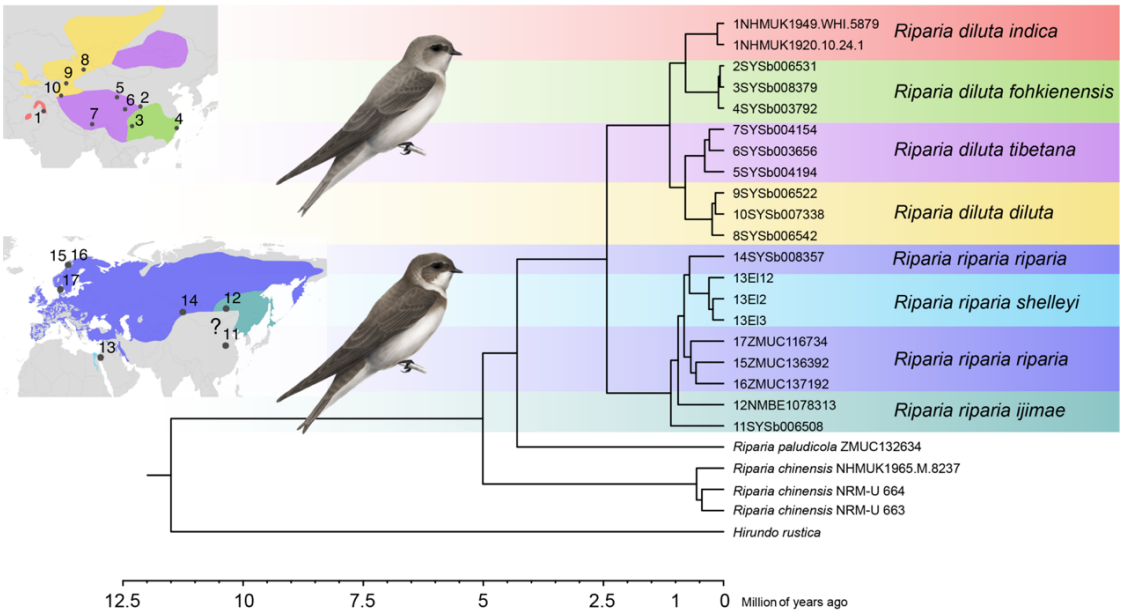


Figure 3. Time-calibrated tree based on concatenated genomic sequences of *Riparia* martins. The split time of 5.01 Ma (million years ago) between *R. chinensis* and the remaining *Riparia* taxa was used as calibration point for time estimation. Labels in the tree indicate the locality ID (the first number) in the map and the sample ID in Table S1. Subspecies are highlighted by colour. Bird illustrations by M. Schweizer.

3.5 Demographic History

PSMC analyses revealed widely different demographic histories within *Riparia*, both at inter- and intraspecific levels (Fig. 4). *Riparia riparia*, except for *R. r. shelleyi*, showed up to an order of magnitude larger effective population sizes than *R. diluta.* Effective population sizes of both *R. r. riparia* and *R. r. ijimae* increased until the beginning of the LGP with not much signal in individual genomes per subspecies afterwards. Starting with the LGP, signals indicated either stable N_e for Central Asian *R. r. riparia* and both *R. r ijimae* samples, or a strong population expansion in European *R. r. riparia*. Effective population size in *R. r. shelleyi* was much lower than in the other *R. riparia* subspecies at a similar level as in *R. diluta*. It experienced a population expansion at the beginning of the late Pleistocene peaking at around 0.5 Ma followed by a long-lasting population decline (Fig. S5, Appendix S1). Overall, bootstrap analyses revealed consistent demographic patterns, but more variation towards the present (Fig. S5, Appendix S1)

In pale sand martin, individuals within subspecies showed largely consistent patterns. *Riparia d. diluta* underwent a population expansion at the beginning of the LGP followed by a trend of decline towards the LGM. In contrast, *R. d. tibetana* populations contracted approximately from the beginning of the LGP onwards after a steep increase in N_e before. There was a recovery with a trend for expansion again towards the LGM. In *R. d. fohkienensis* on the other hand, N_e was generally lower but stable during the LGP with an expansion towards the LGM.

To test for the influence of generation time, we additionally ran PSMC with a generation time of one year instead of two years. Although this shifted the estimates towards the present (Fig. S6, Appendix S1), the overall pattern was not affected and our interpretation of the results would not change.

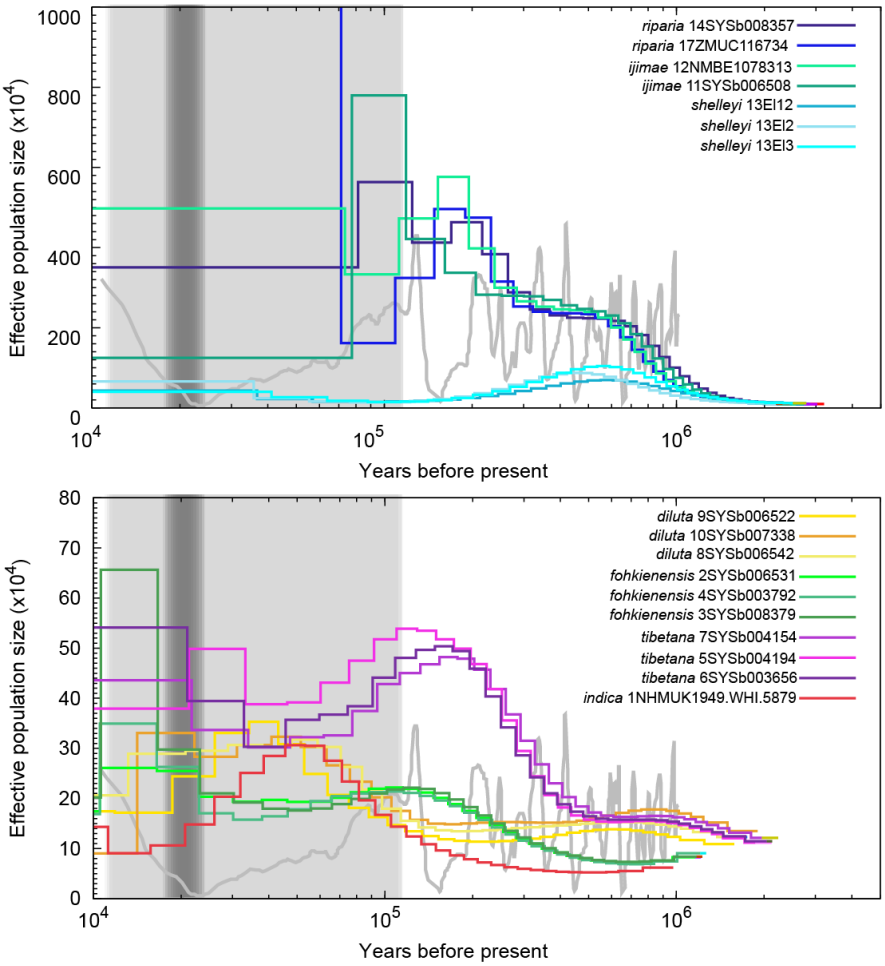


Figure 4. Demographic inference based on PSMC of different subspecies within *R. riparia* (top) and *R. diluta* (bottom). Labels indicate subspecies, locality ID and

sample ID for each individual. Colours correspond to the different subspecies in the distribution map of Figure 1. The grey areas indicate the Last Glacial Period (LGP, 110-12 kya) with Last Glacial Maximum (LGM, 2.3-1.8 kya) shown in dark grey. The grey line illustrates relative changes in Northern Hemisphere air temperature modified from Li et al. (2021).

4. DISCUSSION

Our comparative analyses of whole-genome sequence data provide insights into the diversification history of morphologically almost cryptic collared sand martin species in the genus *Riparia* of the Holarctic and adjacent regions. We revealed contrasting patterns of diversification and demographic history both at inter- and intraspecific levels in pale sand martin *R. diluta* and collared sand martin *R. riparia.* We discuss how regional differences in past climate related environmental change in combination with variation in seasonal migration behaviour could have influenced evolutionary trajectories and range patterns in pale and collared sand martin.

4.1 Phylogeographic History of the collared sand martin

The collared sand martin *R. riparia* represents one of relatively few bird species which have colonized a large Holarctic distribution range without human intervention spanning from the Mediterranean to northern Scandinavia, from Asia Minor over Central Asia to north-eastern Siberia, and from New Mexico to Alaska (Garrison and Turner 2020). The analyses presented here suggest a long-lasting diversification process during the last million years. Individuals were clustered in three different evolutionary lineages in the phylogenomic framework of our species tree consistent with subspecies designation, although separated by mostly very short branches. One lineage comprised the birds from the Levant (*R. r. shelleyi*), the second one the birds from Northern Europe and north-western China (*R. r. riparia*) and the third one the birds from Mongolia and north-eastern China (*R. r. ijimae*). Comparatively minor genomic divergence was found between birds from Europe and Central Asia despite extensive environmental variation across the range. *R. riparia*

from north-western China were genomically much closer to those from Scandinavia than to the geographically adjacent ones from Mongolia and north-eastern China.

Riparia r. shelleyi from Egypt and possibly the Levant (Shirihai and Svensson 2008) is characterized by a demographic history that is independent from the other *R. riparia* from about one million years ago (Fig. 4). The remaining *R. riparia* genomes suggested continuous population expansion approximately from the beginning of the LGP onwards until they provided no clear signal. The latter might be a consequence of long coalescent times in large populations reflecting the large Holarctic range of this species (Garrison and Turner 2020). Only the *R. r. ijimae* genome from north-eastern China suggested a population contraction towards the present. The strong increase in N_e of the European sample of *R. r. riparia* around the beginning of the LGP might be an artefact because recent bottlenecks can potentially cause spurious spikes in N_e in PSMC analyses (Lu et al. 2022). A similar pattern in a single Pied Flycatcher indicating much higher N_e than current estimates was explained by such a history (Nadachowska-Brzyska et al. 2016). We cannot exclude a bottleneck in the history of European *R. riparia*, however, suitable environmental conditions might have been available for *R. riparia* throughout the LGP according to environmental niche modelling (Ponti et al. 2020). Thus, a strong recent bottleneck seems unlikely and the collared sand martin, except *R. r. shelleyi*, could have maintained relatively high N_e during the LGP even in Europe.

Inferred coalescent times of *R. riparia* mtDNA haplotypes at around 0.1 Ma are congruent with the estimates by Pavlova et al. (2008) provided for much shorter mtDNA sequences. This, in combination with only shallow phylogeographic differentiation in mtDNA over the entire Holarctic range of *R. riparia* (Pavlova et al. 2008; Schweizer et al. 2018a), is intriguing given the much longer evolutionary history of divergence inferred from the nuclear genome of our samples. Pavlova et al. (2008) associated mtDNA differentiation between western and eastern Siberian samples with range fragmentation of *R. riparia* during the last glacial maximum. However, our genome-wide data indicates that the east-west split among Eurasian samples is much older - likely several hundred thousand years old. Our result once more shows that inferences of evolutionary histories solely based on mtDNA can potentially yield misleading conclusions not only in terms of phylogenetic

relationships, but also in divergence time estimates and demographic history (e.g. Irwin et al. 2009; Taylor et al. 2021b). Similar mito-nuclear discordances have been found in other Holarctic bird groups and were explained by random genetic drift or selection, or a combination thereof (Irwin et al. 2009; Taylor et al. 2021b). We hypothesize that the collared sand martin might have repeatedly shifted its range as a consequence of climate and associated environmental alterations. The isolated population of *R. r. shelleyi* south of the European range of the species would then represent a remnant of a temporal range shift to the south.

Repeated range shifts of *Riparia* taxa may have led to several phases of introgression. This might have resulted in introgression upon secondary contact with pale sand martin during the initial stages of divergence as indicated by the DSUITE analyses. This may have happened also later between the lineages of collared sand martin today considered different subspecies and resulted in the introgression and fixation of mtDNA haplotypes. Interestingly, no signs for admixture between the different collared sand martin lineages were detected in the Dsuite analyses. However, if hybridization was temporarily limited and mtDNA introgression fostered by selection, genome-wide signs for gene flow might be masked by the species' high effective populations size. Based on the mtDNA analyses of (Pavlova et al. 2008), European and Siberian populations showed signs of recent demographic expansion whereas patterns in birds from the Far East and North America suggested more stable populations. Introgression of mtDNA might thus have happened e.g. from Beringia to the west. However, testing these ideas and putative selective sweeps in mtDNA e.g. as a consequence of thermal adaptations (e.g. Lamb et al. 2018; Melo-Ferreira et al. 2005) will require more extensive population level sampling in the future, including more northern and north-eastern populations in Asia and especially from North America.

4.2 Phylogeographic History of the pale sand martin

In contrast to the collared sand martin, the genomic clades in the geographically much more restricted pale sand martin are largely consistent with the results from limited mtDNA data (Pavlova et al. 2008; Schweizer et al. 2018a) and population genetic data (Tang et al. 2022). The distribution of evolutionary lineages

within pale sand martin is not directly reflected in simple patterns of divergence related to geography (e.g. Dong et al. 2013). It has a circular though not continuous distribution around the hyper-arid region of the Tarim basin and the Gobi Desert and additionally the isolated *R. d. indica* lineage in the north-western part of the Indian Subcontinent. The landscape of the Gobi Desert was already formed at the Plio-/Pleistocene border (Lu et al. 2019), predating the start of the radiation of collared sand martins. Its formation was thus unlikely to have acted as a driver of divergence in pale sand martin, nor was isolation by distance around this hostile area a driver (cf. Tang et al. 2022) as would be expected for a ring-like diversification history (Alcaide et al. 2014). Rather, the species tree revealed a split between the subspecies occurring basically south of the Qinghai-Tibetan plateau (*R. d. fohkienensis* and *R. d. indica*) from those occurring on this plateau and to the north of it (*R. d. tibetana* and *R. r. diluta*). While the breeding ranges of the former two are isolated, the latter are potentially in contact in central, and probably north-western Mongolia with limited introgression from *R. d. diluta* into *R. d. tibetana* (Tang et al. 2022).

Qualitatively different demographic histories were revealed for *R. d. tibetana* and *R. diluta.* While *R. d. tibetana* shows signs for population decline towards the present, an increase in effective population size during the same period was indicated in *R. d. diluta*. This might be a consequence of regionally contrasting environmental effects of climate change during the LGP. The current breeding area of *R. d. tibetana* on the Qinghai-Tibetan plateau was widely glaciated during the cold periods of the late Pleistocene (Cui et al. 1998; Zhou et al. 2006). We thus hypothesize that it might have passed at least part of the late Pleistocene in a refugium to the east or northeast of the Qinghai-Tibetan plateau. In *R. d. diluta* in contrast, range expansion during the LGP might be related to the extension of open arid habitats during dry glacial periods as documented for arid species in different geographic regions (Garcia et al. 2011a, 2011b; Kearns et al. 2014; Alaei Kakhki et al. 2018). The turning point to a decline in N_e in *R. d. diluta* could be explained by increased aridity which might have been too extreme even for arid adapted taxa (cf. Schweizer et al. 2018b; Schweizer and Shirihai 2013). Interestingly, the period of lowest N_e in *R. d. tibetana* coincided with the peak in *R. d. diluta*. Although we

analysed samples with no signs for admixture, it cannot be excluded that the increase in N_e in *R. d. tibetana* afterwards is an artefact in the estimation caused by introgression in secondary contact from *R. d. diluta* after its range expansion (see Tang et al. (2022).

The southern clade comprising *R. d. fohkienensis* and *R. d. indica* represents a unique phylogeographic link in birds between subtropical southern China and the dry north-western part of the Indian Subcontinent. Remarkably, the breeding range of the closely related Grey-throated Martin *R. chinensis* extends continuously between both areas and thus overlaps with both *R. diluta* subspecies (del Hoyo et al. 2020). *Riparia d. indica* breeds in relatively arid regions similar to *R. d. diluta*, and accordingly shows an analogous demographic history. It may have initially profited from an increase in arid open habitats in the north-western part of the Indian Subcontinent (Kar and Kumar 2020) but contracted its range when aridity increased.

Lowland *R. d. fohkienensis* in Southern China is the only taxon within *R. diluta* that does not breed in a comparatively arid environment, and we thus consider its adaptation to a subtropical climate as secondary. Unsurprising, it shows an idiosyncratic demographic history with basically constant N_e followed by an increase towards the present differing strongly from *R. d. tibetana* on the adjacent highlands. Similarly pronounced differences in the demographic history between closely related species on the Qinghai-Tibetan plateau and adjacent Chinese lowlands were shown in Paridae (Cheng et al. 2021). Constant population sizes or even a population expansion during the LGP as in *R. d. fohkienensis* are consistent with minor effects of this period on organisms of subtropical eastern Asia (Zhao et al. 2012; Dong et al. 2017; Cheng et al. 2021).

4.3 Taxonomy

Within the pale sand martin *R. diluta*, our genomic results demonstrate the reliability of earlier taxonomic work integrating very subtle morphological differentiation and biogeographic information, yet our genomic data resolved important features of this radiation. The extended levels of genomic divergence within the pale sand martin might support the existence of multiple species level taxa. The findings by Tang et al. (2022) indicated effective reproductive isolation

between *R. d. tibetana* and *R. d. fohkienensis*, likely through allochrony. Our extended taxon sampling now demonstrates that *R. d. tibetana* and *R. d. fohkienensis* actually belong to different deeper clades within pale sand martin. *Riparia d. fohkienensis* and *R. d. indica* showed similar levels of genomic divergence as *R. d. diluta* and *R. d. tibetana,* and the latter two might interbreed in secondary contact (Tang et al. 2022).

Following an integrative approach towards speciation (Schweizer et al. 2023), we refrain from suggesting to split *R. diluta* into a northern and a southern species before any potential differences in vocalizations have been documented and morphological differentiation in this cryptic radiation has been revisited in the light of our findings. Further studies should moreover investigate the extent of interbreeding between *R. d. diluta* and *R. d. tibetana* and clarify the actual distribution the enigmatic *R. d. indica*. Currently, there is lack of knowledge on its status and distribution (Grimmett et al. 1998; Rasmussen and Anderton 2012). We were only able to include two historical samples of the latter in our analyses including the holotype (NHMUK1920.10.24.1), and further investigations should also cover extant samples.

Within *R. riparia*, we revealed the subspecies *R. r. shelleyi* as genetically distinct. It is also morphologically more differentiated compared to the other subspecies of *R. riparia*, approaching the *R. diluta* species complex in size and plumage (Shirihai and Svensson 2018). The individuals analysed here from the spring season in Southern Israel do not stem from the core range of the taxon and might represent overshooting migrants. Geographically more comprehensive sampling particularly in the regions of potential contact with other *R. riparia* could provide important information about the taxonomic status of *R. r. shelleyi*.

Our results also clarified the taxonomic status of the widely disjunct distributed African *R. paludicola* and Asian *R. chinensis* which have been considered as conspecifics for a long time. However, they were recently split based on differences in size and vocalization (Rasmussen and Anderton 2012). Our genomic data supports their distinctiveness and surprisingly, they turned out to be not even sister species with *R. chinensis* being the sister species of the remaining *Riparia* radiation.

5. CONCLUSIONS

The vast range of the nominate subspecies of the collared sand martin is in stark contrast to the comparatively smaller ranges of the other taxa of *Riparia* analysed in this study. Our data indicate that its Eurasian part is not the result of massive range expansion after the LGP as previously thought, but rather of a long demographic expansion. *Riparia r. riparia* is the taxon with the largest populations sizes during LGP and the longest migration distances, and thus it might have high tolerance for heterogenous climate conditions and a high colonization potential (cf. Thorup et al. 2021). Indeed, 7% of juvenile collared sand martins in Britain were found >199 km away from their natal colonies in successive years (Mead 1979) indicating comparatively low natal philopatry (Tittler et al. 2009). Computer simulations have suggested that migratory birds in North America and the western Palearctic – but less so in Asia – have shifted their breeding range towards LGM progressively in the direction of the equator (Somveille et al. 2020) - probably repeatedly through the late Pleistocene. Especially *R. r. riparia* might thus have adapted to track suitable habitat and ephemeral resources enabling stable or even increased effective populations sizes during the late Pleistocene glacial cycles.

While pronounced dispersal propensity might have allowed the colonization of a large range and hindered lineage diversification in the collared sand martin, distinct seasonal migration behaviour might have fostered lineage divergence and maintained evolutionary diversity in the pale sand martin complex (cf. Tang et al. 2022). Hence, seasonal migration behaviour may have contrasting impacts in the two *Riparia* species: While it might be important in maintaining evolutionary diversity in the pale sand martin, it might be counteracting lineage diversification in nominate collared sand martin and prevent evolutionary subdivision of a large distribution range.

SUPPLEMENTARY MATERIAL

Data available from the Dryad Digital Repository:
https://doi.org/10.5061/dryad.q2bvq83pt

FUNDING

We acknowledge the financial support of Oversea Study Programme of Guangzhou Elite Project (no. JY201726).

ACKNOWLEDEMENTS

We thank Qin Huang, Xinyuan Pan, Dan Liang, Yun Li, 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 we are grateful to Susanne Tellenbach for the help with laboratory work. We thank Peter A. Hosner and J.-B. Kristensen (Zoological Museum, University of Copenhagen), R. Prŷs-Jones and M. Adams (Natural History Museum, Tring), and Ulf Johansson (The Swedish Museum of Natural History) for providing samples. We are grateful to Reto Burri for commenting on the manuscript. We acknowledge the Next Generation Sequencing (NGS) Platform at University of Bern and we especially thank Pamela Nicholson for her support. We thank Alex Suh and Anna Torgasheva for providing a mitochondrial genome of *Riparia riparia* and additional L. Malinovskaya and P. Borodin for the help with sampling the relevant individual.

Author contributions:

Manuel Schweizer and Gerald Heckel conceived the study. Qindong Tang, Manuel Schweizer and Gombobaatar Sundev collected samples with assistance from Yang Liu. Hadoram Shirihai collected samples. Qindong Tang conducted laboratory work and analysed the data together with Gerald Heckel and Manuel Schweizer and assistance from Nilofar Alayee. Manuel Schweizer and Gerald Heckel wrote the manuscript with support from all authors.

REFERENCES

Alaei Kakhki N., Aliabadian M., Förschler M.I., Ghasempouri S.M., Kiabi B.H., Verde Arregoitia L.D., Schweizer M. 2018. Phylogeography of the *Oenanthe hispanica–pleschanka–cypriaca* complex (Aves, Muscicapidae: Saxicolinae): Diversification history of open-habitat specialists based on climate niche models, genetic data, and morphometric data. J. Zool. Syst. Evol. Res.:1–20.

Alcaide M., Scordato E.S.C., Price T.D., Irwin D.E. 2014. Genomic divergence in a ring species complex. Nature. 511:83-U433.

Allio R., Schomaker-Bastos A., Romiguier J., Prosdocimi F., Nabholz B., Delsuc F. 2020. MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. Mol. Ecol. Resour. 20:892–905.

van der Auwera G.A., Connor B.D. 2020. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (1st Edition). O'Reilly Media. .

Baca M., Popović D., Lemanik A., Bañuls-Cardona S., Conard N.J., Cuenca-Bescós G., Desclaux E., Fewlass H., Garcia J.T., Hadravova T., Heckel G., Horáček I., Knul M.V., Lebreton L., López-García J.M., Luzi E., Marković Z., Mauch Lenardić J., Murelaga X., Noiret P., Petculescu A., Popov V., Rhodes S.E., Ridush B., Royer A., Stewart J.R., Stojak J., Talamo S., Wang X., Wójcik J.M., Nadachowski A. 2023. Ancient DNA reveals interstadials as a driver of common vole population dynamics during the last glacial period. J. Biogeogr. 50:183–196.

Beysard M., Heckel G. 2014. Structure and dynamics of hybrid zones at different stages of speciation in the common vole (*Microtus arvalis*). Mol. Ecol. 23:673–687.

Bouckaert R., Vaughan T.G., Barido-Sottani J., Duchêne S., Fourment M., Gavryushkina A., Heled J., Jones G., Kühnert D., De Maio N., Matschiner M., Mendes F.K., Müller N.F., Ogilvie H.A., du Plessis L., Popinga A., Rambaut A., Rasmussen D., Siveroni I., Suchard M.A., Wu C.-H., Xie D., Zhang C., Stadler T., Drummond A.J. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLOS Comput. Biol. 15:e1006650.

Brüniche-Olsen A., Kellner K.F., Belant J.L., DeWoody J.A. 2021. Life-history traits and habitat availability shape genomic diversity in birds: implications for conservation. Proc. R. Soc. B Biol. Sci. 288:20211441.

Carter J.K., Innes P., Goebl A.M., Johnson B., Gebert M., Attia Z., Gabani Z., Li R., Melie T., Dart C., Mares A., Greidanus C., Paterson J., Wall B., Cortese G., Thirouin K., Glime G., Rutten J., Poyd C., Post E., Wall B., Elhadi A.A., Feldmann K., Danz A., Blanchard T., Amato S., Reinert S., Pogoda C.S., Scordato E.S.C., Hund A.K., Safran R.J., Kane N.C. 2020. Complete mitochondrial genomes provide current refined phylogenomic hypotheses for relationships among ten *Hirundo* species. Mitochondrial DNA Part B. 5:2881–2885.

Chattopadhyay B., Garg K.M., Ray R., Rheindt F.E. 2019. Fluctuating fortunes: genomes and habitat reconstructions reveal global climate-mediated changes in bats' genetic diversity. Proc. R. Soc. B Biol. Sci. 286:20190304.

Cheng Y., Miller M.J., Zhang D., Xiong Y., Hao Y., Jia C., Cai T., Li S.-H., Johansson U.S., Liu Y., Chang Y., Song G., Qu Y., Lei F. 2021. Parallel genomic responses to historical climate change and high elevation in East Asian songbirds. Proc. Natl. Acad. Sci. 118:e2023918118.

Cui Z., Wu Y., Liu G., Ge D., Pang Q., Xu Q. 1998. On Kunlun-Yellow River tectonic movement. Sci. China Ser. Earth Sci. 41:592–600.

Dong F., Hung C.-M., Li X.-L., Gao J.-Y., Zhang Q., Wu F., Lei F.-M., Li S.-H., Yang X.-J. 2017. Ice age unfrozen: severe effect of the last interglacial, not glacial, climate change on East Asian avifauna. Bmc Evol. Biol. 17:244.

