Climate-driven flyway changes and memory-based long-distance migration

Zhongru Gu[1,2,3,18], Shengkai Pan[1,2,18], Zhenzhen Lin[1,2,18], Li Hu[1,2,3], Xiaoyang Dai[4], Jiang Chang[5], Yuanchao Xue[6], Han Su[1,2,3], Juan Long[1,2,3], Mengru Sun[3,6], Sergey Ganusevich[7], Vasiliy Sokolov[8], Aleksandr Sokolov[9], Ivan Pokrovsky[9,10,11] , Fen Ji[12], Michael W. Bruford[2,13] , Andrew Dixon[2,14,15,16], Xiangjiang Zhan[1,2,3,17*]

1. Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China.
2. Cardiff University – Institute of Zoology Joint Laboratory for Biocomplexity Research, Chinese Academy of Sciences, Beijing 100101, China.
3. University of the Chinese Academy of Sciences, Beijing 100049, China.
4. School of Biological Sciences, University of Bristol, BS8 1TQ, UK.
5. State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Sciences, Beijing 100012, China.
6. Key Laboratory of RNA Biology, CAS Center for Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China.
7. Wild Animal Rescue Centre, Krasnostudencheskiy pr., 21-45, Moscow 125422, Russia.
8. Institute of Plant and Animal Ecology, Ural Division Russian Academy of Sciences, 202-8 Marta Street, Ekaterinburg 620144, Russia.
9. Arctic Research Station of the Institute of Plant and Animal Ecology, Ural Division Russian Academy of Sciences, 21 Zelenaya Gorka, Labytnangi, Yamalo-Nenetski District 629400, Russia.
10. Department of Migration, Max Planck Institute of Animal Behavior, Am Obstberg 1, D-78315 Radolfzell, Germany.
11. Laboratory of Ornithology, Institute of Biological Problems of the North FEB RAS, 18 Portovaya Street 685000 Magadan, Russia.

12. State Key Laboratory of Stem Cell and Reproductive Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China.

13. School of Biosciences and Sustainable Places Institute, Cardiff University, Cardiff CF10 3AX, Wales, UK.

14. Emirates Falconers' Club, P.O. Box, 47716, Al Mamoura Building (A), Muroor Road, Abu Dhabi, United Arab Emirates.

15. Reneco International Wildlife Consultants, PO Box 61741, Sky Tower, Al Reem, Abu Dhabi, UAE.

16. International Wildlife Consultants, P.O. Box 19, Carmarthen SA33 5YL, UK.

17. Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming 650223, China.

18. These authors contributed equally.

*Corresponding authors:

Xiangjiang Zhan

Email: zhanxj@ioz.ac.cn

Millions of migratory birds occupy seasonally favourable breeding grounds in the Arctic[1], but we know little about the formation, maintenance and future of Arctic bird migration routes and genetic determinants of migratory distance. Here, we established a continental-scale migration system, satellite tracking 56 peregrine falcons (*Falco peregrinus*) from six Eurasian Arctic breeding populations and resequencing 35 genomes from four of these. Different breeding populations used five migration