Dong L., Heckel G., Liang W., Zhang Y. 2013. Phylogeography of Silver Pheasant (*Lophura nycthemera* L.) across China: aggregate effects of refugia, introgression and riverine barriers. Mol. Ecol. 22:3376–3390.

Ellegren H., Smeds L., Burri R., Olason P.I., Backström N., Kawakami T., Künstner A., Mäkinen H., Nadachowska-Brzyska K., Qvarnström A., Uebbing S., Wolf J.B.W. 2012. The genomic landscape of species divergence in Ficedula flycatchers. Nature. 491:756–760.

Garcia J.T., Alda F., Terraube J., Mougeot F., Sternalski A., Bretagnolle V., Arroyo B. 2011a. Demographic history, genetic structure and gene flow in a steppe-associated raptor species. Bmc Evol. Biol. 11:11.

Garcia J.T., Manosa S., Morales M.B., Ponjoan A., de la Morena E.L.G., Bota G., Bretagnolle V., Davila J.A. 2011b. Genetic consequences of interglacial isolation in a steppe bird. Mol. Phylogenet. Evol. 61:671–676.

Garrison B.A., Turner A. 2020. Bank Swallow (*Riparia riparia*). In: Billerman S.M., Keeney B.K., Rodewald P.G., Schulenberg T.S., editors. Birds of the World. Cornell Lab of Ornithology.

Gaston K.J. 1998. Species-range size distributions: products of speciation, extinction and transformation. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 353:219–230.

Gaston K.J. 2003. The structure and dynamics of geographic ranges. Oxford: Oxford University Press.

de Greef E., Brashear W., Delmore K.E., Fraser K.C. 2022. Population structure, patterns of natal dispersal and demographic history in a declining aerial insectivore, the purple martin *Progne subis*. J. Avian Biol. 2022.

Grimmett R., Inskipp C., Inskipp T. 1998. Birds of the Indian subcontinent. London: A & C Black.

del Hoyo J., Collar N., Kirwan G.M. 2020. Gray-throated Martin (*Riparia chinensis*). In: Billerman S.M., Keeney B.K., Rodewald P.G., Schulenberg T.S., editors. Birds of the World. Cornell Lab of Ornithology.

Hung C.M., Drovetski S.V., Zink R.M. 2017. The roles of ecology, behaviour and effective population size in the evolution of a community. Mol. Ecol. 26:3775–3784.

Irwin D.E., Rubtsov A.S., Panov E.N. 2009. Mitochondrial introgression and replacement between yellowhammers (*Emberiza citrinella*) and pine buntings *(Emberiza leucocephalos*) (Aves: Passeriformes). Biol. J. Linn. Soc. 98:422–438.

Junier T., Zdobnov E.M. 2010. The Newick utilities: high-throughput phylogenetic tree processing in the UNIX shell. Bioinformatics. 26:1669–1670.

Kar A., Kumar A. 2020. Evolution of arid landscape in India and likely impact of future climate change. Episodes. 43:511–523.

Kearns A.M., Joseph L., Toon A., Cook L.G. 2014. Australia's arid-adapted butcherbirds experienced range expansions during Pleistocene glacial maxima. Nat. Commun. 5.

Kindler E., Arlettaz R., Heckel G. 2012. Deep phylogeographic divergence and cytonuclear discordance in the grasshopper *Oedaleus decorus*. Mol. Phylogenet. Evol. 65:695–704.

Kumar S., Stecher G., Li M., Knyaz C., Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol. Biol. Evol. 35:1547–1549.

Lamb A.M., Gan H.M., Greening C., Joseph L., Lee Y.P., Morán-Ordóñez A., Sunnucks P., Pavlova A. 2018. Climate-driven mitochondrial selection: A test in Australian songbirds. Mol. Ecol. 27:898–918.

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

Li J., Bian C., Yi Y., Yu H., You X., Shi Q. 2021. Temporal dynamics of teleost populations during the Pleistocene: a report from publicly available genome data. BMC Genomics. 22:490.

Lu C., Yao C., Hung C. 2022. Domestication obscures genomic estimates of population history. Mol. Ecol. 31:752–766.

Lu H., Wang X., Wang X., Chang X., Zhang H., Xu Z., Zhang W., Wei H., Zhang X., Yi S., Zhang W., Feng H., Wang Y., Wang Y., Han Z. 2019. Formation and evolution of Gobi Desert in central and eastern Asia. Earth-Sci. Rev. 194:251–263.

Malinsky M., Matschiner M., Svardal H. 2021. Dsuite - Fast *D* -statistics and related admixture evidence from VCF files. Mol. Ecol. Resour. 21:584–595.

Malinsky M., Svardal H., Tyers A.M., Miska E.A., Genner M.J., Turner G.F., Durbin R. 2018. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. Nat. Ecol. Evol. 2:1940–1955.

Mead C.J. 1979. Colony fidelity and interchange in the Collared sand martin. Bird Study. 26:99–106.

Melo-Ferreira J., Boursot P., Suchentrunk F., Ferrand N., Alves P.C. 2005. Invasion from the cold past: extensive introgression of mountain hare (*Lepus timidus*) mitochondrial DNA into three other hare species in northern Iberia. Mol. Ecol. 14:2459–2464.

Minh B.Q., Schmidt H.A., Chernomor O., Schrempf D., Woodhams M.D., von Haeseler A., Lanfear R. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Mol. Biol. Evol. 37:1530–1534.

Morales-Briones D.F., Kadereit G., Tefarikis D.T., Moore M.J., Smith S.A., Brockington S.F., Timoneda A., Yim W.C., Cushman J.C., Yang Y. 2021. Disentangling Sources of Gene Tree Discordance in Phylogenomic Data Sets: Testing Ancient Hybridizations in Amaranthaceae s.l. Syst. Biol. 70:219–235.

Nadachowska-Brzyska K., Burri R., Smeds L., Ellegren H. 2016. PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white *Ficedula* flycatchers. Mol. Ecol. 25:1058–1072.

Nadachowska-Brzyska K., Li C., Smeds L., Zhang G., Ellegren H. 2015. Temporal Dynamics of Avian Populations during Pleistocene Revealed by Whole-Genome Sequences. Curr. Biol. 25:1375–1380.

Nguyen L.-T., Schmidt H.A., Von Haeseler A., Minh B.Q. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol. Biol. Evol. 32:268–274.

Patterson N., Moorjani P., Luo Y., Mallick S., Rohland N., Zhan Y., Genschoreck T., Webster T., Reich D. 2012. Ancient Admixture in Human History. Genetics. 192:1065–1093.

Pauls S.U., Nowak C., Bálint M., Pfenninger M. 2013. The impact of global climate change on genetic diversity within populations and species. Mol. Ecol. 22:925–946.

Pavlova A., Zink R.M., Drovetski S.V., Rohwer S. 2008. Pleistocene evolution of closely related collared sand martins *Riparia riparia* and *R. diluta*. Mol. Phylogenet. Evol. 48:61–73.

Pedreschi D., García-Rodríguez O., Yannic G., Cantarello E., Diaz A., Golicher D., Korstjens A.H., Heckel G., Searle J.B., Gillingham P., Hardouin E.A., Stewart J.R. 2019. Challenging the European southern refugium hypothesis: Species-specific structures versus general patterns of genetic diversity and differentiation among small mammals. Glob. Ecol. Biogeogr. 28:262–274.

Ponti R., Arcones A., Ferrer X., Vieites D.R. 2020. Lack of evidence of a Pleistocene migratory switch in current bird long-distance migrants between Eurasia and Africa. J. Biogeogr. 47:1564–1573.

Pujolar J.M., Blom M.P.K., Reeve A.H., Kennedy J.D., Marki P.Z., Korneliussen T.S., Freeman B.G., Sam K., Linck E., Haryoko T., Iova B., Koane B., Maiah G., Paul L., Irestedt M., Jønsson K.A. 2022. The formation of avian montane diversity across barriers and along elevational gradients. Nat. Commun. 13:268.

Rabiee M., Sayyari E., Mirarab S. 2019. Multi-allele species reconstruction using ASTRAL. Mol. Phylogenet. Evol. 130:286–296.

Rasmussen P.C., Anderton J.C. 2012. Birds of South Asia. The Ripley Guide. Smithsonian National Museum of Natural History, Washington D.C. and Michigan State University and Lynx Edicions, Barcelona.

Saxenhofer M., Schmidt S., Ulrich R.G., Heckel G. 2019. Secondary contact between diverged host lineages entails ecological speciation in a European hantavirus. PLOS Biol. 17:e3000142.

Schiffels S., Wang K. 2020. MSMC and MSMC2: The Multiple Sequentially Markovian Coalescent. In: Dutheil J.Y., editor. Statistical Population Genomics. New York, NY: Springer US. p. 147–166.

Schweizer M., Liu Y., Olsson U., Shirihai H., Huang Q., Leader P.J., Copete J.L., Kirwan G.M., Chen G.L., Svensson L. 2018a. Contrasting patterns of diversification in two sister species of martins (Aves: Hirundinidae): The Collared sand martin *Riparia riparia* and the Pale Martin *R. diluta*. Mol. Phylogenet. Evol. 125:116–126.

Schweizer M., Marques D.A., Olsson U., Crochet P. 2023. The Howard & Moore Complete Checklist of the Birds of the World: framewo rk for species delimitation. Avian Syst. 1 (IX): N35–N41.

Schweizer M., Shirihai H. 2013. Phylogeny of the *Oenanthe lugens* complex (Aves, Muscicapidae: Saxicolinae): paraphyly of a morphologically cohesive group within a recent radiation of open-habitat chats. Mol. Phylogenet. Evol. 69:450–461.

Schweizer M., Shirihai H., Schmaljohann H., Kirwan G.M. 2018b. Phylogeography of the House Bunting complex: discordance between species limits and genetic markers. J. Ornithol. 159:47–61.

Sexton J.P., McIntyre P.J., Angert A.L., Rice K.J. 2009. Evolution and Ecology of Species Range Limits. Annual Review of Ecology Evolution and Systematics. Palo Alto: Annual Reviews. p. 415–436.

Shirihai H., Svensson L. 2018. Handbook of Western Palearctic Birds. London: A. & C. Black.
Somveille M., Wikelski M., Beyer R.M., Rodrigues A.S.L., Manica A., Jetz W. 2020. Simulation-based reconstruction of global bird migration over the past 50,000 years. Nat. Commun. 11:801.
Suvorov A., Kim B.Y., Wang J., Armstrong E.E., Peede D., D'Agostino E.R.R., Price D.K., Waddell P.J., Lang M., Courtier-Orgogozo V., David J.R., Petrov D., Matute D.R., Schrider D.R., Comeault A.A. 2022. Widespread introgression across a phylogeny of 155 *Drosophila* genomes. Curr. Biol. 32:111-123.e5.
Tamura K., Battistuzzi F.U., Billing-Ross P., Murillo O., Filipski A., Kumar S. 2012. Estimating divergence times in large molecular phylogenies. Proc. Natl. Acad. Sci. 109:19333–19338.
Tang Q., Burri R., Liu Y., Suh A., Sundev G., Heckel G., Schweizer M. 2022. Seasonal migration patterns and the maintenance of evolutionary diversity in a cryptic bird radiation. Mol. Ecol. 31:632–645.
Taylor R.S., Bramwell A.C., Clemente-Carvalho R., Cairns N.A., Bonier F., Dares K., Lougheed S.C. 2021a. Cytonuclear discordance in the crowned-sparrows, *Zonotrichia atricapilla* and *Zonotrichia leucophrys*. Mol. Phylogenet. Evol. 162:107216.
Taylor R.S., Manseau M., Klütsch C.F.C., Polfus J.L., Steedman A., Hervieux D., Kelly A., Larter N.C., Gamberg M., Schwantje H., Wilson P.J. 2021b. Population dynamics of caribou shaped by glacial cycles before the last glacial maximum. Mol. Ecol. 30:6121–6143.
Terhorst J., Kamm J.A., Song Y.S. 2017. Robust and scalable inference of population history from hundreds of unphased whole genomes. Nat. Genet. 49:303–309.
Thompson J.D., Higgins D.G., Gibson T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.
Thorup K., Pedersen L., da Fonseca R.R., Naimi B., Nogués-Bravo D., Krapp M., Manica A., Willemoes M., Sjöberg S., Feng S., Chen G., Rey-Iglesia A., Campos P.F., Beyer R., Araújo M.B., Hansen A.J., Zhang G., Tøttrup A.P., Rahbek C. 2021. Response of an Afro-Palearctic bird migrant to glaciation cycles. Proc. Natl. Acad. Sci. 118:e2023836118.
Tittler R., Villard M.-A., Fahrig L. 2009. How far do songbirds disperse? Ecography. 32:1051–1061.
Turner A. 2020. Congo Martin (*Riparia congica*). In: Billerman S.M., Keeney B.K., Rodewald P.G., Schulenberg T.S., editors. Birds of the World. Cornell Lab of Ornithology.
Willis S.C., Farias I.P., Ortí G. 2014. TESTING MITOCHONDRIAL CAPTURE AND DEEP COALESCENCE IN AMAZONIAN CICHLID FISHES (CICHLIDAE: *CICHLA*): TESTING ANCIENT INTROGRESSION USING UNLINKED LOCI. Evolution. 68:256–268.
Zhang C., Rabiee M., Sayyari E., Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics. 19:153.
Zhao N., Dai C., Wang W., Zhang R., Qu Y., Song G., Chen K., Yang X., Zou F., Lei F. 2012. Pleistocene climate changes shaped the divergence and demography of

Asian populations of the great tit *Parus major*: evidence from phylogeographic analysis and ecological niche models. J. Avian Biol. 43:297–310.

Zhou S., Wang X., Wang J., Xu L. 2006. A preliminary study on timing of the oldest Pleistocene glaciation in Qinghai–Tibetan Plateau. Quat. Int. 154–155:44–51.

APPENDIX S1

Details on Samples

See Table S1 for a list of analysed samples. Two toe pad samples of *R. d. indica* were provided by the Natural History Museum, London, UK (NHMUK), including the holotype (NHMUK1920.10.24.1), and of *R. chinensis* by the Swedish Museum of Natural History (NRM). The toe pad samples were collected between 1913-1914. Three blood samples of *R. riparia* collected in Northern Europe and one blood sample of *R. paludicola* were provided by the Zoological Museum, Natural History Museum of Denmark (ZMUC), and samples of *R. r. shelleyi* were captured and sampled by HS under permits and approvals from the relevant authorities in Israel. The remaining blood samples were collected by QT, MS, YL and GS during 2016-2019 with birds captured and sampled under permits and approvals from the relevant authorities in China and Mongolia. We also included one mitochondrial genome of *Hirundo rustica* (NCBI accession: NC_050295.1) (Carter et al. 2020) and resequencing data from one *H. rustica* individual as outgroup in the analyses (SRA accession: SRS8625592, under BioProject: PRJNA323498 from NCBI).

Table S1. Sample list. Locality ID corresponding to the numbers in the distribution map of Figure 1. Accessions for NCBI sequence read archive (SRA) are under BioProject PRJNA755835 for each individual. NHMUK = Natural History Museum, Tring; NMBE = Natural History Museum of Bern, Switzerland; NRM = The Swedish Museum of Natural History; SYS = Sun Yat-sen University, Guangzhou; ZMUC = Zoological Museum, Natural History Museum of Denmark, University of Copenhagen).

Species	Sample ID	Locality ID	Locality	Tissue	Mean Depth	SRA
R. d. indica	NHMUK1920.10.24.1	1	Jhelum, Punjab, Pakistan	Toe pad	5.3	xxx
R. d. indica	NHMUK1949.WHI.5879	1	Jhelum, Punjab, Pakistan	Toe pad	11.5	xxx
R. d. fohkienensis	SYSb006531	2	Meixian, Shaanxi Province, PR China	Blood	14.5	xxx
R. d. fohkienensis	SYSb008379	3	Yibin, Sichuan, PR China	Blood	14.8	xxx
R. d. fohkienensis	SYSb003792	4	Wenzhou, Zhejiang Province, PR China	Blood	13.9	xxx
R. d. tibetana	SYSb004194	5	Rubber Mountain, Qinghai Province, PR China	Blood	11.4	xxx
R. d. tibetana	SYSb003656	6	Zoige, Sichuan Province, PR China	Blood	11.6	xxx
R. d. tibetana	SYSb004154	7	Lhasa, Tibet Autonomous Region, PR China	Blood	11.8	xxx
R. d. diluta	SYSb006542	8	Wujiaqu, Xinjiang Uygur	Blood	10.1	xxx

R. d. diluta	SYSb006522	9	Alar, Xinjiang Uygur Autonomous Region, PR China	Blood	8.8	xxx
R. d. diluta	SYSb007338	10	Khotan, Xinjiang Uygur Autonomous Region, PR China	Blood	10.6	xxx
R. r. ijimae	SYSb006508	11	Zhengzhou, Henan Province, PR China	Blood	11.6	xxx
R. r. ijimae	NMBE1078313	12	Dzharakhain urto (Jaraakhain urtuu), Mongolia	Blood	11.0	xxx
R. r. shelleyi	El2	13	Eilat, Israel	Blood	10.2	xxx
R. r. shelleyi	El3	13	Eilat, Israel	Blood	10.6	xxx
R. r. shelleyi	El12	13	Eilat, Israel	Blood	11.6	xxx
R. r. riparia	SYSb008357	14	Baisha Mountain, Xinjiang Uygur Autonomous Region, PR China	Blood	12.1	xxx
R. r. riparia	ZMUC136392	15	Kaas, Nordjylland, Denmark	Blood	9.0	xxx
R. r. riparia	ZMUC137192	16	Løkken, Jylland, Denmark	Blood	7.6	xxx

R. r. riparia	ZMUC116734	17	Seivaag, Bodoe, Straumoeya, Norway	Blood	11.0	xxx
R. paludicola	ZMUC132634		Mweya, Queen Elizabeth N.P., Uganda	Blood	13.9	xxx
R. chinensis	NHMUK1965.M.8237		Sukkur, Sind, Pakistan	Toe pad	6.2	xxx
R. chinensis	NRM-U 663		Chieng Rai, Thailand	Toe pad	7.4	xxx
R. chinensis	NRM-U 664		Chieng Rai, Thailand	Toe pad	5.8	xxx

DNA Extraction and Sequencing

Blood samples collected during our fieldwork were preserved in 99% ethanol and later stored at -20 °C. The DNA from toe pad samples of *R. d. indica* and *R. chinensis* were extracted as described in Schweizer et al. (2018). DNA was extracted from the remaining samples with a modified salt extraction protocol (Aljanabi and Martinez 1997). We used two different library construction methods based on the level of DNA fragmentation. For the two toe pad samples of *R. d. indica* and three blood samples of *R. chinensis* with highly fragmented DNA samples (around 200 bp), the quantity, purity and length of the total genomic DNA was assessed using a Thermo Fisher Scientific Qubit 4.0 fluorometer with the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Q32854), a DeNovix DS-11 FX spectrophotometer and an Agilent FEMTO Pulse System with a Genomic DNA 165 kb Kit (Agilent, FP-1002-0275), respectively. Sequencing libraries were made using a *Swift* Accel-NGS *1S* Plus DNA Library Kit (Swift Biosciences,10096) in combination with 1S Unique Dual Indexing Kit (Swift Biosciences, 190384) according to the protocol provided by Swift Biosciences,. Pooled DNA libraries were first sequenced paired-end using an illumina iSeq 100 i1 Reagent v2 (300 cycles, illumina, 20031371) on an illumina iSeq 100 System. This served as a quality control run. Thereafter, the pool was re-balanced and then sequenced paired-end on a NovaSeq 6000 S4 Reagent Kit v1.5 (200 cycles; illumina, 20028313) using an Xp workflow (illumina NovaSeq Xp 4-Lane Kit v1.5, 20043131) on an illumina NovaSeq 6000 instrument. Genomic libraries of the remaining samples were prepared using TruSeq DNA PCR-free sample preparation (Illumina), and then run on an Illumina NovaSeq 6000 (S4 flow cell, 2 x 150 bp).

Raw Reads Processing

The quality of raw reads was checked first using fastqc version 0.10.1 (Andrew 2019). Raw reads from the SP flow cell were then trimmed twice using trimmomatic version 0.39 (Bolger et al. 2014). In the first round, we removed 15 bp from the end of R1 reads (parameters: CROP:82 HEADCROP:4) and 15 bp from the beginning of the R2 reads (parameters: CROP:94 HEADCROP:15) separately to remove the Adaptase tail following the manufacturer's tail trimming protocol. In the second round, trimmed reads were paired (parameters: LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20). Raw reads from the S4 flow cell were trimmed and paired with parameters CROP:145 HEADCROP:8 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20. Those from the resequencing data of *H. rustica* were trimmed and paired with parameters HEADCROP:15 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20. The quality of trimmed and paired reads was further checked using fastqc. Only paired reads were kept for later analysis. We also checked the deamination for historical toepad samples. Only paired reads from each library were mapped separately with reference genome BWA 0.7.17 (bwa men) (Li 2013) and deamination was accessed using mapDamage version 2.0 (Jónsson et al. 2013) to check the frequency of different nucleic acid residues at positions upstream and downstream of the start/end of reads and the read length distribution. No post-

mortem damage were found for toe pad samples and the majority read length is around 80 bp (Figure S7-S10)

Data Preparation for Nuclear Phylogenomic Analysis

The *Riparia riparia* reference genome (Tang et al., 2022, GenBank assembly accession GCA_020917445.1) was first anchored to a chromosome level assembly of a *Hirundo rustica* genome (NCBI assembly bHirRus1.pri.v2) using minimap2 (Li 2018) and the repeat regions were masked using RepeatMasker 4.1.2-p1 (Smit et al. 2013). In order to get a fasta alignment for phylogenomic reconstruction from the resequencing data, we followed the protocol described in Kakhki et al. (2022). Reads of each individual were separately mapped on the chromosome-anchored *Riparia riparia* reference genome using BWA 0.7.17 (bwa men) (Li 2013). For each bam file, only mapped alignments with mapping quality (MAPQ) larger than 30 were retained using SAMTOOLS version 1.9 (Danecek et al. 2021). DUPLICATES were removed using MARKDUPLICATES in PICARD tools version 2.25 (http://broadinstitute.github.io/picard/). Variant calling on all samples was then performed using GATK version 4.1.4.1 (Auwera and O'Connor 2020), following the GATK4 best practice workflow for SNP and indel calling using GenomicsDBImport to merge GVCFs (Genomic Variant Call Format) from multiple samples (https://gatk.broadinstitute.org/hc/en-us/articles/360036883491-GenomicsDBImport). The first round of variant calling was performed using GATK HaplotypeCaller, hard filtering thresholds were applied as suggested by the GATK's Best Practices: QD (QualByDepth) < 2.0, QUAL (Phred quality scores) < 30.0, SOR (StrandOddsRatio) > 3.0, FS (FisherStrand) > 60.0, MQ (RMSMappingQuality) < 40.0, MQRankSum (MappingQualityRankSumTest) < -12.5, ReadPosRankSum (ReadPosRankSumTest) < −8.0. Then we removed SNPs that did not pass the thresholds, and only retained biallelic SNPs with at least one homozygous reference and one homozygous alternative genotype or with at least three observations of reference and alternative alleles. The remaining SNPs were used as input for Base Quality Score Recalibration (BQSR) in GATK to reduce the systematic errors. Afterwards, we did a second round of variant calling with GATK HaplotypeCaller with the recalibrated base quality score and performed joint genotyping with GATK GenotypeGVCFs including invariant sites. We also used RepeatMasker for a second round to check whether there were repeat regions remaining and removed them for the VCF files using bcftools (Danecek et al. 2021). Indels and sites within 10 base pairs of an indel were also removed from the VCF files using bcftools. We then used a perl script from Kakhki et al. (2022) to transform the VCF files into unphased fasta alignments. Only sites with minimum read depth of 5X for each individual and less than 8% missing data were retained. For each alignment, transforming and filtering performed over non-overlapping windows of 10 kb in size along the genome. For heterozygous genotypes, one allele each was randomly written into the fasta alignment. For the gene tree and species tree reconstruction, we subsampled one window from every two windows to obtain unlinked windows. For time tree reconstruction, the input phylogeny was estimated based on all the windows. To compute a time-calibrated phylogeny, we used the unlinked windows from the gene tree construction and only kept alignment with length larger than 6 kb due to the computation limits.

Data Preparation for Mitochondrial Phylogenomic Analysis

We assembled mitochondrial sequences for the 24 *Riparia* samples *de novo* using MitoFinder version 1.4 (Allio et al. 2020). We used one *Riparia riparia* mitochondrial genome (assembled from an unpublished 10X Genomics Chromium data, details see in the acknowledgements) for annotation in MitoFinder.

We assembled eight mitochondrial protein coding genes (ATP6 684 bp, ATP8 168 bp, COX1 1551 bp, COX2 684 bp, COX3 783 bp, CYTB 1143 bp, ND1 978 bp, and ND2 1041) with MitoFinder (Allio et al, 2020,) and aligned them using ClustalW (Thompson et al. 1994) in Mega X (Kumar et al. 2018) including data from *H. rustica* as outgroup. A time-calibrated gene tree was then reconstructed in Beast 2.6.7 (Bouckaert et al. 2019) using published substitution rates for each mitochondrial gene (Lerner et al. 2011). Published substitution rates for each mitochondrial gene were implemented as means of the clock rates in real space of a lognormal distribution with standard deviations of 0.005 (Lerner et al. 2011). We implemented a Yule speciation process for the tree prior and an uncorrelated lognormal relaxed clock model. Ten independent MCMC chains were run for 20 million generations, each with sampling every 1,000 generations with substitution models inferred using bModelTest (Bouckaert and Drummond 2017) implemented in Beast. Tracer 1.7.2 (Rambaut et al. 2018) was used to assess parameter convergence among the runs and to confirm appropriate burn-in and adequate effective sample sizes (ESS) of the posterior distribution. The ten independent runs were combined using LogCombiner 2.6.7 (Bouckaert et al. 2019) with 10% burn-in each and the maximum clade credibility gene tree with 95% highest posterior density (HPD) distributions of each node was estimated with TreeAnnotator 2.6.7 (Bouckaert et al. 2019).

Demographic Inference

To investigate the demographic history of different subspecies of *R. riparia* and *R. diluta,* we used PSMC to estimate change in effective population size (N_e) through time from an unphased diploid genome based on the pairwise sequentially Markovian coalescent model (Li and Durbin 2011).

The whole-genome diploid consensus sequence for each individual (three *R. d. diluta*, three R. *d. fohkienensis*, three *R. d. tibetana,* one *R. d. indica, two R. r. riparia*, two *R. r. ijimae*, three *R. r. shelleyi)* was generated following the pipeline provided with PSMC (https://github.com/lh3/psmc). First, after excluding duplicates, we used bam file (cf. Appendix S1) only including the autosome regions for each individual and applied samtools mpileup with parameters -C50 -Q 30 -q 30 to remove reads with low base quality and low mapping quality. Then we called variants using bcftools and converted VCF files into fastq format using vcfutils.pl vcf2fq with a minimum read depth of 10 (-d 10) as suggested by Nadachowska-Brzyska et al. (2016) and a maximum read depth of twice of the average depth (-D 30). For PSMC inference, we then used a parameter setting usually applied in birds (e.g. Cheng et al. 2021; Nadachowska-Brzyska et al. 2015): total number of iterations (-N) set to 25; upper limit of the TMRCA (time to the most recent common ancestor) (-t) set to 5; initial mutation/recombination ratio (-r) of 5; atomic time interval pattern (-p) of "4+30*2+4+6+10". We scaled results with a mutation rate of 2.3×10^{-9} per generation per site as estimated for Collared Flycatcher (Smeds et al. 2016) and

a generation time of two years (COSEWIC, 2013). We additionally performed 100 bootstrap replicates of PSMC analyses for all individuals.

REFERENCES

Aljanabi S.M., Martinez I. 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. Nucleic acids research. 25:4692–4693.

Allio R., Schomaker-Bastos A., Romiguier J., Prosdocimi F., Nabholz B., Delsuc F. 2020. MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. Mol Ecol Resour. 20:892–905.

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.

Auwera G. van der, O'Connor B.D. 2020. Genomics in the cloud: using Docker, GATK, and WDL in Terra. Sebastopol, CA: O'Reilly Media.

Bolger A.M., Lohse M., Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30:2114–2120.

Bouckaert R., Vaughan T.G., Barido-Sottani J., Duchêne S., Fourment M., Gavryushkina A., Heled J., Jones G., Kühnert D., De Maio N., Matschiner M., Mendes F.K., Müller N.F., Ogilvie H.A., du Plessis L., Popinga A., Rambaut A., Rasmussen D., Siveroni I., Suchard M.A., Wu C.-H., Xie D., Zhang C., Stadler T., Drummond A.J. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PLoS Comput Biol. 15:e1006650.

Bouckaert R.R., Drummond A.J. 2017. bModelTest: Bayesian phylogenetic site model averaging and model comparison. BMC Evol Biol. 17:42.

Carter J.K., Innes P., Goebl A.M., Johnson B., Gebert M., Attia Z., Gabani Z., Li R., Melie T., Dart C., Mares A., Greidanus C., Paterson J., Wall B., Cortese G., Thirouin K., Glime G., Rutten J., Poyd C., Post E., Wall B., Elhadi A.A., Feldmann K., Danz A., Blanchard T., Amato S., Reinert S., Pogoda C.S., Scordato E.S.C., Hund A.K., Safran R.J., Kane N.C. 2020. Complete mitochondrial genomes provide current refined phylogenomic hypotheses for relationships among ten *Hirundo* species. Mitochondrial DNA Part B. 5:2881–2885.

Cheng Y., Miller M.J., Zhang D., Xiong Y., Hao Y., Jia C., Cai T., Li S.-H., Johansson U.S., Liu Y., Chang Y., Song G., Qu Y., Lei F. 2021. Parallel genomic responses to historical climate change and high elevation in East Asian songbirds. Proc. Natl. Acad. Sci. U.S.A. 118:e2023918118.

Danecek P., Bonfield J.K., Liddle J., Marshall J., Ohan V., Pollard M.O., Whitwham A., Keane T., McCarthy S.A., Davies R.M., Li H. 2021. Twelve years of SAMtools and BCFtools. GigaScience. 10:giab008.

Jónsson H., Ginolhac A., Schubert M., Johnson P.L.F., Orlando L. 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics. 29:1682–1684.

Kakhki N.A., Schweizer M., Lutgen D., Bowie R.C.K., Shirihai H., Suh A., Schielzeth H., Burri R. 2022. Abundant Phenotypic Parallelism, Incomplete Lineage Sorting, and Introgression in Open-Habitat Chats. .

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.

Lerner H.R.L., Meyer M., James H.F., Hofreiter M., Fleischer R.C. 2011. Multilocus Resolution of Phylogeny and Timescale in the Extant Adaptive Radiation of Hawaiian Honeycreepers. Current Biology. 21:1838–1844.

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. 34:3094–3100.

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

Nadachowska-Brzyska K., Burri R., Smeds L., Ellegren H. 2016. PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white *Ficedula* flycatchers. Mol Ecol. 25:1058–1072.

Nadachowska-Brzyska K., Li C., Smeds L., Zhang G., Ellegren H. 2015. Temporal Dynamics of Avian Populations during Pleistocene Revealed by Whole-Genome Sequences. Current Biology. 25:1375–1380.

Rambaut A., Drummond A.J., Xie D., Baele G., Suchard M.A. 2018. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Systematic Biology. 67:901–904.

Schweizer M., Liu Y., Olsson U., Shirihai 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.

Smeds L., Qvarnström A., Ellegren H. 2016. Direct estimate of the rate of germline mutation in a bird. Genome Res. 26:1211–1218.

Smit A.F.A., Hubley R., Green P. 2013. RepeatMasker (Open-4.0) [Computer software]. http://www.repeatmasker.org. .

Tang Q., Burri R., Liu Y., Suh A., Sundev G., Heckel G., Schweizer M. 2022. Seasonal migration patterns and the maintenance of evolutionary diversity in a cryptic bird radiation. Molecular Ecology. 31:632–645.

Thompson J.D., Higgins D.G., Gibson T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucl Acids Res. 22:4673–4680.

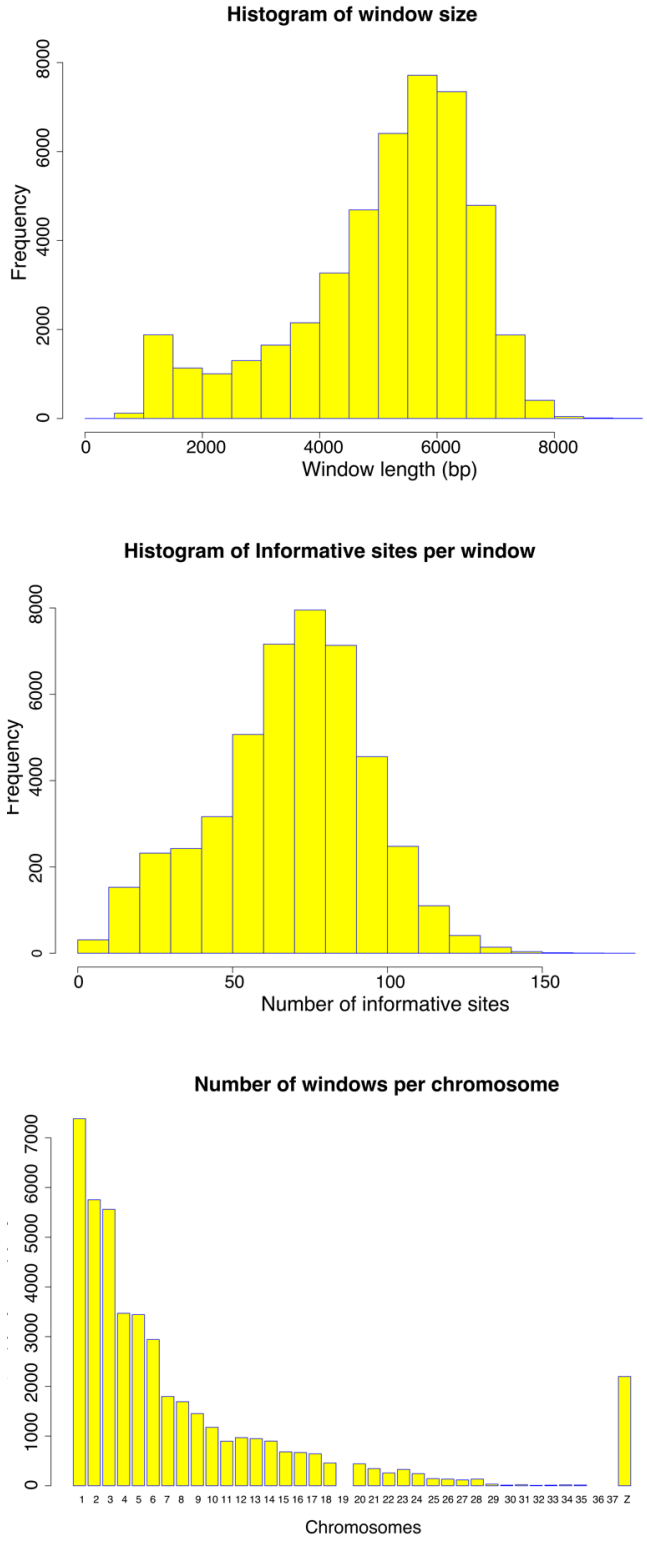


Figure S1. Summary statistics of window size, informative site sand window distribution on the different chromosomes in the final alignment among 25 individuals used for species tree reconstruction using ARSTRAL-III.

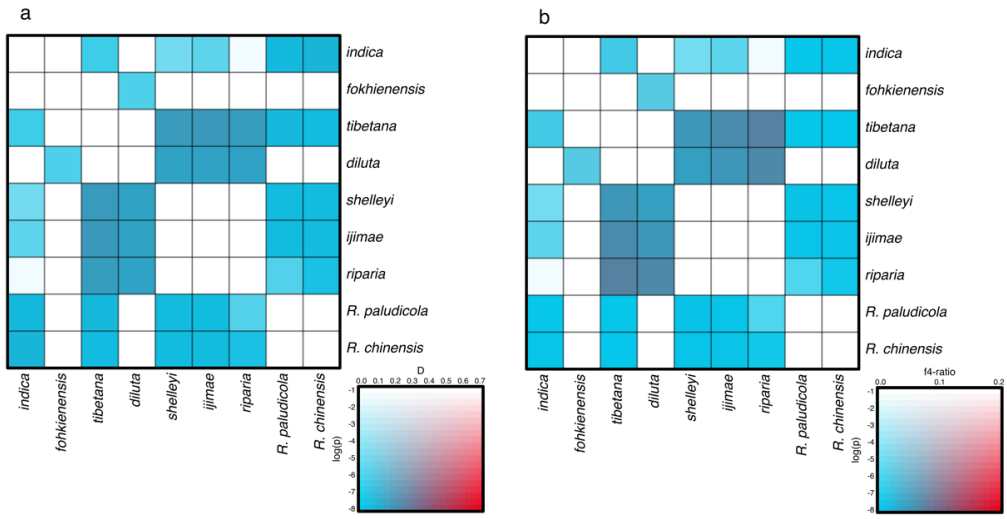


Figure S2. *D*-statistic (a) and f_4-ratio (b) statistic for all possible trios of lineages within *Riparia*

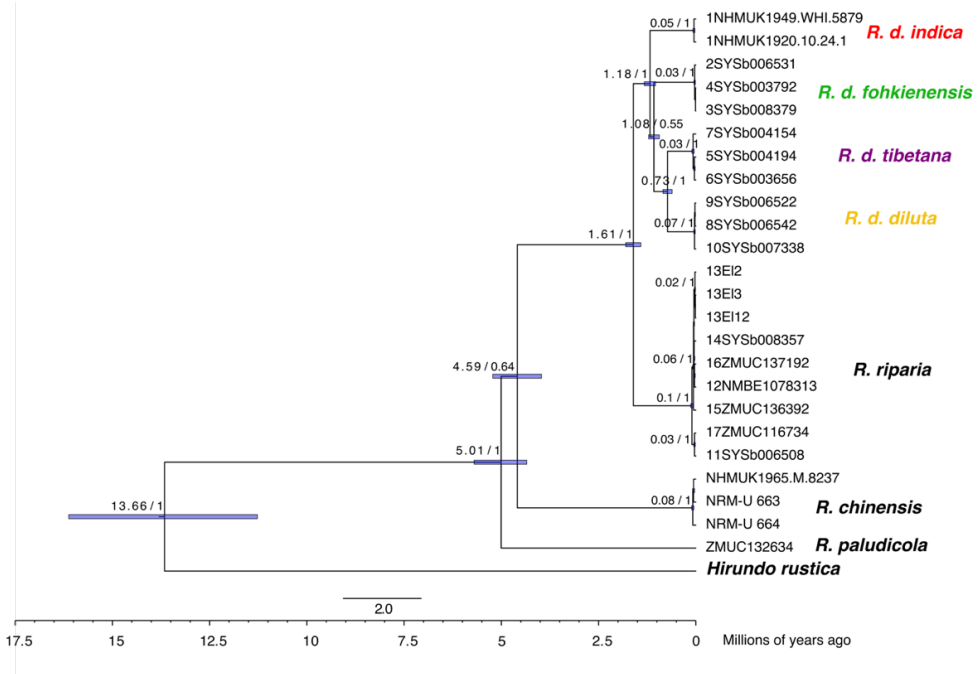


Figure S3. Time-calibrated maximum-clade-credibility tree based on mitochondrial data using BEAST of four *Riparia* species. Mean node ages (value before the slash) are shown for the major nodes and the blue bars represent the 95% highest posterior density distributions for the divergence time estimate of the major nodes. Posterior probability for the major nodes are additionally indicated after the slash. Individual labels in the indicates the locality ID (the first number) in the map (Figure 1) and the sample ID in Table 1. Colors are corresponding to the breeding distribution of different *R. diluta* subspecies in Figure 1 (top right).

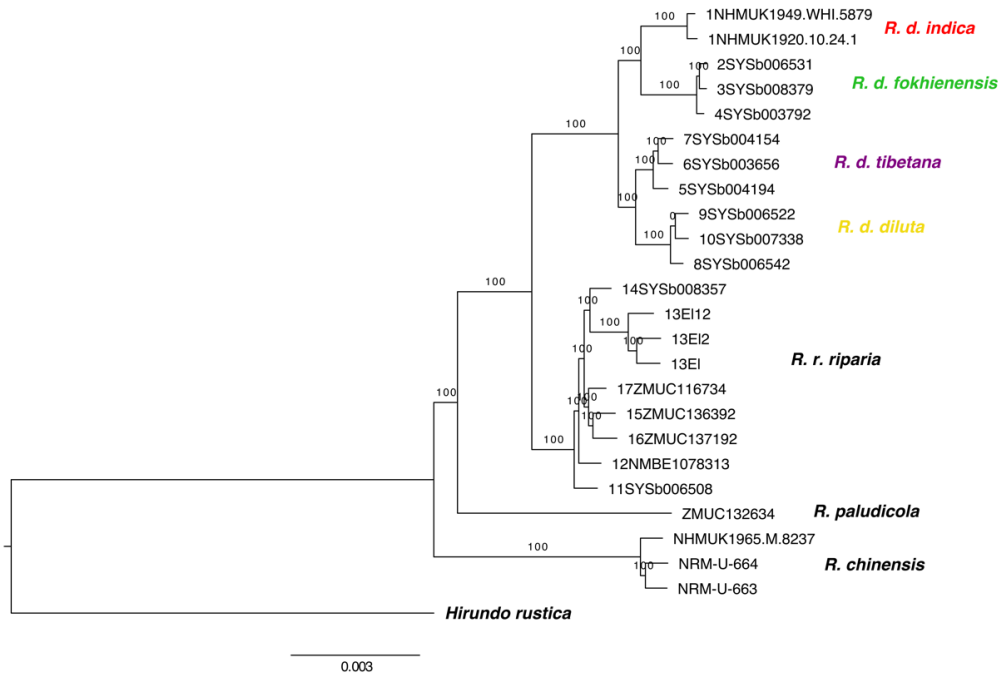
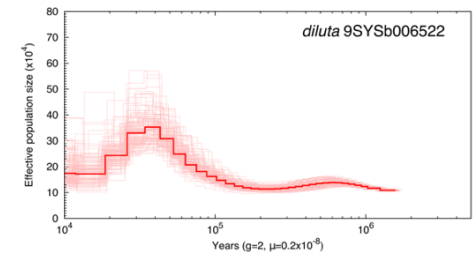
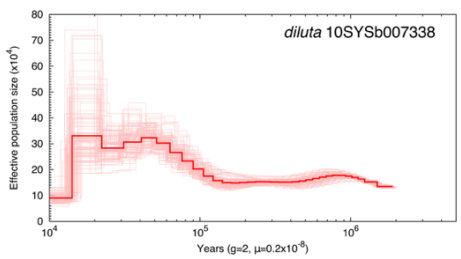
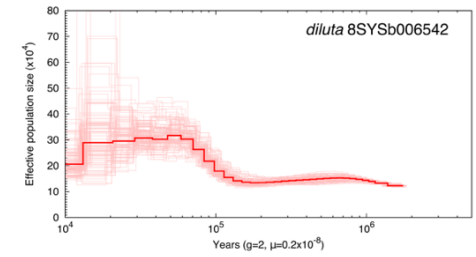
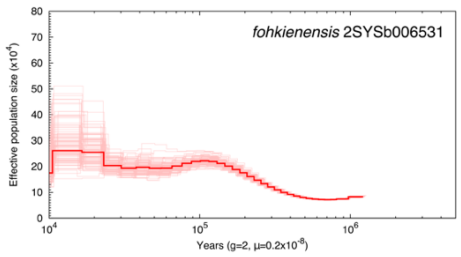


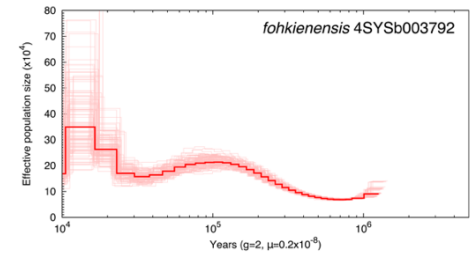
Figure S4. The concatenated tree based 91620 windows of whole genomic data from four *Riparia* species inferred by iqtree. Bootstrap values are shown for each node. Labels in the tree indicates the locality ID (the first number) in the map (Figure 1, top right) and the sample ID in Table 1.

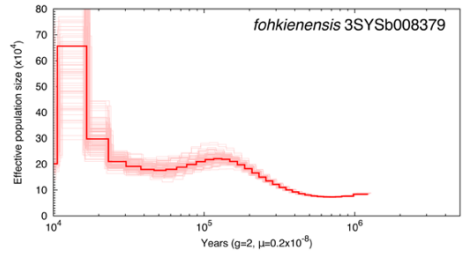

diluta 9SYSb006522
Effective population size (x10^4)
Years (g=2, μ=0.2x10^-8)

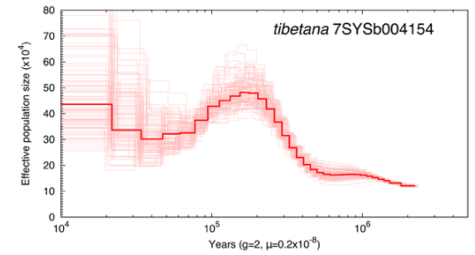

diluta 10SYSb007338
Effective population size (x10^4)
Years (g=2, μ=0.2x10^-8)

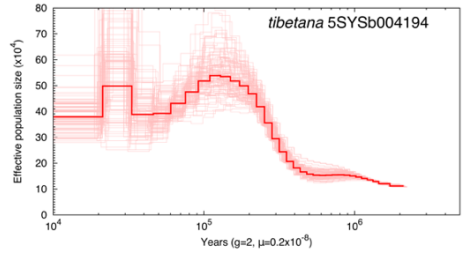

diluta 8SYSb006542
Effective population size (x10^4)
Years (g=2, μ=0.2x10^-8)

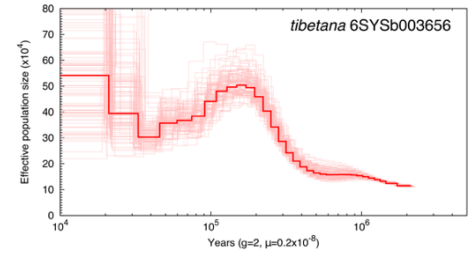

fohkienensis 2SYSb006531
Effective population size (x10^4)
Years (g=2, μ=0.2x10^-8)

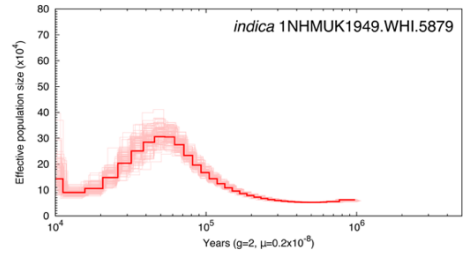

fohkienensis 4SYSb003792
Effective population size (x10^4)
Years (g=2, μ=0.2x10^-8)

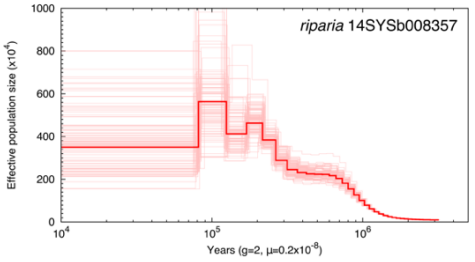

fohkienensis 3SYSb008379
Effective population size (x10^4)
Years (g=2, μ=0.2x10^-8)

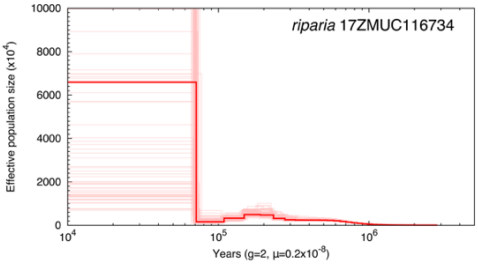

tibetana 7SYSb004154
Effective population size (x10^4)
Years (g=2, μ=0.2x10^-8)


tibetana 5SYSb004194
Effective population size (x10^4)
Years (g=2, μ=0.2x10^-8)


tibetana 6SYSb003656
Effective population size (x10^4)
Years (g=2, μ=0.2x10^-8)


indica 1NHMUK1949.WHI.5879
Effective population size (x10^4)
Years (g=2, μ=0.2x10^-8)


riparia 14SYSb008357
Effective population size (x10^4)
Years (g=2, μ=0.2x10^-8)


riparia 17ZMUC116734
Effective population size (x10^4)
Years (g=2, μ=0.2x10^-8)

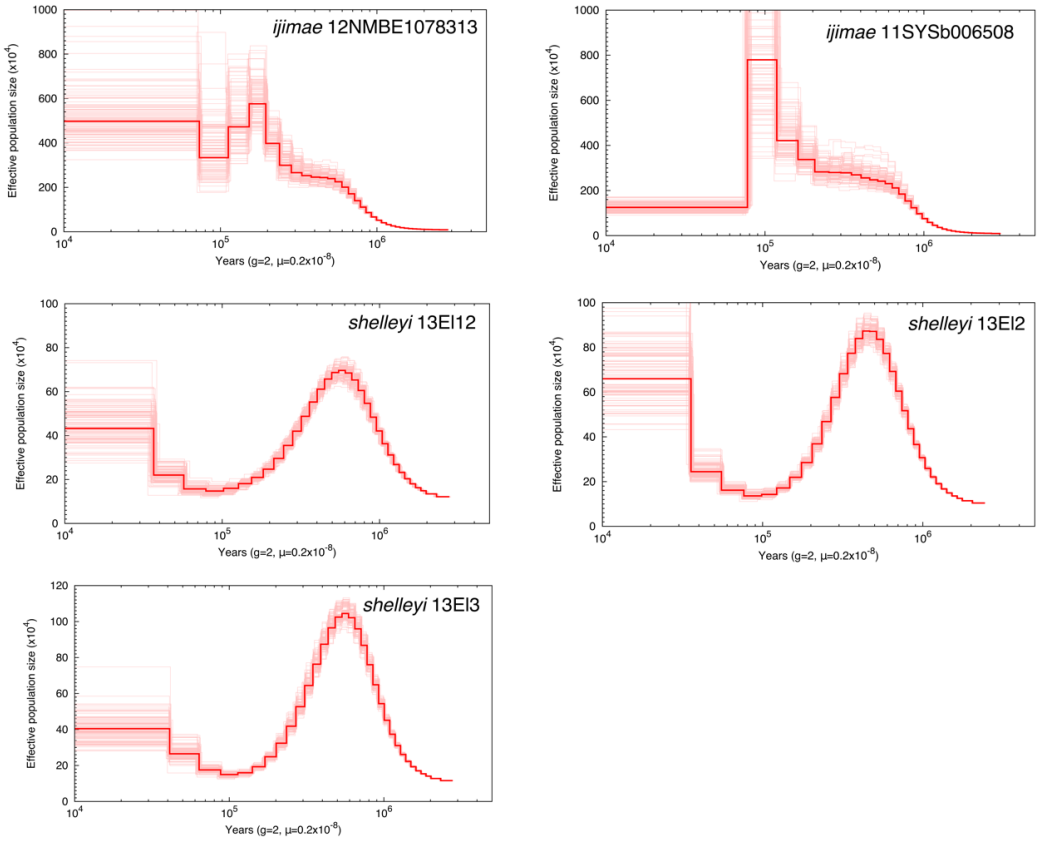


Figure S5. PSMC analysis of 100 bootstrap replicates of each individual in Figure 4.

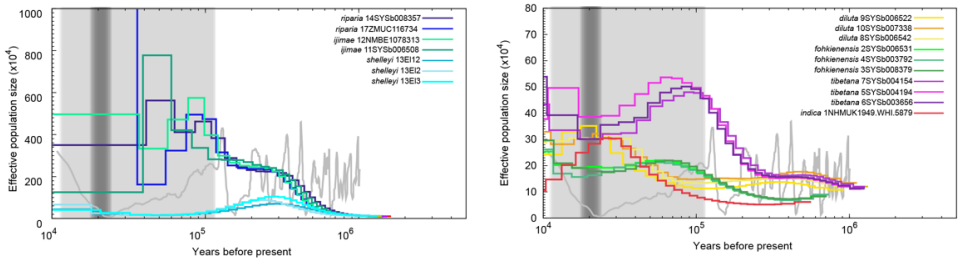


Figure S6. PSMC analysis based on the same parameters as in Figure 4 expect using generation time of one year.

R. d. indica BMMH1949.WHI.5879

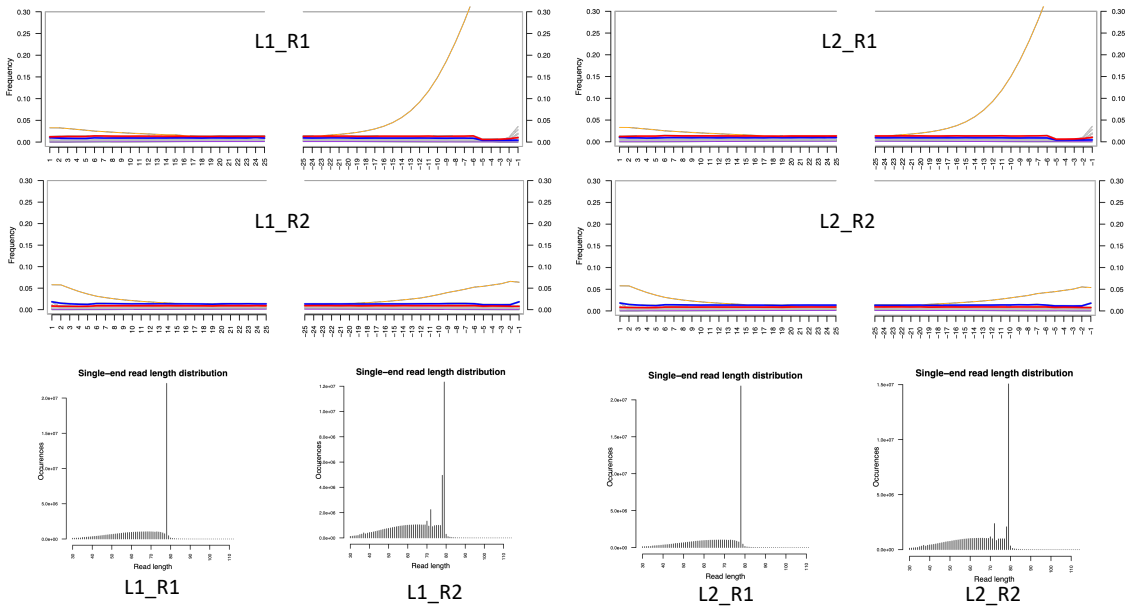


Figure S7. MD: mapDamage deamination plot (upper panel) of sequencing reads of each library from toe pad sample of BMMH1949.WHI.5879 mapped to the reference genome *R. riparia* and the length distribution of reads (lower panel). Red line: C to T substitutions; blue line: G to A substitutions; grey line: all other substitutions; orange line: Soft-clipped bases and purple line: insertions relative to the reference.

R. d. indica NHMUK1920.10.24.1

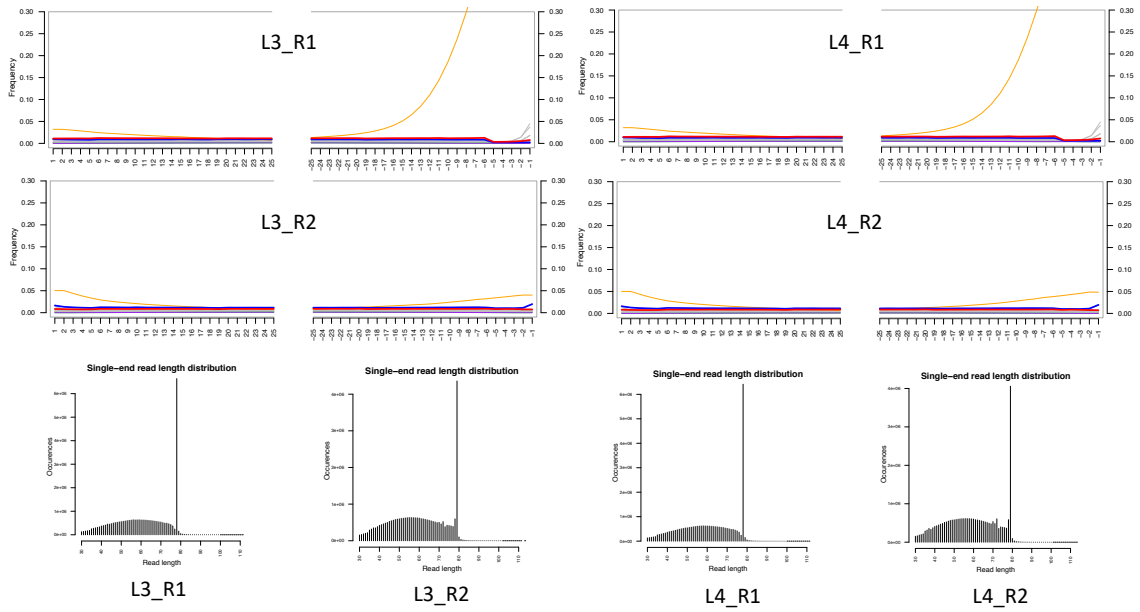


Figure S8. MD: mapDamage deamination plot (upper panel) of sequencing reads of each library from toe pad sample of NHMUK1920.10.24.1 mapped to the reference genome *R. riparia* and the length distribution of reads (lower panel). Red line: C to T substitutions; blue line: G to A substitutions; grey line: all other substitutions; orange line: Soft-clipped bases and purple line: insertions relative to the reference.

R. chinensis NRM-U-663

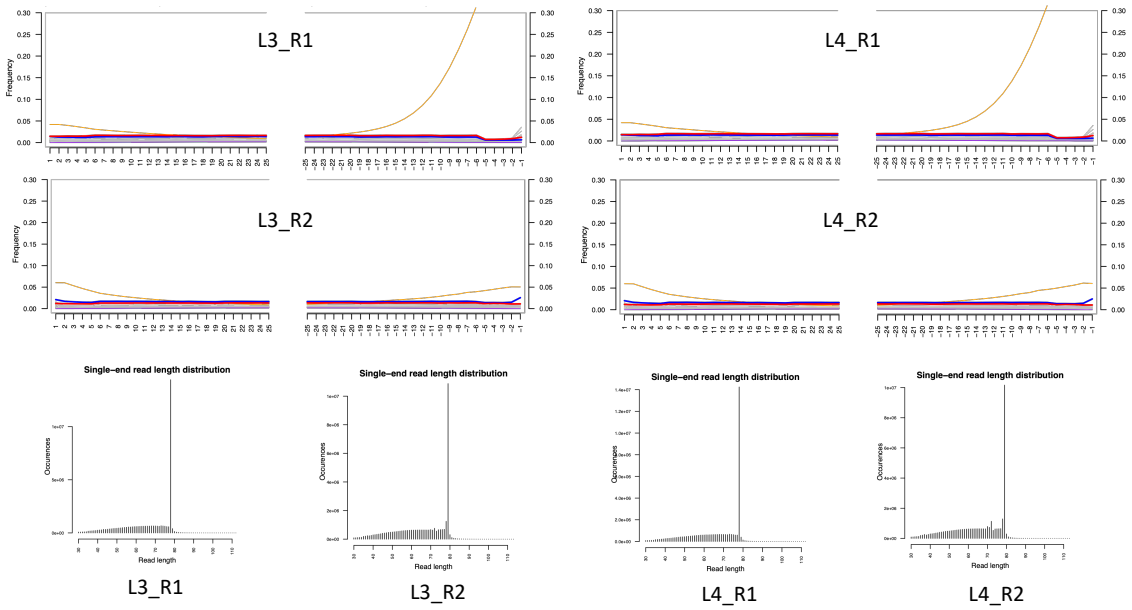


Figure S9. MD: mapDamage deamination plot (upper panel) of sequencing reads of each library from toe pad sample of NRM-U-663 mapped to the reference genome *R. riparia* and the length distribution of reads (lower panel). Red line: C to T substitutions; blue line: G to A substitutions; grey line: all other substitutions; orange line: Soft-clipped bases and purple line: insertions relative to the reference.

R. chinensis NRM-U-664

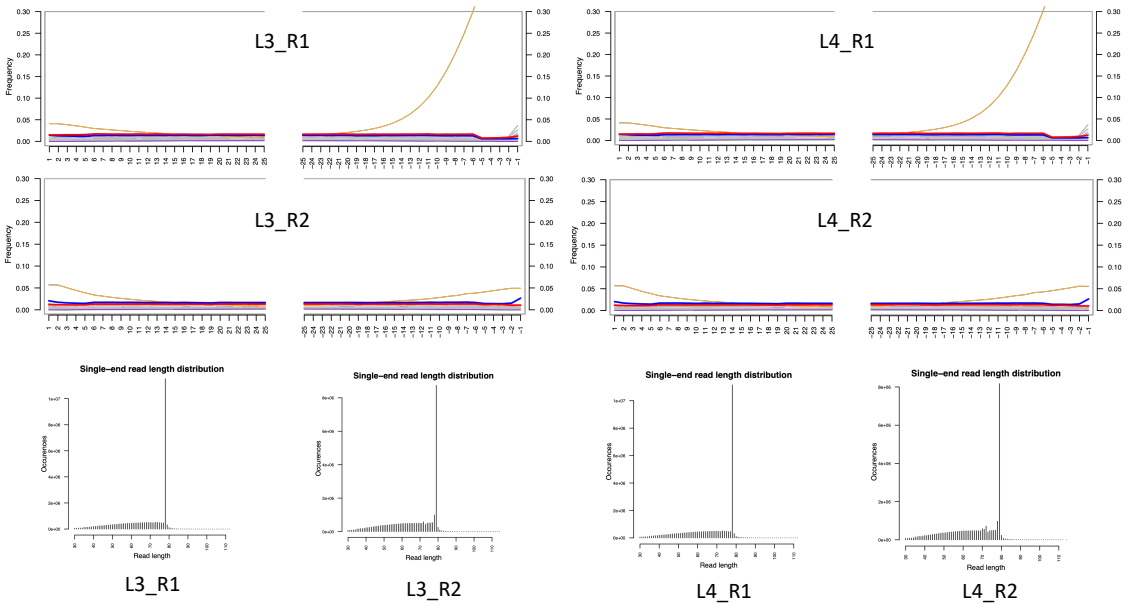


Figure S10. MD: mapDamage deamination plot (upper panel) of sequencing reads of each library from toe pad sample of NRM-U-664 mapped to the reference genome *R. riparia* and the length distribution of reads (lower panel). Red line: C to T substitutions; blue line: G to A substitutions; grey line: all other substitutions; orange line: Soft-clipped bases and purple line: insertions relative to the reference.

Chapter 3

Adaptive introgression of EPAS1 in a migratory bird species

Qindong Tang[1,2], Yang Liu[3], Gombobaatar Sundev[4], Manuel Schweizer[1,2], Gerald Heckel[1*]

[1] *Institute of Ecology and Evolution, University of Bern, Bern, Switzerland*;

[2] *Natural History Museum, Bern, Switzerland*;

[3]*State Key Laboratory of Biocontrol, College of Ecology School of Life Science, Sun Yat-sen University, Guangzhou, China*;

[4] *National University of Mongolia and Mongolian Ornithological Society, Ulaanbaatar, Mongolia*

Manuel Schweizer and Gerald Heckel are joint senior authors

Yang Li and Gombobaatar Sundev are intended to be invited as coauthors and have not revised the manuscript yet.

***Corresponding author:** E-mail: gerald.heckel@unibe.ch

Abstract

Our knowledge on genetic adaptations of resident species to life at high-altitude has significantly increased in the last years yet the potential genetic underpinnings of seasonal high-altitude lifestyle in migratory animals remains poorly understood. Seasonal migration in birds often involves shifts in elevation, imposing acute and often persisting physiological challenges. The pale sand martin species complex, characterized by evolutionary lineages breeding at different altitudes and seasonal migration with large elevation shifts, provides an excellent system to investigate high-altitude adaptation in migratory birds. In this study, we performed whole genome sequencing of populations breeding at high altitudes on the Qinghai-Tibetan plateau, and in the lowlands of Central Mongolia and south eastern China. We applied an F_{st}-based approach accounting for the effects of background and linked selection and haplotype-based methods to uncover genetic signatures associated with potential high-altitude adaptation in pale sand martin genomes. We identified EPAS1 and HBAD as strongest candidate genes for facilitating breeding at high altitudes in these migratory birds. Given that EPAS1 plays an important role in embryotic development which is severely harmed by chronic hypoxia, we hypothesize this state to be particularly relevant for adaptation to breeding at high altitudes. Our contrasts between high and low altitude populations of pale sand martin showed that high altitude haplotypes of EPAS1 might be selected against in the lowland environment. The range expansion of *R. d. tibetana* from the Qinghai-Tibetan plateau into lowland central Mongolia likely resulted in the introgressive replacement of EPAS1 haplotypes from the lowland subspecies *R. d. diluta*. Our finding illustrates that hybridization between divergent lineages can provide adaptive genetic variants to cope with different environments.

Key words: High-altitude adaptation, migration, adaptive introgression, *Riparia diluta*, pale sand martin

Introduction

Understanding how species adapt to extreme environments and the genetic mechanisms underpinning such adaptations has been of particular interest for the past decades. Living at high altitudes imposes the physiologically challenging condition of hypoxia (Hawkes et al., 2011), accompanied by other severe environmental conditions such as low temperature and intense ultraviolet radiation (Londoño et al., 2017; Scott, 2011). Extensive research has been conducted to investigate the genetic basis of high-altitude adaptation across various organisms, including humans residing in highland regions like the Qinghai-Tibetan Plateau and the Andes Mountains (Beall et al., 2010; Bigham et al., 2010; Cheng et al., 2021; Graham & McCracken, 2019; Li et al., 2018; Storz et al., 2007; Wang et al., 2014). Several genes were repeatedly found under strong positive selection for high-altitudes in the hypoxia-inducible factors (HIF) pathways (Bigham et al., 2013; Graham & McCracken, 2019). Interestingly, the specific variants associated with high-altitude adaptation in those genes were found to be distinct between different species and populations (Gou et al., 2014; Graham & McCracken, 2019; Pamenter et al., 2020; Peng et al., 2017; Schweizer et al., 2019). Furthermore, two processes were shown to be responsible for high altitude adaptions. Besides local adaptation caused by *de novo* mutations (Graham & McCracken, 2019; Hao et al., 2019), adaptive introgression was shown to facilitate high altitude adaptation. For example, the adaptive alleles identified in Tibetans were found to be the result of introgression from archaic hominins Denisovans (Huerta-Sánchez et al., 2014; Zhang et al., 2021a). However, these studies have mainly focused on species living permanently at high elevations where they experience chronic hypobaric hypoxia throughout their lifetimes. Less is known about potential genetic adaptation to hypoxia in migratory species that undergo large seasonal elevation shifts and only temporarily live at high altitudes.

Seasonal migration in birds between breeding and wintering grounds often involves shifts in elevation (Williamson & Witt, 2021). For instance, seasonal elevational migration is prevalent in birds breeding at high altitudes: up to 65% of bird species breeding at high elevations of the Himalayas and around 34% breeding on southern and eastern parts of Qinghai-Tibetan Plateau are migratory. These birds spend around three to four months at high altitudes during

the breeding season and move to lower elevations for the rest of the year (Dixit et al., 2016). Such seasonal changes over large altitude differences require pronounced physiological adjustments to different environments and impose a significant challenge for migratory birds. However, potential genetic adaptations underpinning such annual elevation shifts are still largely unknown (Campagna & Toews, 2022). Bar headed geese *Anser indicus* breeding on the Qinghai-Tibatan plateau and Central Asian Mountains and wintering chiefly on the Indian Subcontinent, as the best-studied species, showed adaptations to improve oxygen transport efficiency (Petschow et al., 1977; Scott et al., 2009) and a number of genes in the hypoxia inducible factor (HIF) pathway are under strong positive selection (Wang et al., 2020). While this might chiefly comprise adaptations to migrating at high altitudes in this species, the breeding period additionally represents one of the physiologically most demanding period in the annual cycle of birds (Dunn, 2004) and is thus prone for selection. Thus, we hypothesized that signs for genetic adaptations to hypoxia might be prevalent in migratory birds with seasonal elevation shifts that need to cope with temporary high energy demands under hypoxia environment during migration and breeding.

The genetic basis underlying adaptation to high-altitudes is probably best studied in migratory species complexes with variation among evolutionary lineages in both, elevation of breeding ranges and levels of divergence. In such a setting, the variability in genome-wide divergence and its contrast to divergence levels at specific loci, could make the detection of candidates for adaptation likely. The pale sand martin is a bird species complex that underwent a morphologically largely cryptic radiation into four evolutionary lineages occurring in Central, East and South Asia (Tang et al., 2022; Tang et al, in revision). The lineage of the subspecies *R. d. tibetana* is of particular interest, as it consists of highland populations breeding on the Qinghai-Tibetan plateau at altitudes of more than 3200 m asl, and others breeding much lower in central Mongolia below 1400m asl (Tang et al. 2022). The latter shows limited mixed ancestries with the subspecies *R. d. diluta* of lowland Central Asia. The area of contact between these subspecies lies in a well-known avian migratory divide where populations to the west take a western migratory route around the Qinghai-Tibetan plateau and those to the east an eastern one with hybrids with unfavourable intermediate migration routes being selected against (Irwin & Irwin, 2005; Scordato et al., 2020). Such a scenario might also be

prevalent in pale sand martins (Tang et al. 2022). The breeding range of highland *R. d. tibetana* abuts at the eastern edge of the Qinghai-Plateau with the subspecies *R. d. fohkienensis* that occupies the subtropical Chinese lowlands below 700 m asl (Figure 1). No signs for gene flow were detected between these two subspecies which might be caused by allochronic breeding with *R. d. fohkienensis* as a partial or short-distance migrant breeding much earlier than *R. d. tibetana* (Tang et. 2022). This combination of variation in elevation of breeding ranges and genomic diverge among evolutionary lineages in pale sand martin makes it a promising system to investigate the genetic basis of high-altitude adaptation in migratory birds.

We sequenced whole genomes of populations of *R. d. tibetana* breeding on the high elevation Qinghai-Tibetan plateau (> 3200m asl) and in the lowlands (< 1400m asl) of Central Mongolia. For comparison, we added genome data from populations of *R. d. diluta* and *R. d. fohkienensis* breeding at much lower elevation in distinct regions in north western China and south eastern China. We investigated the genomic landscape of differentiation between different highland and lowland populations with the aim to detect regions under strong selection that could be related to high-altitude adaptation.

Material and methods

Sampling and resequencing

We analyzed 100 individuals from 10 populations of *R. diluta* and *R. fohkienensis* on or adjacent to the Tibetan plateau covering an altitudinal range of 80 to 3599 meters above sea level (Figure 1, Table 1). This included three highland populations of *R. d. tibetana* on the Qinghai-Tibetan plateau, one lowland population of *R. d. tibetana* in central Mongolia, three lowland populations of *R. d. diluta* in northwestern China, and three lowland populations of *R. d. fohkienensis* from southeastern China (Figure 1, Table S1). Here, we define breeding grounds on the Qinghai-Tibetan plateau as highland and the remaining (< 1500 m) as lowland. Blood samples were preserved in 99% ethanol and later stored at −20°C. DNA was extracted with a modified salt extraction protocol (Aljanabi & Martinez, 1997). Genomic libraries were prepared using TruSeq DNA PCR-free sample preparation (Illumina), and then processed for paired-end 150 bp read sequencing on an Illumina NovaSeq 6000 (S4 flow cell).

De novo reference genome sequencing, assembly and gene annotation

First, a reference genome from *R. d. fohkienensis* was sequenced and assembled *de novo* analogous to Tang et al. (2022). DNA was extracted using the Qiagen MagAttract HMW DNA kit and linked-read sequencing was performed as in Lutgen et al (2020). A single library was prepared using the 10× Genomics Chromium Genome library kit and sequenced on half an S4 lane on a NovaSeq 6000 instrument. The genome was assembled *de novo* with supernova version 2.1.0 (Weisenfeld et al., 2017), resulting in a pseudohaploid assembly with a total length of 1.14 Gb, an effective read coverage of 45.0x and a scaffold N50 of 15.8 Mb. Evaluation of the draft genome with Benchmarking Universal Single-Copy Orthologue (BUSCO version 4.1.4) (Manni et al., 2021) based on avian ortholog data set (aves_odb9, N=4,915) yielded a completeness score of 91.4%. After anchoring the draft assembly to a chromosome level assembly of a *Hirundo rustica* (barn swallow) genome (NCBI assembly bHirRus1.pri.v2) using minimap2 (H. Li, 2018) and masking the repeat regions using RepeatMasker 4.1.2-p1 (Smit et al., 2013), we obtained a chromosome level reference genome of *R. d. fohkienensis* of 992 Mb. We used the homology-based gene prediction program GeMoMa (Gene Model Mapper) version 1.7.1 (Keilwagen et al., 2016, 2018) to annotate our reference genome using gene models from the well annotated chicken (*Gallus gallus*) genome assembly (GenBank assembly accession: GCA_016699485.1) which resulted in a total number of 9831 predicted/identified genes.

Data preparation and SNPs calling

The quality of raw reads was first checked using fastqc version 0.10.1 (Andrew 2019). Reads were then trimmed and paired using trimmomatic version 0.39 (Bolger et al., 2014) with parameters CROP:145 HEADCROP:8 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20. The quality of trimmed and paired reads was further checked and showed good quality using fastqc. Paired reads were then mapped on the chromosome-anchored *Riparia f. fohkienensis* reference genome using BWA 0.7.17 (bwa men) (H. Li, 2013). For each bam file, only mapped alignments with mapping quality (MAPQ) larger than 30 were retained using SAMTOOLS version 1.9 (Danecek et al., 2021) and duplicates were then removed using MarkDuplicates in PICARD tools version 2.25 (http://broadinstitute.github.io/picard/). Variants were called with

HaplotypeCaller using GATK version 4.1 (Auwera & O'Connor, 2020), following the GATK4 best practice workflow for SNP and indel calling using GenomicsDBImport to merge GVCFs (Genomic Variant Call Format) from multiple samples (https://gatk.broadinstitute.org/hc/en-us/articles/360036883491-GenomicsDBImport). Joint genotyping was conducted for all individuals and we kept invariable sites for F_{st} estimation. We used vcftools version 0.1.15 (Danecek et al., 2011) to keep only the sites with minimum read depth of 5x per individual, less than 10% missing genotypes and a maximum of two alleles for the analysis.

PCA and admixture analysis

We first wanted to check whether the population structure within Pale Sand Martins was consistent between resequencing data and GBS (genotyping by sequencing) data of our previous study (Tang et al., 2022). Only variable sites from autosomes with MAF (Minor allele frequency) > 0.05 and located at least 10 kb apart and thus considered as unlinked were kept using vcftools version 0.1.15 (Danecek et al., 2011). We conducted a principal component analysis (PCA) using PCAngsd (Meisner & Albrechtsen, 2018) with eigenvectors computed from the covariance matrix with the function 'eigen' in R 3.6.2 (R Core Team, 2019). We also performed an admixture analysis using ADMIXTURE version 1.3.0 (Alexander et al., 2009) to check for mixed ancestries of individuals, with the number of ancestral populations K from 2 to 10, and we chose the best fit for K based on cross-validation. Input files for PCAngsd and ADMIXTURE were generated with plink version 1.90 (http://pngu.mgh.harvard.edu/purcell/plink/) (Purcell et al., 2007).

Sliding window scan for outlier regions of highland-lowland population differentiation

In order to capture potential selective signatures of highland-lowland differentiation and to reduce the effects of inflated population differentiation as a consequence of heterogeneous recombination rates across the genome and background selection (Burri, 2017), we calculated Delta F_{st} (Roesti et al., 2014) as F_{st} between highland and lowland populations scaled by F_{st} among all lowland populations. We estimated highland-lowland differentiation as the mean of pairwise Fst for non-overlapping 50 kb windows along the autosomes between three highland *R. d. tibetana* populations and three lowland populations each of *R. d. diluta* and *R. d. fohkienensis* (in total 18 pairs). Lowland-lowland differentiation was estimated for

each window as the mean of the pairwise comparisons between 3 lowland populations of *R. d. dilu*ta and 3 lowland populations of *R. d. fohkienensis* (in total 9 pairs). We then calculated Delta F_{st} (DeltaFst) for each 50 kb window as the mean highland-lowland F_{st} minus mean lowland-lowland F_{st}. Windows with DeltaFst larger than the 99% percentile of the genome-wide DeltaFst were considered as peaks. At this step, we excluded the lowland population of *R. d. tibetana* from the DeltaFst calculation due to their known history of introgression (Figure 1, Tang et al. (2022)).

To investigate highland-lowland differentiation within *R. d. tibetana* and the potential influence of introgression from *R. d. diluta* into the lowland *R. d. tibetana* population, we additionally calculated pairwise Fst-values between the three highland *R. d. tibetana* populations and the lowland *R. d. tibetana* population, and between the three lowland *R. d. diluta* populations and the lowland *R. d. tibetana* population separately. Genomic regions that introgressed from *R. d. diluta* into the lowland *R. d. tibetana* population would show lower F_{st}-values in the comparison of *R. d. diluta* and lowland *R. d. tibetana* than the F_{st}-values from the comparison of highland *R. d. tibetana* and lowland *R. d. tibetana*. All F_{st}-values were calculated using vcftools version 0.1.15 (Danecek et al., 2011).

Haplotype-based scan for regions under selective sweeps for high altitude adaptation

We used a second approach to detect signatures of selective sweeps indicating local adaptation. To this end, we tested for different haplotype structure between populations relying on the assumption that regions under selective sweeps in one population would retain longer homozygous haplotypes at high frequency compared to a population where these selective processes are not prevalent (Szpiech et al., 2021). We calculated XP-nSL (Ferrer-Admetlla et al., 2014; Szpiech et al., 2021), a haplotype-based statistic which does not require a genetic recombination map, using selscan version 1.3.0 (Szpiech & Hernandez, 2014). We first phased SNPs using shapeit4 version 4.2.2 (Delaneau et al., 2019) combined with whatshap version 1.7 (Martin et al., 2016) as suggested in the shapeit4 protocol (https://odelaneau.github.io/shapeit4/). Each individual was first phased by whatshap as a pre-processing step to extract phase information from bam files separately. Then we grouped the individuals from the same subspecies separately to compute haplotypes using shapeit4.

All 30 individuals from *R. d. fohkienensis* were also phased for the haplotype network reconstruction (see below). Raw XP-nSL scores across the autosomes were calculated by selscan (selscan flags -xpnsl) for two different highland-lowland population pairs: the first (referred to as xpnsl_t) used all 30 individuals from highland *R. d. tibetana* populations against the 10 individuals from the lowland *R. d. tibetana* population as reference population. The second (referred to as xpnsl_d) used the same highland *R. d. tibetana* individuals against the 30 individuals from *R. d. diluta* as reference population. Positive XP-nSL scores correspond to long homozygous haplotypes and a potential sweep in highland populations compared to the lowland population, and vice versa for negative XP-nSL scores (Szpiech et al., 2021). Then the XP-nSL scores were normalized, the genome divided into non-overlapping 50k windows and highest positive and lowest negative scores were assigned to each window using selscan's norm version 1.3.0 (norm flag -xpnsl -bp-win -winsize 50,000). We identified the windows having the top 1% extreme positive scores as peaks representing potential regions under selective sweep indicating high altitude adaptation.

Gene Ontology enrichment analysis

We obtained three datasets of annotated genes in the outlier windows identified based on DeltaFst, xpnsl_d and xpnsl_t. A Venn diagram summarizing the overlap in the results was generated using the online version of BioVenn: http://www.biovenn.nl/index.php (Hulsen et al., 2008). A functional enrichment analysis was performed for each gene dataset separately with GO (Gene Ontology) using g: GOSt in g: Profiler (Raudvere et al., 2019). We applied the multiple testing correction method g: SCS, excluded the electronic GO annotations as suggested by gProfiler (https://biit.cs.ut.ee/gprofiler/page/docs#electronic_annotations_iea) and used the annotated genes from the *R. d. fohkienensis* reference genome as background gene set.

Haplotype network reconstruction

We extracted the CDS regions of genes from the phased data set of all 100 individuals located within the identified outlier windows shared among the three analyses: DeltaFst, xpnsl_d and xpnsl_t. We used a python script fasta2vcf.py (https://github.com/santiagosnchez/vcf2fasta) to transform vcf into fasta format and aligned the CDS regions using MEGA X (Kumar et al.,

2018) with CLUSTAL W (Thompson et al., 1994). We then reconstructed minimum spanning networks using PopART (Leigh & Bryant, 2015) based on input nexus files generated with DNAsp 6 (Rozas et al., 2017). DNA alignments were also translated into amino acid (AA) alignments, and we then reconstructed median-joining (MJ) networks for genes located in the outlier windows with more than two different amino acid (AA) haplotypes using Network 10.2.0.0 (fluxus-engineering.com)(Bandelt et al., 1999).

Topology weighting analysis

We performed topology weighting to check for gene tree discordancy along the chromosomes with shared outlier windows among the three analyses (DeltaFst, xpnsl_d and xpnsl_t) using Twisst (Martin & Van Belleghem, 2017) based on sliding windows containing 50 SNPs. Individual-locus neighbor-joining trees were reconstructed with PhyML version 3.3 (Guindon et al., 2010) based on phased genotypes using the script (phyml_sliding_windows.py) provided in the Twisst pipeline. All 100 individuals were included for the tree reconstruction and we defined four taxa as: *R. d. fohkienensis*, *R. d. diluta*, Lowland *R. d. tibetana* and Highland *R. d. tibetana*. The trees were rooted using *R. d. fohkienensis* as outgroup.

Results

PCA and admixture

The average mean read depth per individual after SNPs calling and filtering was 13.2X (range: 10.6-18.4X). We obtained 87,966 unlinked SNPs from autosomes for the PCA and admixture analysis. *R. d. fohkienensis*, *R. d. diluta* and *R. d. tibetana* were well separated from each other in the PCA, and all *R. d. tibetana* individuals from lowland Mongolia clustered close to the other individuals of the subspecies (Figure 1B). Admixture analyses resulted in K = 3 as the best-fitting number of ancestral populations consistent with strict individual level clustering into the three subspecies (Figure 1C). In agreement with our earlier analyses with a much smaller dataset (Tang et al. 2022), the population of lowland *R. d. tibetana* in Mongolia showed evidence of limited mixed ancestries with north-western Chinese *R. d. diluta*.

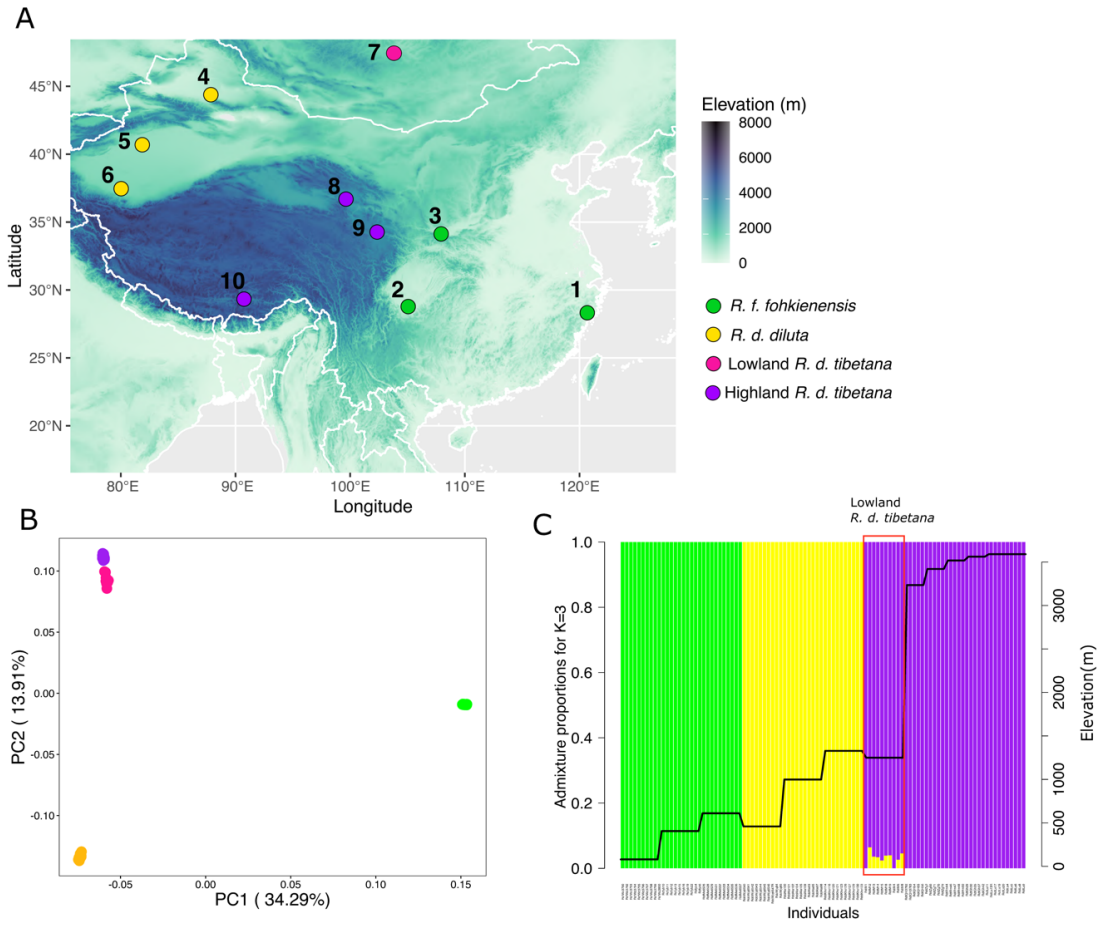


Figure 1. A) Sampling map of *R. d. tibetana* (purple), *R. d. diluta* (yellow), and *R. d. fohkienensis* (green) populations on or adjacent to the Tibetan plateau. Elevation is indicated with background color. Numbers correspond to population identifiers (PopID, see Table S1). B) Scatterplot of a principal component analysis (PCA) based on 87,966 unlinked autosomal SNPs. C) Bar plots showing individual ancestry assignments from an admixture analysis with K=3. The lowland population of *R. d. tibetana* showing evidence of admixture with R. d. diluta is marked with a red frame.

Genomic signatures of high-altitude adaptation in *R. d. tibetana*

DeltaFst analyses yielded outliers within the top 1% windows in which we found 81 annotated genes. The outlier windows detected with the haplotype scan with XP-nSL detected 136 annotated genes for the *tibetana*-high vs *diluta*-low contrast and 143 annotated genes for the *tibetana*-high vs. *tibetana*-low contrast (Table S2). No GO terms were found significantly enriched among these genes. However, three outlier windows were shared among the three approaches (Figure 2) and these contained 7 annotated genes: EPAS1 on chromosome 3; MAJIN, GOLGB1, LOC430443 on chromosome 4; HBAD, HBZ and NPRL3 on chromosome 15

(Figure 2, Figure 5). The window containing EPAS1 also had the genome-wide highest F_{st} value ranging from 0.69-0.74 for each comparison between highland *R. d. tibetana* populations on the Qinghai-Tibetan Plateau and the lowland *R. d. tibetana* population from Mongolia (genome-wide average F_{st}= 0.04; Figure 3). However, the average pairwise F_{st} of the window containing EPAS1 between populations from lowland Mongolia *R. d. tibetana* and north-western Chinese *R. d. diluta* was 0.09 and thus much lower. In contrast, their genome-wide average pairwise F_st was 0.23. The other two shared outlier windows did not show very low F_{st} between lowland *R. d. tibetana* and *R. d. diluta* or elevated F_{st} between highland and lowland comparisons within *R. d. tibetana*. The F_{st} value of these two outlier windows between lowland *R. d. tibetana* and *R. d. diluta* ranged from 0.19-0.25. For highland and lowland comparisons within *R. d. tibetana*, the F_{st} value of the outlier window on chromosome 4 ranged from 0.08-0.25, and the one on chromosome 15 from 0.03-0.07.

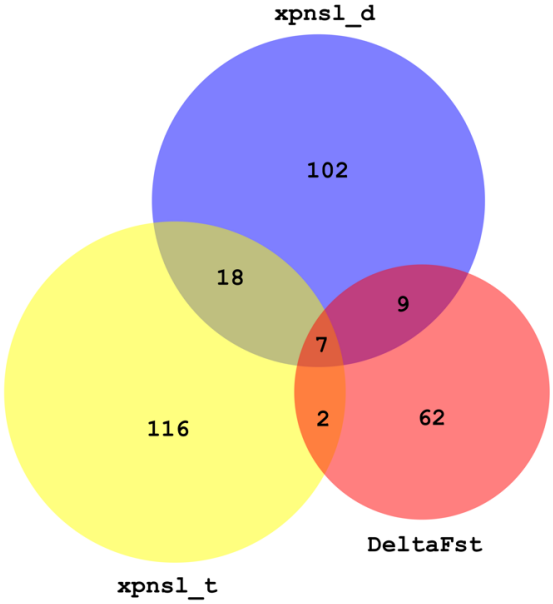


Figure 2. Venn diagram of annotated genes in the genomic outlier windows identified by three different analyses (xpnsl_d, xpnsl_t, DeltaFst). Seven genes in three 50 kb windows were shared among all the three analyses: EPAS1 (chromosome 3), MAJIN, LOC430443, GOLGB1 (chromosome 4), HBAD, HBZ and NPRL3 (chromosome 15).

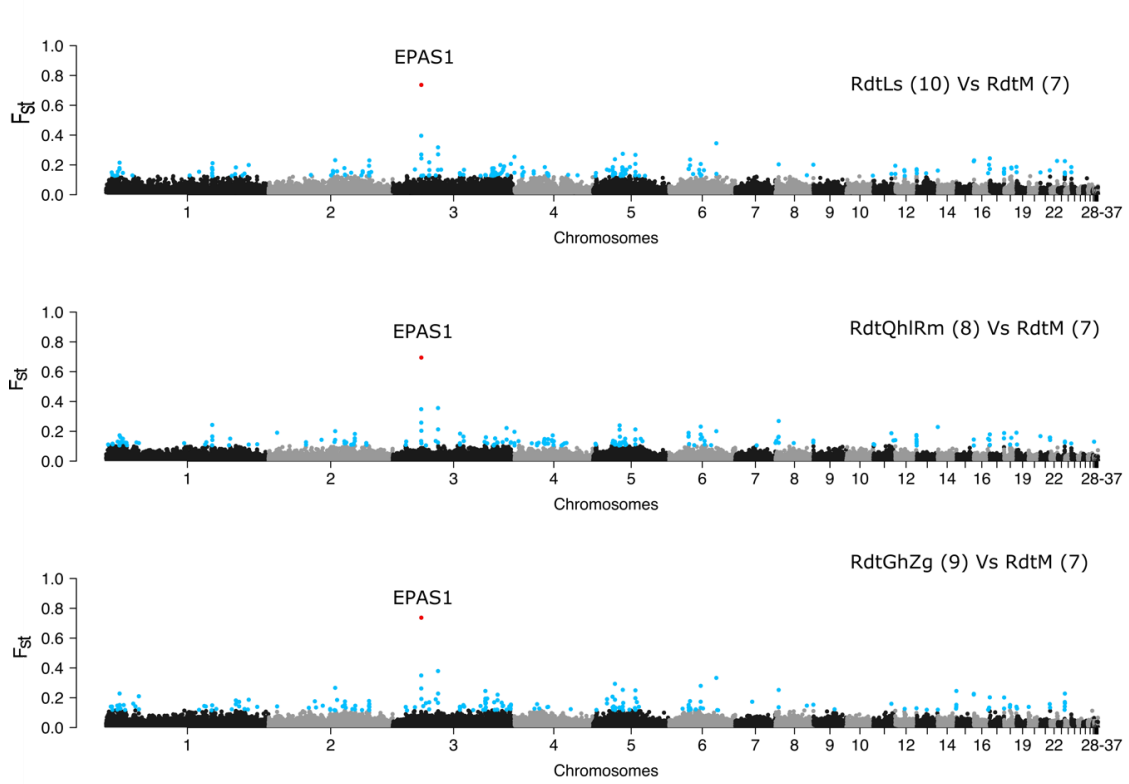


Figure 3. Pairwise F_{st} between highland and lowland populations of *R. d. tibetana* in non-overlapping 50kb windows along the genome. Each dot indicates the mean F_{st} for each window. Mean F_{st} values in the top 1% percentile were marked in blue and the window containing EPAS1 with the genome-wide highest mean F_{st} value was marked in red. Numbers in the brackets correspond to the PopID in the sampling map in Figure 1.

Haplotype network and gene tree discordance

Haplotype networks for the CDS of the seven annotated genes found in outlier windows showed stark differences. We obtained 2574 bp of the CDS region of EPAS1. Comparison with the barn swallow EPAS1 AA sequence from NCBI (Sequence ID: XP_0399136858.1; best match according to BLAST) revealed one deletion of six amino acids (AA) and an insertion of one AA. The nucleotide network of EPAS1 from all 100 individuals (Figure 4) showed three major clusters of haplotypes consisting of the following populations: highland *R. d. tibetana*, *R. d. diluta* together with all lowland *R. d. tibetana,* and *R. d. fohkienensis*. Thus, lowland *R. d. tibetana* shared several haplotypes with *R. d. diluta* and were different from highland *R. d. tibetana* contrasting with the overall genetic differentiation between these populations (compare Figure 4 and Figure 1). Lowland populations of all three subspecies shared AA haplotypes of EPAS1, while highland *R. d. tibetana* possessed distinct haplotypes (Figure S1).

The most common highland AA haplotype (Hap10 in Figure S1) differed by four amino acid changes from to the most common lowland EPAS1 AA haplotype (Hap1 in Figure S1): His (H) to Tyr (Y) on position 613, Gln (Q) to Lys (K) on position 662, Ser (S) to Gly (G) on position 776 and V (Val) to M (Met) on position 837. All these four AA changes were shared with the remaining two highland AA haplotypes (H11 and H12 in Figure S1). However, each of these AA changes were not unique to highland population, but also occurred in *R. d. diluta*.

For the remaining six candidate genes, highland and lowland *R. d. tibetana* largely shared nucleotide haplotypes and AA haplotypes were shared between all the three subspecies (Figure 4, Figure S1). No AA changes were found among all the samples for MAJIN as well as for the CDS region of NPRL3 within the outlier window. Only two AA haplotypes were found in LOC430443 and HBZ with one AA haplotype shared by all populations. GOLGB1 had the highest haplotype diversity but most haplotypes of lowland *R. d. tibetana* clustered with highland *R. d. tibetana* (Figure S1). For HBAD, we found a common unique AA change to A for highland *R. d. tibetana* at position 38 (Hap3 in Figure S1). Highland birds were heterozygous with amino acid of Q and A at this site except for three individuals heterozygous with Q and P. However, the lowland birds of *R. d. tibetana* and *R. d. diluta* were heterozygous with Q and P at this site except for two *R. d. diluta* individuals homozygous for P.

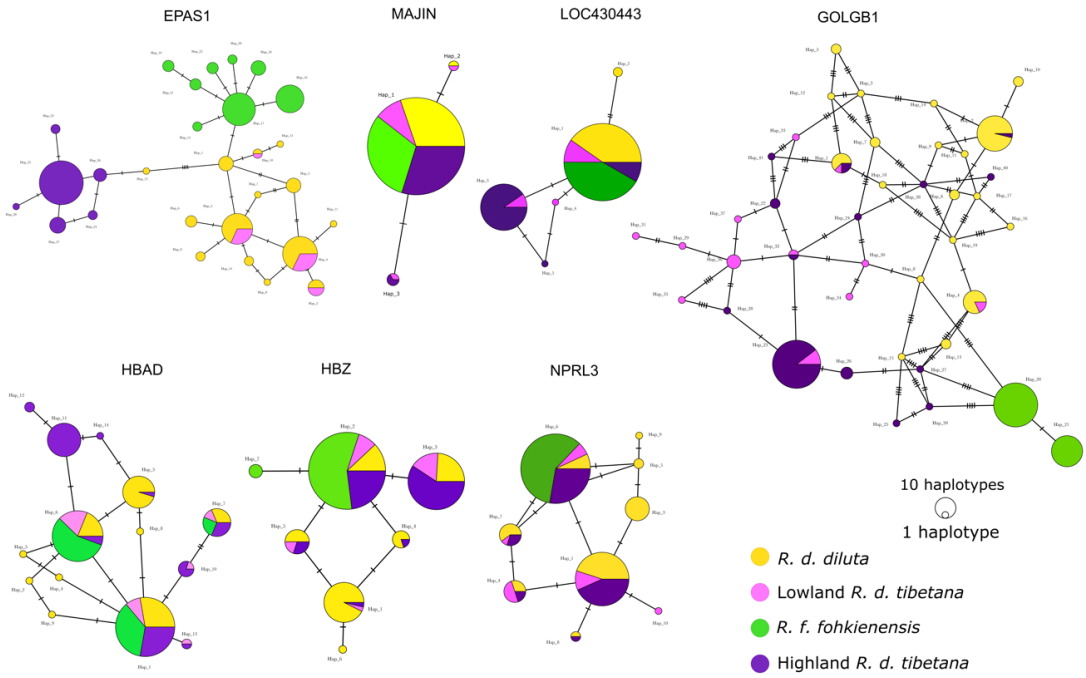


Figure 4. Haplotype network based on the CDS (coding sequence) regions of 7 candidate genes for high altitude adaptation in *R. d. fohkienensis*, *R d. diluta* and *R. d. tibetana*. Only CDS regions within the identified outlier window were extracted for the haplotype network reconstruction.

Topology weighting analysis revealed that the majority of the windows had topologies congruent with the likely species tree: *R. d. diluta* as sister group to a clade consisting of lowland and highland populations of *R. d. tibetana* (Chapter 2; topology 1, Figure 5). The larger genomic region around EPAS1 showed the alternative topology with lowland *R. d. tibetana* clustering with *R. d. diluta* as the dominant topology (topology 3) only in the window with this gene and its immediate vicinity (Figure 5). The remaining two shared outlier windows had the species tree as dominant topology for the candidate genes consistent with the surrounding genomic regions (Figure 5).

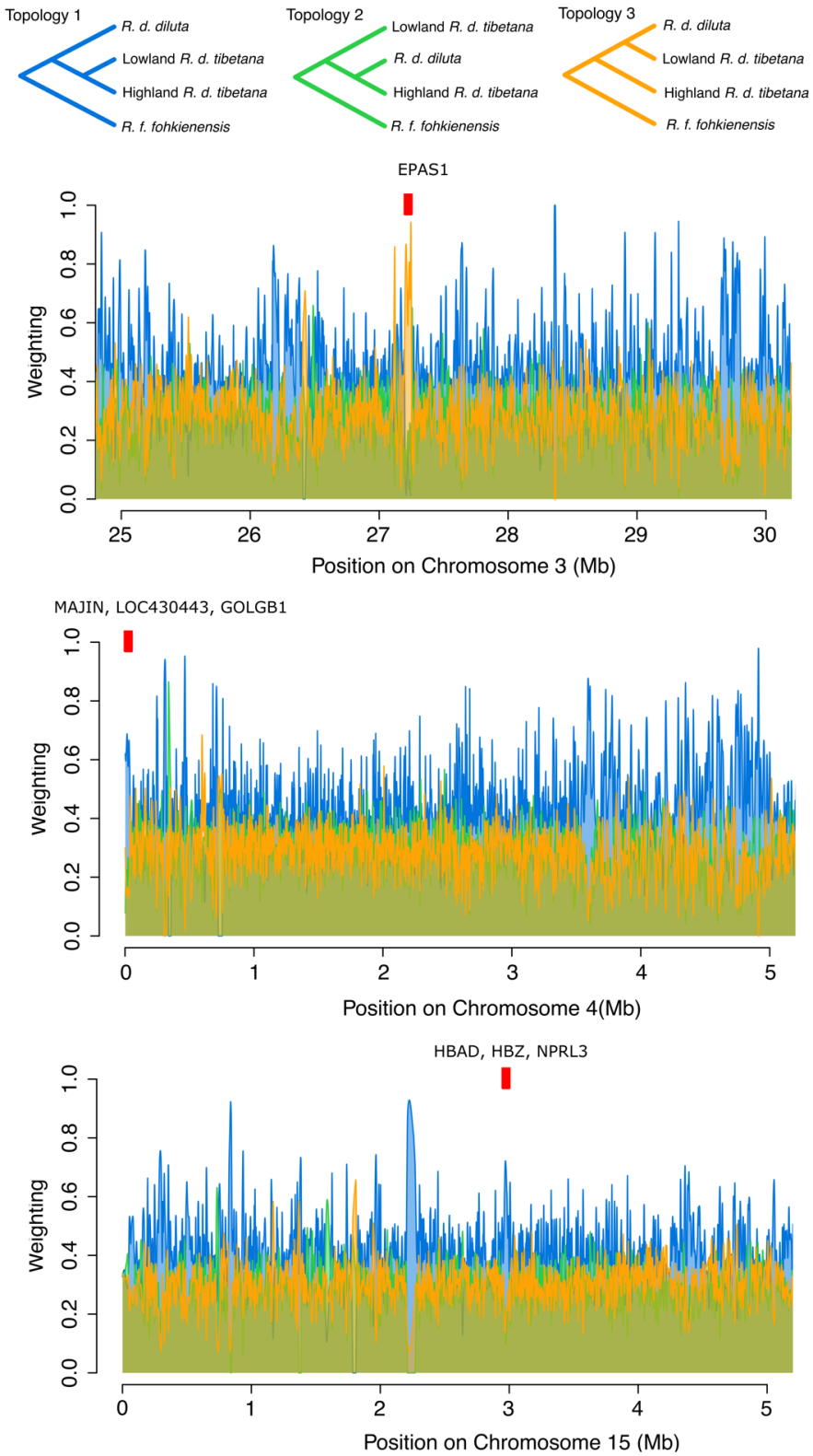

Topology 1
R. d. diluta
Lowland R. d. tibetana
Highland R. d. tibetana
R. f. fohkienensis
Topology 2
Lowland R. d. tibetana
R. d. diluta
Highland R. d. tibetana
R. f. fohkienensis
Topology 3
R. d. diluta
Lowland R. d. tibetana
Highland R. d. tibetana
R. f. fohkienensis
EPAS1
Weighting
0.0
0.2
0.4
0.6
0.8
1.0
25
26
27
28
29
30
Position on Chromosome 3 (Mb)
MAJIN, LOC430443, GOLGB1
Weighting
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
Position on Chromosome 4(Mb)
HBAD, HBZ, NPRL3
Weighting
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
Position on Chromosome 15 (Mb)

Figure 5. Topology weighting along 5 Mb on chromosomes 3, 4 and 15 of *Riparia* populations containing the outlier regions for high altitude adaptation. Weights are smoothed using loess smoothing (span = 20 kbp) in R. Three alternative potential topologies of populations are shown on top: topology 1 is the species tree (Chapter 2), topology 2 clusters highland *R. d .tibetana* with *R. d. diluta*, topology 3 clusters lowland *R. d. tibetana* with *R. d. diluta*. The positions of the three shared outlier windows with candidate genes are marked in red.

Discussion

The aim of this study was to test for signs for potential genetic adaptation to high altitude in the genomic architecture of migratory birds with large seasonal elevation shifts. By studying the genomic differentiation between lowland and highland populations within the pale sand martin species complex, we uncovered regions that showed strong indications of selection for high-altitude adaptation. Among these regions, we found strongest support for EPAS1 as candidate gene. Interestingly, EPAS1 haplotypes in a lowland population of *R. d. tibetana* were completely replaced by haplotypes from the lowland sister lineage *R. d. diluta* indicating adaptive introgression.

Genetic signatures of high-altitude adaptation

By applying two methodological approaches, we detected genomic regions that might be involved in adaptation to cope with harsh conditions of high altitudes. The DeltaFst approach was applied to capture the overall genetic differentiation between highland and lowland populations. Regions with particularly high DeltaFst thus might be associated with high-altitude adaptation. Meanwhile, in the second approach, high positive XP-nSL scores should indicate potential regions under selective sweeps in the highland population of *R. d. tibetana* compared with lowland populations of *R. d. tibetana* and *R. d. diluta*. We expected regions under strong selection for high-altitude adaptation to be identified as outliers in both approaches. However, there was considerable variation among the genes identified by different analyses (Figure 2). At least two factors could explain the variation of identified genes between the different approaches. First, demographic history could confound the inferences of selection (Excoffier et al., 2009). Lowland *R. d. tibetana* likely recently expanded

to Central Mongolia from the Tibetan Plateau (Tang et al. in revision). Different alleles could be fixed in lowland *R. d. tibetana* during the range expansion into Central Mongolia just due to drift and mimic signatures of selection such as a high F_{st} value when compared to the highland population. Second, adaptive signals unrelated to adaption to high elevation conditions might confound the results. For example, the migration divide might play a key role in maintaining the genetic differentiation between *R. d. tibetana* and *R. d. diluta* (Tang et al., 2022). If the migration route of populations on the Qinghai-Tibetan plateau differs from the lowland *R. d. tibetana* (Tang et al., 2022), signals of selection from the comparisons between *R. d. tibetana* and *R. d. diluta* could be associated with the migration divide rather than high altitude adaptation.

We found three outlier windows shared by the three analyses and there were seven annotated genes within these windows (Figure 2). Three of them are directly related to hypoxia or oxygen delivery (Gou et al., 2007; Taylor et al., 2016): EPAS1 (Endothelial PAS domain-containing protein 1), HBAD (Hemoglobin alpha, subunit D) and HBZ (Hemoglobin subunit Zeta). One gene NPRL3 is related to cardiac muscle tissue development (Kowalczyk et al., 2012). The remaining genes are involved in meiotic telomere clustering (MAJIN) (Wang et al., 2019), regulation of calcium ion binding (LOC430443) (Braunewell & Gundelfinger, 1999) and protein glycosylation regulation (GOLGB1) (Lan et al., 2016). Interestingly, among the three genes with function related to hypoxia or oxygen delivery, only EPAS1 and HBAD showed amino acid changes between highland and lowland populations and EPAS1 showed the largest differences between highland and lowland populations (Figure S1).

Selective sweeps on EPAS1 on the Qinghai-Tibetan Plateau were suggested from both selection scans and haplotype networks. Two major clusters were found for EPAS1 in both nucleotide and amino acid haplotype networks. Highland EPAS1 haplotypes were separated from most lowland haplotypes and showed lower haplotype diversity than lowland EPAS1 (Figure 4, Figure S1). Considering that all single amino changes found in the highland EPAS1 haplotypes also occurred in lowland *R. d. diluta*, we hypothesized that the highland population first expanded to the Qinghai-Tibetan plateau from the lowland and natural

selection then favored recombined alleles leading to haplotypes with adaptation for living at high altitudes.

Numerous studies (e.g. Graham & McCracken, 2019; Storz et al., 2007; Wang et al., 2014) have shown that a number of genes in the hypoxia inducible factor (HIF) pathway are under strong positive selection for chronic hypoxia stress. Among those, EPAS1 has been found repeatedly to be a target for selection across a variety of species (Beall et al., 2010; Graham & McCracken, 2019; Storz et al., 2007). However, there is variation among species in how distinct EPAS1 haplotypes from highlands are from those of the lowland. All EPAS1 variants found in Tibetan humans are located in noncoding regions (Peng et al., 2017) while amino acid changes were identified in Tibetan dogs (Gou et al., 2014), deer mice (Schweizer et al., 2019) and two Andean duck species, speckled teal and yellow-billed pintail (Graham & McCracken, 2019). Most nonsynonymous changes found in EPAS1 between lowland and highland populations from the Andean ducks occurred in exon 12 (Graham & McCracken, 2019). Interestingly, two out of four amino acid changes for highland EPAS1 in *R. d. tibetana* were also located in exon 12. This region encodes ODD (O_2-dependent degradation)/NTAD (N-terminal transactivation) domain of EPAS1 protein which is crucial for the HIF function (Jaakkola et al., 2001). Under normoxic conditions, HIF-2α (Hypoxia-inducible factor 2-alpha) encoded by EPAS1 is degraded via post-translational modifications. The degradation is activated by hydroxylation of specific proline residues within the ODD domain (Appelhoff et al., 2004; Jaakkola et al., 2001; Ivan et al., 2001; Maxwell et al., 1999). However, under hypoxia conditions, oxygen-dependent degradation is arrested; accumulated HIF-2α then forms HIF complexes which can recognize HREs (hypoxia response elements) within the promoters and activates the expression of hypoxia-associated genes. High conservation of exon boundaries has been shown between human and bird EPAS1 (Graham & McCracken, 2019). Mutations in human EPAS1 exon 12 were found to be associated with erythrocytosis (Percy et al., 2008) which involves an excessive increase in red blood cell production, enhancing the oxygen-carrying capacity of the blood in response to reduced oxygen availability. Further studies are needed to demonstrate whether AA changes in our system might have any functional consequences related to hypoxia.

Furthermore, we also identified a common unique amino acid change in HBAD that only occurred in highland populations. Increased oxygen affinity in hemoglobin has been observed in various organisms inhabiting high altitudes, such as bar-headed geese (Liang et al., 2001) and deer mice (Storz et al., 2009). In Tibetan chickens, HBAD has been associated with high-altitude adaptation (Zhong et al., 2022) and a specific amino acid change in this protein was proposed to enhance oxygen affinity (Gou et al., 2007). However, the position of this amino acid change is different from the one we identified. Experiments and protein structure prediction need to be conducted to further check whether the highland HBAD that we found in *R. d. tibetana* also shows higher oxygen affinity.

Unlike mammals, no maternal support of adequate amounts of oxygen is provided to the avian embryo during incubation. Given the demanding conditions of low oxygen experienced at high-altitudes, chronic hypoxia can have deleterious impacts on cardiovascular development in avian embryos (Salinas et al., 2010). Even though migratory birds only spend a few months on the plateau during the breeding season, high altitude adaptation might especially be important during ontogenesis. Experimental studies of embryonic development in mice indicated EPAS1 to be essential to prevent heart failure (Tian et al., 1998) and it plays an important role in remodelling the primary vascular network into a mature hierarchy pattern (Peng et al., 2000). Thus, we hypothesize that EPAS1 together with HBAD could provide important effects especially during embryotic development to cope with hypoxia in highland *R. d. tibetana* populations.

Adaptive introgression of low altitude variants in EPAS1

Interestingly, unlike in HBAD, all nucleotide haplotypes of EPAS1 from lowland *R. d. tibetan*a were shared with *R. d. diluta* (Figure 4), and the most frequent AA haplotype was shared between all lowland populations (Figure S1). Topology weighting analysis also revealed a constrasting pattern of the region containing EPAS1 from the rest of genome. The outlier window containing EPAS1 showed a dominant gene tree topology with lowland *R. d. tibetana* and *R. d. diluta* as sister lineages while topology of the other two outlier windows had the species tree as the dominant gene tree topology (Figure 5). Our results thus indicate that the

lowland EPAS1 haplotypes of *R. d. tibetana* likely introgressed from *R. d. diluta* and were positively selected for (Figure 1).

Contrasting to findings that adaptive introgression of EPAS1 from Denisovans might help Tibetans adapt to high altitudes (Huerta-Sánchez et al., 2014; Zhang et al., 2021), our results suggest that adaptive introgression of EPAS1 likely facilitated pale sand martins from high altitudes to colonize the lowland. We hypothesized that *R. d. tibetana* might have passed at least part of the late Pleistocene in a refugium to the east or northeast of the Qinghai-Tibetan plateau during the glacial expansion (Tang et al in revision). Strong flexibility has been shown in bird migration (Somveille et al., 2020) and migratory birds can change to different wintering or breeding grounds to establish a novel migration route within a short period of time (Irwin, 2009). Thus the colonisation of lowlands by *R. d. tibetana* might have occurred rather recent probably after Last Glacial Maximum. Limited genome-wide admixed ancestry of *R. d. diluta* was only found in lowland *R. d. tibetana* (Figure 1). Thus, we think adaptive introgression would be the more likely scenario than convergent selection of EPAS1 in the lowland populations.

The role of introgressive hybridization in evolution is controversial. Most hybrid genotypes are considered less fit (Barton & Hewitt, 1985; Burke & Arnold, 2001). However introgression of a few loci or establishment of a recombinant genotype through hybridization may also promote adaptation (Abbott et al., 2013; Seehausen, 2004). An increasing number of studies shows that standing genetic variation played a key role in adaptive radiations through introgression that facilitates species to occupy new niches (Marques et al., 2019; Meier et al., 2017; Rubin et al., 2022). *De novo* mutation rates are typically low and might not be efficient to cope with constantly changing environments. Recruitment or recombining of ancestral variation could be more efficient to induce adaptation in a short period of time (Meier et al., 2017; Rubin et al., 2022). Our study adds an example by indicating that adaptive introgression could facilitate elevation shifts and the colonization of new environment.

Conclusion

Our study sheds light on the genetic adaptation in a migratory bird species breeding at high elevations. Unlike resident species at high elevation that endure chronic hypobaric hypoxia throughout lifetimes, these birds only occur on high-altitudes during the breeding season, raising questions whether genetic adaptations are involved. We identified EPAS1 and HBAD associated high-altitude adaption in pale sand martin and we hypothesize they might play an important role in embryotic development to cope with hypoxia. Furthermore, our results indicated adaptive introgression of EPAS1 might have facilitated the colonization of the lowlands. Our findings suggested important role of introgressive hybridization in adaptation.

Acknowledgements

We thank Qin Huang, Xinyuan Pan, Dan Liang, Yun Li, 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 we are grateful to Susanne Tellenbach for the help with laboratory work. We acknowledge the Next Generation Sequencing (NGS) Platform at University of Bern and we especially thank Pamela Nicholson for her support. 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. We acknowledge the financial support of Oversea Study Programme of Guangzhou Elite Project (no. JY201726)

References

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J. E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C. A., Buggs, R., Butlin, R. K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S. H., Hermansen, J. S., Hewitt, G., Hudson, A. G., Jiggins, C., … Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*(2), 229–246. https://doi.org/10.1111/j.1420-9101.2012.02599.x

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Aljanabi, SM., & Martinez, I. (1997). Universal and rapid salt-extraction of high quality genomic DNA for PCR- based techniques. *Nucleic Acids Research*, *25*(22), 4692–4693. https://doi.org/10.1093/nar/25.22.4692

Appelhoff, R. J., Tian, Y.-M., Raval, R. R., Turley, H., Harris, A. L., Pugh, C. W., Ratcliffe, P. J., & Gleadle, J. M. (2004). Differential Function of the Prolyl Hydroxylases PHD1, PHD2, and PHD3 in the Regulation of Hypoxia-inducible Factor. *Journal of Biological Chemistry*, *279*(37), 38458–38465. https://doi.org/10.1074/jbc.M406026200

Auwera, G. van der, & O'Connor, B. D. (2020). *Genomics in the cloud: Using Docker, GATK, and WDL in Terra* (First edition). O'Reilly Media.

Bandelt, H. J., Forster, P., & Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, *16*(1), 37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036

Barton, N. H., & Hewitt, G. M. (1985). Analysis of Hybrid Zones. *Annual Review of Ecology and Systematics*, *16*(1), 113–148. https://doi.org/10.1146/annurev.es.16.110185.000553

Beall, C. M., Cavalleri, G. L., Deng, L., Elston, R. C., Gao, Y., Knight, J., Li, C., Li, J. C., Liang, Y., McCormack, M., Montgomery, H. E., Pan, H., Robbins, P. A., Shianna, K. V., Tam, S. C., Tsering, N., Veeramah, K. R., Wang, W., Wangdui, P., … Zheng, Y. T. (2010). Natural selection on *EPAS1* (*HIF2α*) associated with low hemoglobin concentration in Tibetan highlanders. *Proceedings of the National Academy of Sciences*, *107*(25), 11459–11464. https://doi.org/10.1073/pnas.1002443107

Bigham, A., Bauchet, M., Pinto, D., Mao, X., Akey, J. M., Mei, R., Scherer, S. W., Julian, C. G., Wilson, M. J., López Herráez, D., Brutsaert, T., Parra, E. J., Moore, L. G., & Shriver, M. D. (2010). Identifying Signatures of Natural Selection in Tibetan and Andean Populations Using Dense Genome Scan Data. *PLoS Genetics*, *6*(9), e1001116. https://doi.org/10.1371/journal.pgen.1001116

Bigham, A. W., Wilson, M. J., Julian, C. G., Kiyamu, M., Vargas, E., Leon-Velarde, F., Rivera-Chira, M., Rodriquez, C., Browne, V. A., Parra, E., Brutsaert, T. D., Moore, L. G., & Shriver, M. D. (2013). Andean and Tibetan patterns of adaptation to high altitude: Andean and Tibetan Altitude Adaptation. *American Journal of Human Biology*, *25*(2), 190–197. https://doi.org/10.1002/ajhb.22358

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Braunewell, K.-H., & Gundelfinger, E. D. (1999). Intracellular neuronal calcium sensor proteins: A family of EF-hand calcium-binding proteins in search of a function. *Cell and Tissue Research*, *295*(1), 1–12. https://doi.org/10.1007/s004410051207

Burke, J. M., & Arnold, M. L. (2001). Genetics and the Fitness of Hybrids. *Annual Review of Genetics*, *35*(1), 31–52. https://doi.org/10.1146/annurev.genet.35.102401.085719

Burri, R. (2017). Interpreting differentiation landscapes in the light of long-term linked selection: DIFFERENTIATION AND LONG-TERM LINKED SELECTION. *Evolution Letters*, *1*(3), 118–131. https://doi.org/10.1002/evl3.14

Campagna, L., & Toews, D. P. L. (2022). The genomics of adaptation in birds. *Current Biology*, *32*(20), R1173–R1186. https://doi.org/10.1016/j.cub.2022.07.076

Cheng, Y., Miller, M. J., Zhang, D., Xiong, Y., Hao, Y., Jia, C., Cai, T., Li, S.-H., Johansson, U. S., Liu, Y., Chang, Y., Song, G., Qu, Y., & Lei, F. (2021). Parallel genomic responses to

historical climate change and high elevation in East Asian songbirds. *Proceedings of the National Academy of Sciences*, *118*(50), e2023918118. https://doi.org/10.1073/pnas.2023918118

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008. https://doi.org/10.1093/gigascience/giab008

Delaneau, O., Zagury, J.-F., Robinson, M. R., Marchini, J. L., & Dermitzakis, E. T. (2019). Accurate, scalable and integrative haplotype estimation. *Nature Communications*, *10*(1), 5436. https://doi.org/10.1038/s41467-019-13225-y

Dixit, S., Joshi, V., & Barve, S. (2016). Bird diversity of the Amrutganga Valley, Kedarnath, Uttarakhand, India with an emphasis on the elevational distribution of species. *Check List*, *12*(2), 1874. https://doi.org/10.15560/12.2.1874

Dunn, P. (2004). Breeding Dates and Reproductive Performance. In *Advances in Ecological Research* (Vol. 35, pp. 69–87). Elsevier. https://doi.org/10.1016/S0065-2504(04)35004-X

Excoffier, L., Foll, M., & Petit, R. J. (2009). Genetic Consequences of Range Expansions. *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 481–501. https://doi.org/10.1146/annurev.ecolsys.39.110707.173414

Ferrer-Admetlla, A., Liang, M., Korneliussen, T., & Nielsen, R. (2014). On Detecting Incomplete Soft or Hard Selective Sweeps Using Haplotype Structure. *Molecular Biology and Evolution*, *31*(5), 1275–1291. https://doi.org/10.1093/molbev/msu077

Gou, X., Li, N., Lian, L., Yan, D., Zhang, H., Wei, Z., & Wu, C. (2007). Hypoxic adaptations of hemoglobin in Tibetan chick embryo: High oxygen-affinity mutation and selective expression. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *147*(2), 147–155. https://doi.org/10.1016/j.cbpb.2006.11.031

Gou, X., Wang, Z., Li, N., Qiu, F., Xu, Z., Yan, D., Yang, S., Jia, J., Kong, X., Wei, Z., Lu, S., Lian, L., Wu, C., Wang, X., Li, G., Ma, T., Jiang, Q., Zhao, X., Yang, J., … Li, Y. (2014). Whole-genome sequencing of six dog breeds from continuous altitudes reveals adaptation to high-altitude hypoxia. *Genome Research*, *24*(8), 1308–1315. https://doi.org/10.1101/gr.171876.113

Graham, A. M., & McCracken, K. G. (2019). Convergent evolution on the hypoxia-inducible factor (HIF) pathway genes EGLN1 and EPAS1 in high-altitude ducks. *Heredity*, *122*(6), 819–832. https://doi.org/10.1038/s41437-018-0173-z

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Systematic Biology*, *59*(3), 307–321. https://doi.org/10.1093/sysbio/syq010

Hao, Y., Xiong, Y., Cheng, Y., Song, G., Jia, C., Qu, Y., & Lei, F. (2019). Comparative transcriptomics of 3 high-altitude passerine birds and their low-altitude relatives. *Proceedings of the National Academy of Sciences*, *116*(24), 11851–11856. https://doi.org/10.1073/pnas.1819657116

Hawkes, L. A., Balachandran, S., Batbayar, N., Butler, P. J., Frappell, P. B., Milsom, W. K., Tseveenmyadag, N., Newman, S. H., Scott, G. R., Sathiyaselvam, P., Takekawa, J. Y., Wikelski, M., & Bishop, C. M. (2011). The trans-Himalayan flights of bar-headed geese (*Anser indicus*). *Proceedings of the National Academy of Sciences*, *108*(23), 9516–9519. https://doi.org/10.1073/pnas.1017295108

Huerta-Sánchez, E., Jin, X., Asan, Bianba, Z., Peter, B. M., Vinckenbosch, N., Liang, Y., Yi, X., He, M., Somel, M., Ni, P., Wang, B., Ou, X., Huasang, Luosang, J., Cuo, Z. X. P., Li, K., Gao, G., Yin, Y., … Nielsen, R. (2014). Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature*, *512*(7513), 194–197. https://doi.org/10.1038/nature13408

Hulsen, T., De Vlieg, J., & Alkema, W. (2008). BioVenn – a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics*, *9*(1), 488. https://doi.org/10.1186/1471-2164-9-488

Irwin, D. E. (2009). Speciation: New Migratory Direction Provides Route toward Divergence. *Current Biology*, *19*(24), R1111–R1113. https://doi.org/10.1016/j.cub.2009.11.011

Irwin, D. E., & Irwin, J. H. (2005). Siberian migratory divides: The role of seasonal migration in speciation. In *R. Greenberg & P. P. Marra (Eds.), Birds of two worlds: The ecology and evolution of migration.* (pp. 27–40). Johns Hopkins University Press.

Jaakkola, P., Mole, D. R., Tian, Y.-M., Wilson, M. I., Gielbert, J., Gaskell, S. J., Kriegsheim, A. V., Hebestreit, H. F., Mukherji, M., Schofield, C. J., Maxwell, P. H., Pugh, ‡ Christopher W., & Ratcliffe, ‡ Peter J. (2001). Targeting of HIF-α to the von Hippel-Lindau Ubiquitylation Complex by O_2 -Regulated Prolyl Hydroxylation. *Science*, *292*(5516), 468–472. https://doi.org/10.1126/science.1059796

Keilwagen, J., Hartung, F., Paulini, M., Twardziok, S. O., & Grau, J. (2018). Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinformatics*, *19*(1), 189. https://doi.org/10.1186/s12859-018-2203-5

Keilwagen, J., Wenk, M., Erickson, J. L., Schattat, M. H., Grau, J., & Hartung, F. (2016). Using intron position conservation for homology-based gene prediction. *Nucleic Acids Research*, *44*(9), e89–e89. https://doi.org/10.1093/nar/gkw092

Kowalczyk, M. S., Hughes, J. R., Babbs, C., Sanchez-Pulido, L., Szumska, D., Sharpe, J. A., Sloane-Stanley, J. A., Morriss-Kay, G. M., Smoot, L. B., Roberts, A. E., Watkins, H., Bhattacharya, S., Gibbons, R. J., Ponting, C. P., Wood, W. G., & Higgs, D. R. (2012). Nprl3 is required for normal development of the cardiovascular system. *Mammalian Genome*, *23*(7–8), 404–415. https://doi.org/10.1007/s00335-012-9398-y

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*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Lan, Y., Zhang, N., Liu, H., Xu, J., & Jiang, R. (2016). Golgb1 regulates protein glycosylation and is crucial for mammalian palate development. *Development*, dev.134577. https://doi.org/10.1242/dev.134577

Leigh, J. W., & Bryant, D. (2015). POPART: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, *6*(9), 1110–1116. https://doi.org/10.1111/2041-210X.12410

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. https://doi.org/10.1093/bioinformatics/bty191

Li, J.-T., Gao, Y.-D., Xie, L., Deng, C., Shi, P., Guan, M.-L., Huang, S., Ren, J.-L., Wu, D.-D., Ding, L., Huang, Z.-Y., Nie, H., Humphreys, D. P., Hillis, D. M., Wang, W.-Z., & Zhang, Y.-P. (2018). Comparative genomic investigation of high-elevation adaptation in ectothermic snakes. *Proceedings of the National Academy of Sciences*, *115*(33), 8406–8411. https://doi.org/10.1073/pnas.1805348115

Liang, Y., Hua, Z., Liang, X., Xu, Q., & Lu, G. (2001). The crystal structure of bar-headed goose hemoglobin in deoxy form: The allosteric mechanism of a hemoglobin species with high oxygen affinity 1 1Edited by A. Klug. *Journal of Molecular Biology*, *313*(1), 123–137. https://doi.org/10.1006/jmbi.2001.5028

Lutgen, D., Ritter, R., Olsen, R., Schielzeth, H., Gruselius, J., Ewels, P., García, J. T., Shirihai, H., Schweizer, M., Suh, A., & Burri, R. (2020). Linked-read sequencing enables haplotype-resolved resequencing at population scale. *Molecular Ecology Resources*, *20*(5), 1311–1322. https://doi.org/10.1111/1755-0998.13192

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., & Zdobnov, E. M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution*, *38*(10), 4647–4654. https://doi.org/10.1093/molbev/msab199

Marques, D. A., Meier, J. I., & Seehausen, O. (2019). A Combinatorial View on Speciation and Adaptive Radiation. *Trends in Ecology & Evolution*, *34*(6), 531–544. https://doi.org/10.1016/j.tree.2019.02.008

Martin, M., Patterson, M., Garg, S., O Fischer, S., Pisanti, N., Klau, G. W., Schöenhuth, A., & Marschall, T. (2016). *WhatsHap: Fast and accurate read-based phasing* [Preprint]. Bioinformatics. https://doi.org/10.1101/085050

Martin, S. H., & Van Belleghem, S. M. (2017). Exploring Evolutionary Relationships Across the Genome Using Topology Weighting. *Genetics*, *206*(1), 429–438. https://doi.org/10.1534/genetics.116.194720

Meier, J. I., Marques, D. A., Mwaiko, S., Wagner, C. E., Excoffier, L., & Seehausen, O. (2017). Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature Communications*, *8*(1), 14363. https://doi.org/10.1038/ncomms14363

Meisner, J., & Albrechtsen, A. (2018). Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. *Genetics*, *210*(2), 719–731. https://doi.org/10.1534/genetics.118.301336

Pamenter, M. E., Hall, J. E., Tanabe, Y., & Simonson, T. S. (2020). Cross-Species Insights Into Genomic Adaptations to Hypoxia. *Frontiers in Genetics*, *11*, 743. https://doi.org/10.3389/fgene.2020.00743

Peng, J., Zhang, L., Drysdale, L., & Fong, G.-H. (2000). The transcription factor EPAS-1/hypoxia-inducible factor 2α plays an important role in vascular remodeling. *Proceedings of the National Academy of Sciences*, *97*(15), 8386–8391. https://doi.org/10.1073/pnas.140087397

Peng, Y., Cui, C., He, Y., Ouzhuluobu, Zhang, H., Yang, D., Zhang, Q., Bianbazhuoma, Yang, L., He, Y., Xiang, K., Zhang, X., Bhandari, S., Shi, P., Yangla, Dejiquzong, Baimakangzhuo, Duojizhuoma, Pan, Y., … Su, B. (2017). Down-Regulation of *EPAS1* Transcription and

Genetic Adaptation of Tibetans to High-Altitude Hypoxia. *Molecular Biology and Evolution*, msw280. https://doi.org/10.1093/molbev/msw280

Percy, M. J., Beer, P. A., Campbell, G., Dekker, A. W., Green, A. R., Oscier, D., Rainey, M. G., Van Wijk, R., Wood, M., Lappin, T. R. J., McMullin, M. F., & Lee, F. S. (2008). Novel exon 12 mutations in the HIF2A gene associated with erythrocytosis. *Blood*, *111*(11), 5400–5402. https://doi.org/10.1182/blood-2008-02-137703

Petschow, D., Wurdinger, I., Baumann, R., Duhm, J., Braunitzer, G., & Bauer, C. (1977). Causes of high blood O2 affinity of animals living at high altitude. *Journal of Applied Physiology*, *42*(2), 139–143. https://doi.org/10.1152/jappl.1977.42.2.139

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

R Core Team. (2019). *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Https://www.R-project.org/* [Computer software].

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., & Vilo, J. (2019). g:Profiler: A web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Research*, *47*(W1), W191–W198. https://doi.org/10.1093/nar/gkz369

Roesti, M., Gavrilets, S., Hendry, A. P., Salzburger, W., & Berner, D. (2014). The genomic signature of parallel adaptation from shared genetic variation. *Molecular Ecology*, *23*(16), 3944–3956. https://doi.org/10.1111/mec.12720

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. https://doi.org/10.1093/molbev/msx248

Rubin, C.-J., Enbody, E. D., Dobreva, M. P., Abzhanov, A., Davis, B. W., Lamichhaney, S., Pettersson, M., Sendell-Price, A. T., Sprehn, C. G., Valle, C. A., Vasco, K., Wallerman, O., Grant, B. R., Grant, P. R., & Andersson, L. (2022). Rapid adaptive radiation of Darwin's finches depends on ancestral genetic modules. *Science Advances*, *8*(27), eabm5982. https://doi.org/10.1126/sciadv.abm5982

Salinas, C. E., Blanco, C. E., Villena, M., Camm, E. J., Tuckett, J. D., Weerakkody, R. A., Kane, A. D., Shelley, A. M., Wooding, F. B. P., Quy, M., & Giussani, D. A. (2010). Cardiac and vascular disease prior to hatching in chick embryos incubated at high altitude. *Journal of Developmental Origins of Health and Disease*, *1*(1), 60–66. https://doi.org/10.1017/S2040174409990043

Schweizer, R. M., Velotta, J. P., Ivy, C. M., Jones, M. R., Muir, S. M., Bradburd, G. S., Storz, J. F., Scott, G. R., & Cheviron, Z. A. (2019). Physiological and genomic evidence that selection on the transcription factor Epas1 has altered cardiovascular function in high-altitude deer mice. *PLOS Genetics*, *15*(11), e1008420. https://doi.org/10.1371/journal.pgen.1008420

Scordato, E. S. C., Smith, C. C. R., Semenov, G. A., Liu, Y., Wilkins, M. R., Liang, W., Rubtsov, A., Sundev, G., Koyama, K., Turbek, S. P., Wunder, M. B., Stricker, C. A., & Safran, R. J. (2020). Migratory divides coincide with reproductive barriers across replicated avian

hybrid zones above the Tibetan Plateau. *Ecology Letters*, *23*(2), 231–241. https://doi.org/10.1111/ele.13420

Scott, G. R., Egginton, S., Richards, J. G., & Milsom, W. K. (2009). Evolution of muscle phenotype for extreme high altitude flight in the bar-headed goose. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1673), 3645–3653. https://doi.org/10.1098/rspb.2009.0947

Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends in Ecology & Evolution*, *19*(4), 198–207. https://doi.org/10.1016/j.tree.2004.01.003

Smit, A. F. A., Hubley, R., & Green, P. (2013). *RepeatMasker (Open-4.0) [Computer software]. Http://www.repeatmasker.org*.

Somveille, M., Wikelski, M., Beyer, R. M., Rodrigues, A. S. L., Manica, A., & Jetz, W. (2020). Simulation-based reconstruction of global bird migration over the past 50,000 years. *Nature Communications*, *11*(1), 801. https://doi.org/10.1038/s41467-020-14589-2

Storz, J. F., Runck, A. M., Sabatino, S. J., Kelly, J. K., Ferrand, N., Moriyama, H., Weber, R. E., & Fago, A. (2009). Evolutionary and functional insights into the mechanism underlying high-altitude adaptation of deer mouse hemoglobin. *Proceedings of the National Academy of Sciences*, *106*(34), 14450–14455. https://doi.org/10.1073/pnas.0905224106

Storz, J. F., Sabatino, S. J., Hoffmann, F. G., Gering, E. J., Moriyama, H., Ferrand, N., Monteiro, B., & Nachman, M. W. (2007). The Molecular Basis of High-Altitude Adaptation in Deer Mice. *PLoS Genetics*, *3*(3), e45. https://doi.org/10.1371/journal.pgen.0030045

Szpiech, Z. A., & Hernandez, R. D. (2014). selscan: An Efficient Multithreaded Program to Perform EHH-Based Scans for Positive Selection. *Molecular Biology and Evolution*, *31*(10), 2824–2827. https://doi.org/10.1093/molbev/msu211

Szpiech, Z. A., Novak, T. E., Bailey, N. P., & Stevison, L. S. (2021). Application of a novel haplotype-based scan for local adaptation to study high-altitude adaptation in rhesus macaques. *Evolution Letters*, *5*(4), 408–421. https://doi.org/10.1002/evl3.232

Tang, Q., Burri, R., Liu, Y., Suh, A., Sundev, G., Heckel, G., & Schweizer, M. (2022). Seasonal migration patterns and the maintenance of evolutionary diversity in a cryptic bird radiation. *Molecular Ecology*, *31*(2), 632–645. https://doi.org/10.1111/mec.16241

Tang, Q., Alayee, N., Liu, Y., Suh, A., Sundev, G., Heckel, G., & Schweizer, M. (n.d.). *The Hourglass Runs at Different Speeds in Riparia Sand Martins: Phylogenomic History and Diversification of a Cryptic Radiation (in revision with Journal of Biogeography).*

Taylor, S. E., Bagnall, J., Mason, D., Levy, R., Fernig, D. G., & See, V. (2016). Differential sub-nuclear distribution of hypoxia-inducible factors (HIF)-1 and -2 alpha impacts on their stability and mobility. *Open Biology*, *6*(9), 160195. https://doi.org/10.1098/rsob.160195

Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, *22*(22), 4673–4680. https://doi.org/10.1093/nar/22.22.4673

Tian, H., Hammer, R. E., Matsumoto, A. M., Russell, D. W., & McKnight, S. L. (1998). The hypoxia-responsive transcription factor EPAS1 is essential for catecholamine homeostasis and protection against heart failure during embryonic development. *Genes & Development*, *12*(21), 3320–3324. https://doi.org/10.1101/gad.12.21.3320

Wang, G.-D., Fan, R.-X., Zhai, W., Liu, F., Wang, L., Zhong, L., Wu, H., Yang, H.-C., Wu, S.-F., Zhu, C.-L., Li, Y., Gao, Y., Ge, R.-L., Wu, C.-I., & Zhang, Y.-P. (2014). Genetic Convergence in the Adaptation of Dogs and Humans to the High-Altitude Environment of the Tibetan Plateau. *Genome Biology and Evolution*, *6*(8), 2122–2128. https://doi.org/10.1093/gbe/evu162

Wang, W., Wang, F., Hao, R., Wang, A., Sharshov, K., Druzyaka, A., Lancuo, Z., Shi, Y., & Feng, S. (2020). First de novo whole genome sequencing and assembly of the bar-headed goose. *PeerJ*, *8*, e8914. https://doi.org/10.7717/peerj.8914

Wang, Y., Chen, Y., Chen, J., Wang, L., Nie, L., Long, J., Chang, H., Wu, J., Huang, C., & Lei, M. (2019). The meiotic TERB1-TERB2-MAJIN complex tethers telomeres to the nuclear envelope. *Nature Communications*, *10*(1), 564. https://doi.org/10.1038/s41467-019-08437-1

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. https://doi.org/10.1101/gr.214874.116

Williamson, J. L., & Witt, C. C. (2021). Elevational niche-shift migration: Why the degree of elevational change matters for the ecology, evolution, and physiology of migratory birds. *Ornithology*, *138*(2), ukaa087. https://doi.org/10.1093/ornithology/ukaa087

Zhang, X., Witt, K. E., Bañuelos, M. M., Ko, A., Yuan, K., Xu, S., Nielsen, R., & Huerta-Sanchez, E. (2021a). The history and evolution of the Denisovan- *EPAS1* haplotype in Tibetans. *Proceedings of the National Academy of Sciences*, *118*(22), e2020803118. https://doi.org/10.1073/pnas.2020803118

Zhang, X., Witt, K. E., Bañuelos, M. M., Ko, A., Yuan, K., Xu, S., Nielsen, R., & Huerta-Sanchez, E. (2021b). The history and evolution of the Denisovan- *EPAS1* haplotype in Tibetans. *Proceedings of the National Academy of Sciences*, *118*(22), e2020803118. https://doi.org/10.1073/pnas.2020803118

Zhong, H., Kong, X., Zhang, Y., Su, Y., Zhang, B., Zhu, L., Chen, H., Gou, X., & Zhang, H. (2022). Microevolutionary mechanism of high-altitude adaptation in Tibetan chicken populations from an elevation gradient. *Evolutionary Applications*, *15*(12), 2100–2112. https://doi.org/10.1111/eva.13503

Supplements

Table S1. Sampling list. PopID corresponds to the numbers in the sampling map in Figure 1.

PopID	Species	Sample size	Elevation(m)	Locality	Coordinate	PopName
1	*R. d. fohkienensis*	10	80	Yongjia Academy, Yongjia County, Wenzhou, Zhejiang Province, PR China	28°16'9.66"N, 120°40'54.89"E	RdfWz
2	*R. d. fohkienensis*	10	405	Lanxi County, Yibin, Sichuan PR China	28°48'45.90"N, 104°57'1.65"E	RdfYb
3	*R. d. fohkienensis*	10	610	Meixian county, Shaanxi Province, PR China	34°14'N, 107°54'E	RdfMx
4	*R. d. diluta*	10	459	Wujiaqu, Changji, Xinjiang Uygur Autonomous Region, PR China	44°22'N, 87°53'E	RddWjq
5	*R. d. diluta*	10	999	Alar, Xinjiang Uygur Autonomous Region, PR China	40°35.433'N, 81°43.39'E	RddAla
6	*R. d. diluta*	10	1328	Khotan, Xinjiang Uygur Autonomous Region, PR China	37°11'39.51"N, 79°57'41.05''E	RddKho
7	*R. d. tibetana*	3	1211	Argalant, Mongolia	47°52.4902'N, 105°47.7577'E	RdtM
7	*R. d. tibetana*	7	1279	Elsen Tasarkhai, Mongolia	47°21.5352'N, 103°41.9692'E	RdtM
8	*R. d. tibetana*	5	3235	Qinghai Lake, Qinghai Province, PR China	36°34'15‘'N, 100°44'39‘'E	RdtQhl
8	*R. d. tibetana*	5	3517	Rubber Mountain, Qinghai Province, PR China	36°45'55''N, 99°38'20''E	RdtRm
9	*R. d. tibetana*	5	3420	Zoige, Sichuan Province, PR China	33°27'12”N, 102°28'55”E	RdtZg
9	*R. d. tibetana*	5	3561	Gahai Lake, Gansu Province, PR China	34°14'50‘'N, 102°20'32‘'E	RdtGh
10	*R. d. tibetana*	10	3599	Qushui County, Lhasa, Tibet autonomous region, PR China	29°20'13"N, 90°52'36''E	RdtLs

Table S2. Lists of annotated genes found in the outlier windows for three analysis based on Delta divergence and XP-nSL score.

xpnsl_d	xpnsl_t	DeltaFst
ABHD18	ABHD10	FNDC4
ACOT6	AGAP1	LOC112530482
AGAP1	AKAP9	ALOX15B
AKAP9	ALDH18A1	BF2
ALDH1L2	ANGPTL2	CHCHD7
ANKRD34C	ANXA8L1	CNTN5
ANPEP	ARFIP2	CRACD
AP3S2	B4GALT5	CUEDC2
AQP11	BCL9L	CYB5R2
ATP1A1	BFAR	DOCK4
BCL9L	BLACAT1	DUSP6
BICD2	C1orf210	EPAS1
BLOC1S1	C8orf59	FAM19A1
BMP8A	CA13	FBXO15
C12orf45	CA4	FGD5
C2CD2	CAND2	GABRA4
CA4	CCKBR	GABRB2
CAND2	CEP70	GIMD1
CAPZA1	CHEK2	GOLGA4
CD1C	CHKA	GOLGB1
CD44	CLDN23	HBAD
CERS1	CLDN25	HBZ
CHD7	COX10	HNF4G
CHID1	CPNE3	HUS1
CLK3	CPT1A	IGSF10
CTSG	CYP51A1	ITGA9
CYP51A1	DDX41	ITIH5
CYTH1	DOCK9	LOC101748344
DNAH17	DPH5	LOC101748553
DNAJC14	DUSP14	LOC101751319
DPP9	DYNC1I2	LOC107049274
DST	E2F5	LOC107049393
DTX2	EPAS1	LOC107052718
DUOXA2	EVI2A	LOC107053122
EBF1	EVI2B	LOC112530482
EDC3	FAIM	LOC112530492
EFCAB5	FBXW5	LOC121107939
EHF	FHIP1B	LOC121108002

EPAS1	FLNB	LOC121108005
FAM126A	FLT1	LOC121108045
FBXW2	GABRP	LOC121108225
FGD5	GAL	LOC121110182
GARNL3	GOLGB1	LOC124417063
GBE1	GP1BB	LOC124417490
GID8	GPR182	LOC124418371
GMPS	GPRIN2	LOC421982
GOLGB1	H1FOO	LOC426526
GOSR2	HBAD	LOC430443
H1FOO	HBZ	LOC770705
HAUS2	HCLS1	LOC776232
HBAD	HNF4G	LRRC14B
HBZ	HSPA9	LRRFIP2
HCLS1	IFITM10	MAJIN
HOXC10	INHBA	ME1
INPP1	KCNIP1	MRPS25
KANSL3	KCNMB1	NEUROD6
LIMK1	KDM6A	NFX1
LIN7A	KMT2A	NPRL3
LOC101748913	LOC101748913	NR2C2
LOC101751162	LOC101751878	NUDT13
LOC101751605	LOC107053928	OIT3
LOC107052718	LOC124416959	OR14J1L67
LOC107053714	LOC419074	OR14J1L81
LOC112530482	LOC430303	OVOB
LOC112530492	LOC430443	P4HA1
LOC121108216	LOC769726	PAK3
LOC121108225	LRRCC1	PGLYRP2
LOC121111295	LRRN1	PIM
LOC121111296	MADD	PLA2G12B
LOC124416959	MAJIN	PPP1R12A
LOC124417490	MFHAS1	PTPN5
LOC124417894	MIF	RAMP3
LOC428872	MLPH	RBSN
LOC430443	MMP17	ROS1
LOC771876	MRPS14	RPL39L
MAJIN	MRRFP1	SDR16C5
MAPK8IP3	MSH4	SH3D19
7-Mar	MYO1A	SH3YL1
MATN2	NAB2	TNS1
MESP1	NCOA4	TRHDE

MFSD8
MLXIPL
MORC1
MOV10
MRC2
MRPS25
MRPS14
MXD1
NDUFS5
NPRL3
NR2C2
NRBP1
OR5AS1
OR9Q1
PDPR
PHC2
PIK3R3
PLCH2
PLIN4
PLXND1
RACGAP1
RALGPS1
RBSN
RDH5
RHO
RHOC
RIOX1
RPL30
RPL32
S100A10
SCLT1
SEC11A
SFSWAP
SLC17A6
SLC17A9
ST3GAL2
ST7
SUCNR1
TANC2
TBL2
TFAP2E
TGFA
NEMP1
NPRL3
NPY4R
NSD1
OMG
OPN1MSW
OR9Q1
PARN
PARP9
PDCL
PHKA1
PHLDB2
PIGF
PLA2G5
PLSCR5
PLXND1
PPP1R3B
PRAG1
PRPH2L
PSPC1
PTGDS
PTGES1
PTGS1
RAB24
RABGAP1L
RABL2A
RC3H2
RCN1
REN
RHO
RHOQ
RNF166
RPL32
SEPTIN5
SFSWAP
SH3D19
SLC25A12
SLC2A11L4
SLC30A7
SLC35F5
SLC45A1
SLC6A1
ZNF512
TMTC3
TNS1
TRH
TRIP13
TRPS1
VDAC3
ZNF512

TIA1	SMARCAL1
TIMP2	SPATA16
TMCC1	SPATA18
TNFAIP8L1	SULT6B1L
TRH	TAC3
TSSK3	TADA2A
UGT1A1	TCTN3
USP10	THSD4
USP34	TIE1
USP36	TMCC1
USP39	TMEM125
USP6	TNFSF4
ZNF362	TRAF2
ZNF592	TRH
	UBE4A
	UBXN2B
	USP6
	WWP1
	ZBTB34
	ZBTB43
	ZNF692

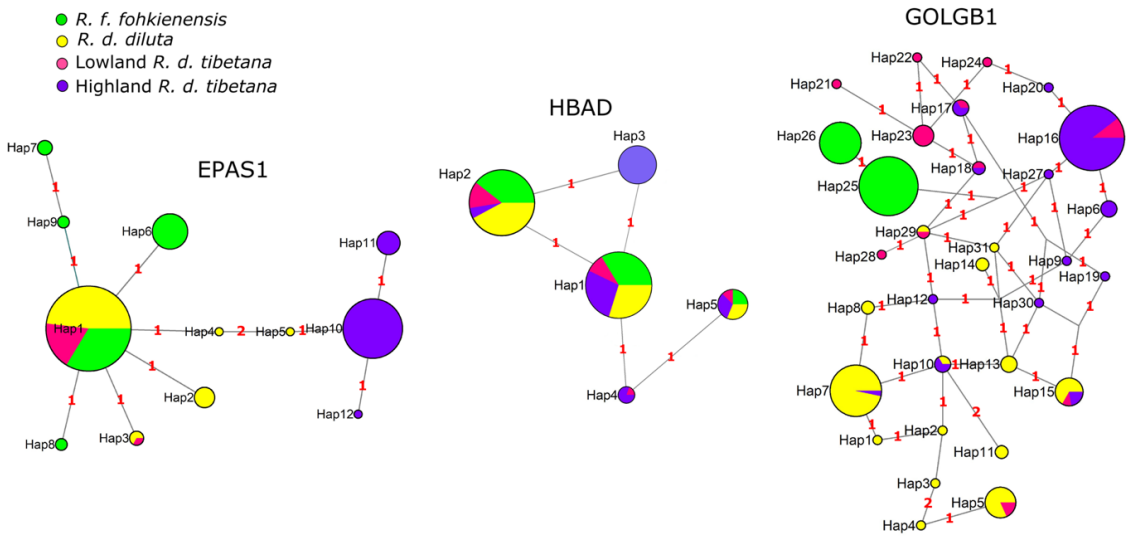


Figure S1. Haplotype network of amino acid sequences for the CDS region of EPAS1, HBAD and GOLGB1 within the outlier window. Numbers indicates the number of amino acid changes.

General discussion

General discussion

By utilising whole genomic data and comprehensive geographical sampling for pale sand martin, I applied a comparative approach including its sister species collared sand martin to study their cryptic diversification and demographic history. I revealed contrasting patterns of evolutionary diversity between the two sister species: shallow genomic divergence over a large geographic are with mitochondrial-nuclear discordance in Holarctic collared sand martin versus multiple morphologically cryptic evolutionary lineages in pale sand martin with relatively restricted ranges in Asia. Furthermore, I identified potential genes associated with high-altitude adaptation in pale sand martin.

Divergence with mitochondrial-nuclear discordance in Eurasian collared sand martin

Contrasting to the cryptic radiation found in pale sand martin, comparatively minor genomic divergence was found in collared sand martin between birds from Europe and Central Asia despite extensive environmental variation across the range. Three different evolutionary lineages were revealed in the phylogenomic framework of the species tree consistent with subspecies designation, although separated by mostly very short branches (**Chapter 2**). Time calibration based on genomic divergence indicated a long-lasting diversification process within the *R. riparia* during the last million years. Furthermore, except for *R. r. shelleyi* from Egypt and possibly the Levant (Shirihai and Svensson 2008) characterized by a distinct demographic history, the remaining *R. riparia* genomes suggested continuous population expansion approximately from the beginning of the LGP onwards. My data indicated that the large distribution of *R. riparia* of Eurasian part is not due to the massive range expansion after the LGP as previously thought based on interpretation of mtDNA variation (Pavlova et al. 2008; Schweizer et al. 2018), but rather of a long demographic expansion. *R. riparia* is the taxon with the largest populations sizes during LGP and the longest migration distances, and thus it might have high tolerance for heterogenous climate conditions and a high colonization potential (cf. Thorup et al. 2021). Intriguingly, coalescent times for the taxon inferred from multi mtDNA haplotypes are only around 0.1 Ma. Similar mito-nuclear discordances have been found in other Holarctic bird groups and were explained

by random genetic drift or selection, or a combination thereof (Irwin et al. 2009; Taylor et al. 2021). Based on the mtDNA analyses of (Pavlova et al. 2008), European and Siberian populations showed signs of recent demographic expansion whereas patterns in birds from the Far East and North America suggested more stable populations. Introgression of mtDNA might thus have happened e.g. from Beringia to the west. Whereas, strong selection on mitochondrial genome such as a consequence of thermal adaptations (e.g. Lamb et al. 2018; Melo-Ferreira et al. 2005) might also occur in collared sand martin. More biogeographic samplings including more northern and north-eastern populations in Asia and especially from North America are required to further test these mechanisms underlying the mito-nuclear discordances.

Simulations have suggested that migratory birds in North America and the western Palearctic – but less so in Asia – have shifted their breeding range towards LGM progressively in the direction of the equator (Somveille et al., 2020) and probably repeatedly through the late Pleistocene. I hypothesized that the collared sand martin might have repeatedly shifted its range as a consequence of climate and associated environmental alterations and pronounced dispersal propensity might have allowed the colonization of a large range and hindered their lineage diversification over a large geographic area.

Cryptic radiation in pale sand martin and regionally past climate impacts

First in **Chapter 1**, I revised the breeding distribution of subspecies *R. d. tibetana* which was thought to be restricted to Qinghai-Tibetan plateau (e.g. del Hoyo & Collar, 2016). Our genome-wide data showed its disjunct breeding range extends into central Mongolia where *R. d. diluta* was formerly believed to occur (e.g. del Hoyo & Collar, 2016). However, it remains to be examined how farther *R. d. tibetana* extends to the north and west from central Mongolia. In **Chapter 2**, two deep phylogenetic clades were revealed by strong genomic divergence within pale sand martin consisting of two evolutionary lineages at species level divergence (Tang et al., 2022; Tang et al., in revision). Each lineage includes two phylogeographic units broadly reflecting the distribution area of previously described subspecies (Schweizer et al., 2018; Tang et al., 2022). One clade comprising the *R. d.*

fohkienensis and *R. d. indica* represents a unique phylogeographic link in birds between subtropical China and the dry northern Indian Subcontinent while more northerly distributed *R. d. tibetana* from Qinghai-Tibetan plateau and central Mongolia and *R. d. diluta* in Central Asia clustered as sister group in the other clade.

In **Chapter 2**, distinct demography histories were also revealed within pale sand martin under the impacts of large regional differences in past Pleistocene climate change in east Asia. Consistent with other bird species in eastern Asia (Song et al., 2016), south eastern Chinese *R. f. fohkienensis* showed a stable effective population size during the Last Glacial Period (LGP) and experienced expansion towards the Last Glacial Maximum (LGM). Meanwhile, *Riparia d. diluta* underwent a population expansion at the beginning of the LGP followed by a trend of decline towards the LGM. In contrast, *R. d. tibetana* populations contracted approximately from the beginning of the LGP onwards after a steep increase in N_e before. This could be due to different impacts of Pleistocene glacial cycling on high altitude and lowland. During the late Pleistocene, the current breeding area of *R. d. tibetana* on the Qinghai-Tibetan plateau was largely covered by glaciers (Cui et al., 1998; Zhou et al., 2006), we hypothesized that *R. d. tibetana* may have stayed in a refugium in areas to the east or northeast of the Qinghai-Tibetan plateau. On the other hand, *R. d. diluta* likely expanded its range during the Last Glacial Period due to the extension of open arid habitats, a phenomenon observed in arid-adapted species in various geographic region (Alaei Kakhki et al., 2018; Kearns et al., 2014). The decline in Ne (effective population size) in *R. d. diluta* could be attributed to increased aridity, which might have become too extreme even for species adapted to arid conditions. *R. d. indica* breeds in relatively arid regions similar to *R. d. diluta*, and accordingly shows an analogous demographic history. It may have initially profited from an increase in arid open habitats in the north-western part of the Indian Subcontinent (Kar & Kumar, 2020) but contracted its range when aridity increased.

Mechanisms maintaining the cryptic diversity

I used genome-wide data to check whether gene flow is restricted between parapatric populations of different evolutionary lineages within pale sand martin. No gene flow was

detected between parapatric population of *R. d. tibetana* and *R. f. fohkienensis* near the eastern edge of Qinghai-Tibetan plateau. Furthermore, phenology data showed large differentiation in breeding timing between them indicating allochrony might act as prezygotic isolation to prevent the gene flow.

More interestingly, I found limited mixed ancestries of lowland population of *R. d. tibetana* in central Mongolia with *R. d. diluta* using both GBS (genotyping by sequencing) (**Chapter 1**) and resequencing data (**Chapter 3**). This area is located in a region of a well-known migratory divide for different bird species complexes that populations from the east of the region take a different migratory route than the populations from the west of the area (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 and Mongolia species use just one migratory route – east or west – around it or show different routes in different subspecies (Irwin & Irwin, 2005; Scordato et al., 2020). Limited hybridization was also found between two subspecies of barn swallow *Hirundo rustica* in the same region which show contrasting migration routes around the Qinghai-Tibetan plateau (Scordato et al., 2020). *R. d. diluta* was recorded to winter in the northwestern Indian Subcontinent and occasionally on the Arabian Peninsula, likely following a western migration route around the Qinghai-Tibetan plateau (Rasmussen et al., 2005; Shirihai & Svensson, 2018). In contrast, *R. d. tibetana* is thought to winter in Southeast Asia, potentially circling the Qinghai-Tibetan plateau through the east or using river valleys along its southeastern edge to winter in the northern, central, and northeastern Indian Subcontinent (own data; Rasmussen et al., 2005). We hypothesized that hybrids might have non-optimal intermediate migration routes towards Qinghai-Tibetan Plateau and selection against them might restrict gene flow. Our results indicate that seasonal migration behaviour might have played an important role in maintaining evolutionary diversity in birds under morphological stasis. However, more sampling would be needed to check whether there is a secondary contact zone between the *R. d. tibetana* and *R. d. diluta* in the western Mongolia. Furthermore, the migration behaviour also remains to be studied between populations in pale sand martin, especially in the region of the potential migratory divide. As birds molt, new feathers are formed, preserving a record of the chemical composition of their diet which reflect local environment during feather growth. Different

migration routes were revealed by different chemical composition of feathers grown in different breeding range or wintering ground based on stable isotope and trace element within various bird species (Chamberlain et al., 1997; Scordato et al., 2020; Szép et al., 2003). However, different feathers would moult at different time and it varies a lot between species (Jenni & Winkler, 2020). Moult was less studied in pale sand martin, thus makes it difficult to choose which feather to be applied stable isotope or trace element analyses (Szép et al., 2003). Thus, geolocator might be more applicable to apply to study their migration route.

High-altitude adaptation and adaptive introgression of EPAS1

In Chapter 3, I studied the genomic architecture of sand martins to check for signs indicating adaptation to. In order to find the genomic regions that might associated to high-altitude adaptation, I applied Fst-based estimation accounting for the effects of background and linked selection and haplotype-based analysis between different pairs of lowland and highland populations from three lineages in east Asia based on resequencing data. I identified genes including EPAS1 and HBAD potentially associated with high-altitude adaptation within *R. d. tibetana*. The Genomic basis of high-attitude adaptation was extensively studied in different organisms including humans but mainly focused on resident species (Beall et al., 2010; Bigham et al., 2010; Cheng et al., 2021; Graham & McCracken, 2019; Li et al., 2018; Storz et al., 2007; Wang et al., 2014). Less is known about the genetic basis of how migratory birds adapt to the high altitudes with large elevation shift except studies in the bar headed goose showing adaptations to improve oxygen transport efficiency (Petschow et al., 1977; Scott et al., 2009; Wang et al., 2020) or whether there is genetic adaptation involved since migratory birds only spend few months on the high-altitudes during breeding season. Our results indicated selective sweep on both HBAD and EPAS1 on the Qinghai-Tibetan Plateau in *R. d. tibetana*. I identified one unique amino acid change in HBAD (hemoglobin alpha, subunit D) that only occurred on highland populations. Increased oxygen affinity in hemoglobin has been observed in various organisms inhabiting high altitudes, such as bar-headed geese (Liang et al., 2001), deer mice (Storz et al., 2009). In Tibetan chickens, HBAD has been associated with high-altitude adaptation (Zhong et al., 2022) and we hypothesized the unique amino acid change we identified in highland *R. d. tibetana* might also enhance oxygen affinity (Gou

et al., 2007) However experiments and protein structure prediction need to be conducted to further check whether the highland HBAD that we found in *R. d. tibetana* also shows high oxygen affinity and the variant is indeed adaptive.

Two clusters were found for EPAS1 in the nucleotide haplotype network reconstruction (Figure 4): one only consisted of haplotypes found in highland *R. d. tibetana* populations (except one haplotype from *R. d. diluta*) with lowest haplotype diversity and the second cluster included all the haplotypes from lowland populations including *R. d. fohkienensis*, *R. d. diluta* and Mongolian *R. d. tibetana*. There are two dominant amino acid haplotypes found in lowland and highland population separately. EPAS1 has been found repeatedly to be a target for selection to cope with hypoxia on the high-altitudes across a variety of species (Beall et al., 2010; Graham & McCracken, 2019; Storz et al., 2007). Two out of four novel amino acid changes we identified for highland EPAS1 in *R. d. tibetana* were in exon 12. This region encodes ODD (O_2-dependent degradation)/NTAD (N-terminal transactivation) domain of EPAS1 protein which is crucial for the function that ODD is the target of oxygen-dependent degradation and NTAD is thought to confer target specificity (Jaakkola et al., 2001). Convergent evolution of EPAS1 for high altitude adaptation were found between two resident bird species Andean speckled teal and yellow-billed pintail and most nonsynonymous AA changes in EPAS1 also occurred in exon 12 (Graham & McCracken, 2019). Experimental studies in mice embryonic development showed EPAS1 to be essential for catecholamine homeostasis to prevent heart failure (Tian et al., 1998) as well as playing an important role in remodeling of the primary vascular network into a mature hierarchy pattern (Peng et al., 2000). Given the demanding conditions of low oxygen experienced at high-altitudes, chronic hypoxia can have deleterious impacts on cardiovascular development in avian embryos (Salinas et al., 2010). We hypothesised that EPAS1 together with HBAD would provide high altitude adaptation especially during embryotic development to cope with hypoxia in highland *R. d. tibetana* populations. Experiment and protein structure prediction need to be conducted to further check whether the highland EPAS1 and HBAD we found in *R. d. tibetana* would facilitate embryotic development under chronic hypoxia condition.

Furthermore, my results also indicated adaptive introgression of EPAS1 from *R. d. diluta* into lowland *R. d. tibetana* in central Mongolia. Topology weighting analysis also showed that the outlier window containing EPAS1 with dominant gene tree topology clustering lowland *R. d. tibetana* and *R. d. diluta* as sister group while the other two outlier window with species tree topology as the dominant gene tree topology. Limited genome-wide admixture of *R. d. diluta* only occurred in parapatric population of lowland *R. d. tibetana* and I hypothesized that lowland *R. d. tibetana* might recently expanded from the Qinghai-Tibetan Plateau during or after LGM and there might be secondary contact between *R. d. diluta* and *R. d. tibetana* in the lowland. The findings in Tibetan people showed that the high-altitude EPAS1 haplotypes were introgressed from Denisovans (Huerta-Sánchez et al., 2014; Zhang et al., 2021). However, our results indicated adaptive introgression of EPAS1 happened in lowland: lowland EPAS1 haplotypes from *R. d. diluta* were likely introgressed into *R. d. tibetana* during the secondary contact when high-altitude population expanded into lowland facilitating the high-altitude adaptive *R. d. tibetana* to colonize into the lowland. Considering the genome-wide divergence between *R. d. diluta* and lowland *R. d. tibetana*, other mechanisms especially migratory divide might play an important role to restrict the gene flow between them that only allow introgression happened with few adaptive loci, pale sand martin thus would provide a potential system to study the genetic basis underling the migration orientation which has been a big puzzle in bird evolution (Delmore et al., 2020; Delmore et al., 2016; Justen & Delmore, 2022; Sokolovskis et al., 2022).

In this thesis, I conducted a comprehensive analysis of genome-wide data, revealed cryptic radiation within a complex of migratory bird species and unravelled evidence of genomic adaptations in these migratory birds to the high altitudes. My findings suggested the presence of intriguing mechanisms that limit gene flow within these cryptic diversifications among migratory bird species. We hypothesize that pronounced dispersal propensity in the strongly migratory nominate subspecies of collared sand martin has hindered lineage divergence across its vast distribution range. In contrast, differential seasonal migratory behaviours appear to play a pivotal role in maintaining evolutionary diversity within the pale sand martin species complex under morphological stasis. Furthermore, my work provides

another example of important role of adaptive introgression to facilitate migratory birds in colonizing new environments successfully.

Reference

Alaei Kakhki, N., Aliabadian, M., Förschler, M. I., Ghasempouri, S. M., Kiabi, B. H., Verde Arregoitia, L. D., & Schweizer, M. (2018). Phylogeography of the *Oenanthe hispanica-pleschanka-cypriaca* complex (Aves, Muscicapidae: Saxicolinae): Diversification history of open-habitat specialists based on climate niche models, genetic data, and morphometric data. *Journal of Zoological Systematics and Evolutionary Research*, *56*(3), 408–427. https://doi.org/10.1111/jzs.12206

Beall, C. M., Cavalleri, G. L., Deng, L., Elston, R. C., Gao, Y., Knight, J., Li, C., Li, J. C., Liang, Y., McCormack, M., Montgomery, H. E., Pan, H., Robbins, P. A., Shianna, K. V., Tam, S. C., Tsering, N., Veeramah, K. R., Wang, W., Wangdui, P., … Zheng, Y. T. (2010). Natural selection on *EPAS1* (*HIF2α*) associated with low hemoglobin concentration in Tibetan highlanders. *Proceedings of the National Academy of Sciences*, *107*(25), 11459–11464. https://doi.org/10.1073/pnas.1002443107

Bigham, A., Bauchet, M., Pinto, D., Mao, X., Akey, J. M., Mei, R., Scherer, S. W., Julian, C. G., Wilson, M. J., López Herráez, D., Brutsaert, T., Parra, E. J., Moore, L. G., & Shriver, M. D. (2010). Identifying Signatures of Natural Selection in Tibetan and Andean Populations Using Dense Genome Scan Data. *PLoS Genetics*, *6*(9), e1001116. https://doi.org/10.1371/journal.pgen.1001116

Chamberlain, C. P., Blum, J. D., Holmes, R. T., Feng, X., Sherry, T. W., & Graves, G. R. (1997). The use of isotope tracers for identifying populations of migratory birds. *Oecologia*, *109*(1), 132–141. https://doi.org/10.1007/s004420050067

Cheng, Y., Miller, M. J., Zhang, D., Xiong, Y., Hao, Y., Jia, C., Cai, T., Li, S.-H., Johansson, U. S., Liu, Y., Chang, Y., Song, G., Qu, Y., & Lei, F. (2021). Parallel genomic responses to historical climate change and high elevation in East Asian songbirds. *Proceedings of the National Academy of Sciences*, *118*(50), e2023918118. https://doi.org/10.1073/pnas.2023918118

Cui, Z., Wu, Y., Liu, G., Ge, D., Pang, Q., & Xu, Q. (1998). On Kunlun-Yellow River tectonic movement. *Sci. China Ser. Earth Sci. 41:592–600.*

del Hoyo, J., & Collar, N. J. (2016). *HBW and BirdLife International illustrated checklist of the birds of the world: Vol. 2: Passerines.* Lynx Edicions.

Delmore, K. E., Toews, D. P. L., Germain, R. R., Owens, G. L., & Irwin, D. E. (2016). The Genetics of Seasonal Migration and Plumage Color. *Current Biology*, *26*(16), 2167–2173. https://doi.org/10.1016/j.cub.2016.06.015

Delmore, K., Illera, J. C., Pérez-Tris, J., Segelbacher, G., Lugo Ramos, J. S., Durieux, G., Ishigohoka, J., & Liedvogel, M. (2020). The evolutionary history and genomics of European blackcap migration. *eLife*, *9*, e54462. https://doi.org/10.7554/eLife.54462

Gou, X., Li, N., Lian, L., Yan, D., Zhang, H., Wei, Z., & Wu, C. (2007). Hypoxic adaptations of hemoglobin in Tibetan chick embryo: High oxygen-affinity mutation and selective expression. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *147*(2), 147–155. https://doi.org/10.1016/j.cbpb.2006.11.031

Graham, A. M., & McCracken, K. G. (2019). Convergent evolution on the hypoxia-inducible factor (HIF) pathway genes EGLN1 and EPAS1 in high-altitude ducks. *Heredity*, *122*(6), 819–832. https://doi.org/10.1038/s41437-018-0173-z

Huerta-Sánchez, E., Jin, X., Asan, Bianba, Z., Peter, B. M., Vinckenbosch, N., Liang, Y., Yi, X., He, M., Somel, M., Ni, P., Wang, B., Ou, X., Huasang, Luosang, J., Cuo, Z. X. P., Li, K., Gao, G., Yin, Y., … Nielsen, R. (2014). Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature*, *512*(7513), 194–197. https://doi.org/10.1038/nature13408

Irwin, D. E., & Irwin, J. H. (2005). Siberian migratory divides: The role of seasonal migration in speciation. In *In R. Greenberg & P. P. Marra (Eds.), Birds of two worlds: The ecology and evolution of migration* (pp. 27–40). Johns Hopkins University Press.

Irwin, D. E., Rubtsov, A. S., & Panov, E. N. (2009). Mitochondrial introgression and replacement between yellowhammers (Emberiza citrinella) and pine buntings (Emberiza leucocephalos) (Aves: Passeriformes). *Biological Journal Of The Linnean Society*, *98*(2), 422–438.

Jaakkola, P., Mole, D. R., Tian, Y.-M., Wilson, M. I., Gielbert, J., Gaskell, S. J., Kriegsheim, A. V., Hebestreit, H. F., Mukherji, M., Schofield, C. J., Maxwell, P. H., Pugh, ‡ Christopher W., & Ratcliffe, ‡ Peter J. (2001). Targeting of HIF-α to the von Hippel-Lindau Ubiquitylation Complex by O $_2$ -Regulated Prolyl Hydroxylation. *Science*, *292*(5516), 468–472. https://doi.org/10.1126/science.1059796

Jenni, L., & Winkler, R. (2020). *Moult and ageing of European passerines* (Second edition). Helm.

Justen, H., & Delmore, K. E. (2022). The genetics of bird migration. *Current Biology*, *32*(20), R1144–R1149. https://doi.org/10.1016/j.cub.2022.07.008

Kar, A., & Kumar, A. (2020). Evolution of arid landscape in India and likely impact of future climate change. *Episodes*, *43*(1), 511–523. https://doi.org/10.18814/epiiugs/2020/020033

Kearns, A. M., Joseph, L., Toon, A., & Cook, L. G. (2014). Australia's arid-adapted butcherbirds experienced range expansions during Pleistocene glacial maxima. *Nature Communications*, *5*(1), 3994. https://doi.org/10.1038/ncomms4994

Lamb, A. M., Gan, H. M., Greening, C., Joseph, L., Lee, Y. P., Morán-Ordóñez, A., Sunnucks, P., & Pavlova, A. (2018). Climate-driven mitochondrial selection: A test in Australian songbirds. *Molecular Ecology*, *27*(4), 898–918. https://doi.org/10.1111/mec.14488

Li, J.-T., Gao, Y.-D., Xie, L., Deng, C., Shi, P., Guan, M.-L., Huang, S., Ren, J.-L., Wu, D.-D., Ding, L., Huang, Z.-Y., Nie, H., Humphreys, D. P., Hillis, D. M., Wang, W.-Z., & Zhang, Y.-P. (2018). Comparative genomic investigation of high-elevation adaptation in ectothermic snakes. *Proceedings of the National Academy of Sciences*, *115*(33), 8406–8411. https://doi.org/10.1073/pnas.1805348115

Liang, Y., Hua, Z., Liang, X., Xu, Q., & Lu, G. (2001). The crystal structure of bar-headed goose hemoglobin in deoxy form: The allosteric mechanism of a hemoglobin species with high oxygen affinity 1 1Edited by A. Klug. *Journal of Molecular Biology*, *313*(1), 123–137. https://doi.org/10.1006/jmbi.2001.5028

Melo-Ferreira, J., Boursot, P., Suchentrunk, F., Ferrand, N., & Alves, P. C. (2005). Invasion from the cold past: Extensive introgression of mountain hare (Lepus timidus) mitochondrial DNA into three other hare species in northern Iberia. *Molecular Ecology*, *14*(8), 2459–2464. https://doi.org/10.1111/j.1365-294X.2005.02599.x

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. https://doi.org/10.1016/j.ympev.2008.03.030

Peng, J., Zhang, L., Drysdale, L., & Fong, G.-H. (2000). The transcription factor EPAS-1/hypoxia-inducible factor 2α plays an important role in vascular remodeling. *Proceedings of the National Academy of Sciences*, *97*(15), 8386–8391. https://doi.org/10.1073/pnas.140087397

Petschow, D., Wurdinger, I., Baumann, R., Duhm, J., Braunitzer, G., & Bauer, C. (1977). Causes of high blood O2 affinity of animals living at high altitude. *Journal of Applied Physiology*, *42*(2), 139–143. https://doi.org/10.1152/jappl.1977.42.2.139

Rasmussen, P. C., Anderton, J. C., & Edicions, L. (2005). *Birds of south Asia: The Ripley guide.* Barcelona: Lynx Edicions.

Salinas, C. E., Blanco, C. E., Villena, M., Camm, E. J., Tuckett, J. D., Weerakkody, R. A., Kane, A. D., Shelley, A. M., Wooding, F. B. P., Quy, M., & Giussani, D. A. (2010). Cardiac and vascular disease prior to hatching in chick embryos incubated at high altitude. *Journal of Developmental Origins of Health and Disease*, *1*(1), 60–66. https://doi.org/10.1017/S2040174409990043

Schweizer, M., Liu, Y., Olsson, U., Shirihai, H., Huang, Q., Leader, P. J., Copete, J. L., Kirwan, G. M., Chen, G. L., & Svensson, L. (2018a). 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, M., Liu, Y., Olsson, U., Shirihai, H., Huang, Q., Leader, P. J., Copete, J. L., Kirwan, G. M., Chen, G., & Svensson, L. (2018b). 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

Scordato, E. S. C., Smith, C. C. R., Semenov, G. A., Liu, Y., Wilkins, M. R., Liang, W., Rubtsov, A., Sundev, G., Koyama, K., Turbek, S. P., Wunder, M. B., Stricker, C. A., & Safran, R. J. (2020a). Migratory divides coincide with reproductive barriers across replicated avian hybrid zones above the Tibetan Plateau. *Ecology Letters*, *23*(2), 231–241. https://doi.org/10.1111/ele.13420

Scordato, E. S. C., Smith, C. C. R., Semenov, G. A., Liu, Y., Wilkins, M. R., Liang, W., Rubtsov, A., Sundev, G., Koyama, K., Turbek, S. P., Wunder, M. B., Stricker, C. A., & Safran, R. J. (2020b). Migratory divides coincide with reproductive barriers across replicated avian hybrid zones above the Tibetan Plateau. *Ecology Letters*, *23*(2), 231–241. https://doi.org/10.1111/ele.13420

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.

Scott, G. R., Egginton, S., Richards, J. G., & Milsom, W. K. (2009). Evolution of muscle phenotype for extreme high altitude flight in the bar-headed goose. *Proceedings of the Royal Society B: Biological Sciences*, *276*(1673), 3645–3653. https://doi.org/10.1098/rspb.2009.0947

Shirihai, H., & Svensson, L. (2018). *Handbook of western palearctic birds*. A. & C. Black.

Sokolovskis, K., Lundberg, M., Åkesson, S., Willemoes, M., Zhao, T., Caballero-Lopez, V., & Bensch, S. (2022). *Migration direction in a songbird explained by two loci* (Version 6, p. 183855940 bytes) [dataset]. Dryad. https://doi.org/10.5061/DRYAD.STQJQ2C6T

Somveille, M., Wikelski, M., Beyer, R. M., Rodrigues, A. S. L., Manica, A., & Jetz, W. (2020). Simulation-based reconstruction of global bird migration over the past 50,000 years. *Nature Communications*, *11*(1), 801. https://doi.org/10.1038/s41467-020-14589-2

Song, G., Zhang, R., DuBay, S. G., Qu, Y., Dong, L., Wang, W., Zhang, Y., Lambert, D. M., & Lei, F. (2016). East Asian allopatry and north Eurasian sympatry in Long-tailed Tit lineages despite similar population dynamics during the late Pleistocene. *Zoologica Scripta*, *45*(2), 115–126. https://doi.org/10.1111/zsc.12148

Storz, J. F., Runck, A. M., Sabatino, S. J., Kelly, J. K., Ferrand, N., Moriyama, H., Weber, R. E., & Fago, A. (2009). Evolutionary and functional insights into the mechanism underlying high-altitude adaptation of deer mouse hemoglobin. *Proceedings of the National Academy of Sciences*, *106*(34), 14450–14455. https://doi.org/10.1073/pnas.0905224106

Storz, J. F., Sabatino, S. J., Hoffmann, F. G., Gering, E. J., Moriyama, H., Ferrand, N., Monteiro, B., & Nachman, M. W. (2007). The Molecular Basis of High-Altitude Adaptation in Deer Mice. *PLoS Genetics*, *3*(3), e45. https://doi.org/10.1371/journal.pgen.0030045

Szép, T., Møller, A. P., Vallner, J., Kovács, B., & Norman, D. (2003). Use of trace elements in feathers of sand martin Riparia riparia for identifying moulting areas. *Journal of Avian Biology*, *34*(3), 307–320.

Tang, Q., Burri, R., Liu, Y., Suh, A., Sundev, G., Heckel, G., & Schweizer, M. (2022). Seasonal migration patterns and the maintenance of evolutionary diversity in a cryptic bird radiation. *Molecular Ecology*, *31*(2), 632–645. https://doi.org/10.1111/mec.16241

Tang, Q., Alayee, N., Liu, Y., Suh, A., Sundev, G., Heckel, G., & Schweizer, M. (n.d.). *The Hourglass Runs at Different Speeds in Riparia Sand Martins: Phylogenomic History and Diversification of a Cryptic Radiation (in revision with Journal of Biogeography).*

Taylor, R. S., Bramwell, A. C., Clemente-Carvalho, R., Cairns, N. A., Bonier, F., Dares, K., & Lougheed, S. C. (2021). Cytonuclear discordance in the crowned-sparrows, Zonotrichia atricapilla and Zonotrichia leucophrys. *Molecular Phylogenetics and Evolution*, *162*, 107216. https://doi.org/10.1016/j.ympev.2021.107216

Thorup, K., Pedersen, L., da Fonseca, R. R., Naimi, B., Nogués-Bravo, D., Krapp, M., Manica, A., Willemoes, M., Sjöberg, S., Feng, S., Chen, G., Rey-Iglesia, A., Campos, P. F., Beyer, R., Araújo, M. B., Hansen, A. J., Zhang, G., Tøttrup, A. P., & Rahbek, C. (2021). Response of an Afro-Palearctic bird migrant to glaciation cycles. *Proceedings of the National Academy of Sciences*, *118*(52), e2023836118. https://doi.org/10.1073/pnas.2023836118

Tian, H., Hammer, R. E., Matsumoto, A. M., Russell, D. W., & McKnight, S. L. (1998). The hypoxia-responsive transcription factor EPAS1 is essential for catecholamine homeostasis and protection against heart failure during embryonic development. *Genes & Development*, *12*(21), 3320–3324. https://doi.org/10.1101/gad.12.21.3320

Wang, G.-D., Fan, R.-X., Zhai, W., Liu, F., Wang, L., Zhong, L., Wu, H., Yang, H.-C., Wu, S.-F., Zhu, C.-L., Li, Y., Gao, Y., Ge, R.-L., Wu, C.-I., & Zhang, Y.-P. (2014). Genetic Convergence in the Adaptation of Dogs and Humans to the High-Altitude

Environment of the Tibetan Plateau. *Genome Biology and Evolution*, *6*(8), 2122–2128. https://doi.org/10.1093/gbe/evu162

Wang, W., Wang, F., Hao, R., Wang, A., Sharshov, K., Druzyaka, A., Lancuo, Z., Shi, Y., & Feng, S. (2020). First de novo whole genome sequencing and assembly of the bar-headed goose. *PeerJ*, *8*, e8914. https://doi.org/10.7717/peerj.8914

Zhang, X., Witt, K. E., Bañuelos, M. M., Ko, A., Yuan, K., Xu, S., Nielsen, R., & Huerta-Sanchez, E. (2021). The history and evolution of the Denisovan- *EPAS1* haplotype in Tibetans. *Proceedings of the National Academy of Sciences*, *118*(22), e2020803118. https://doi.org/10.1073/pnas.2020803118

Zhong, H., Kong, X., Zhang, Y., Su, Y., Zhang, B., Zhu, L., Chen, H., Gou, X., & Zhang, H. (2022). Microevolutionary mechanism of high-altitude adaptation in Tibetan chicken populations from an elevation gradient. *Evolutionary Applications*, *15*(12), 2100–2112. https://doi.org/10.1111/eva.13503

Zhou, S., Wang, X., Wang, J., & Xu, J. (2006). A preliminary study on timing of the oldest Pleistocene glaciation in Qinghai–Tibetan Plateau. *Quat. Int. 154–155:44–51*.

Acknowledgements

I would first like to acknowledge the Guangzhou Elite Scholarship and my beloved parents Mrs. Qian Rao and Mr. Yan Tang for their financial support to complete my PhD studies in Switzerland. I am profoundly thankful to my two advisors, Gerald Heckel and Manuel Schweizer, for their invaluable guidance in conducting scientific research with meticulous care. Their mentorship has been instrumental in nurturing my commitment to precision and thoroughness in the pursuit of knowledge. I would like to thank Darren Irwin for the evaluation of my thesis and Maddy Prakash Thakur for being the chair at my defence.

I would like to express my deepest gratitude to Prof. Yang Liu and the students and postdocs from his lab in Sun Yat-sen University and people from bird watching society of China for all the support and help to accomplish an excellent fieldwork in China. Especially I am thankful to Qin Huang, Xinyuan Pan, Xia Zhan, Chentao Wei and Wenjie Cheng who assisted me in the field on Qinghai-Tibetan Plateau and north western China around Taklamakan dessert and Dan Liang and Yun Li have collected samples in Sichuan. I would also like to thank Sarangua Bayrgerel (National University of Mongolia), Turmunbaatar Damba, Tuvshin Unenbat (Mongolian Ornithological Society), Paul Walser Schwyzer and Silvia Zumbach were of invaluable help with the fieldwork conducted by my advisor Manuel in Mongolia. Unfortunately, because of pandemic I was not able to participate the second fieldtrip to Mongolia.

It was an invaluable experience for me to both work at the Institute of Ecology and Evolution and Natural History Museum of Bern where I have met great colleagues and friends. I would like to express my heartfelt appreciation for Susanne Tellenbach not only for the help of lab work but for her family-like care through my whole PhD in Bern. I am grateful to Tamara Spasojevic, Pooja Singh and Sandra Oliveira for their indispensable support and guidance on pursuing a career in science. I would like to thank Reto Burri, Nilofar Alayee and Dave Lutgen for the scientific discussions. And a special thanks to David Marques and Nicolas Auchli for those amazing birding moments in Switzerland. Finally, to all my dear friends, Marion Talbi, Carlos Ramirez, Catia Chanfana, Ludovico Formenti, Hiranya Sudasinghe, Juliana Fuentes, Bárbara Calegari, Erina Balma, Laurance Etter, Beat Pfarre, Luyao Tu and Yin Jia, thanks for all the company in this wonderful journey.

Declaration of consent

on the basis of Article 18 of the PromR Phil.-nat. 19

Name/First Name: Qindong Tang

Registration Number: 17-127-838

Study program: Ecology and Evolution

Bachelor ☐ Master ☐ Dissertation ☐

Title of the thesis: Unravelling Cryptic Radiation and High-Altitude Adaptation in a Migratory Bird Species Complex

Supervisor: Gerald Heckel

I declare herewith that this thesis is my own work and that I have not used any sources other than those stated. I have indicated the adoption of quotations as well as thoughts taken from other authors as such in the thesis. I am aware that the Senate pursuant to Article 36 paragraph 1 litera r of the University Act of September 5th, 1996 and Article 69 of the University Statute of June 7th, 2011 is authorized to revoke the doctoral degree awarded on the basis of this thesis.
For the purposes of evaluation and verification of compliance with the declaration of originality and the regulations governing plagiarism, I hereby grant the University of Bern the right to process my personal data and to perform the acts of use this requires, in particular, to reproduce the written thesis and to store it permanently in a database, and to use said database, or to make said database available, to enable comparison with theses submitted by others.

Bern / 27.09.2023

Place/Date

Signature Tang Qindong .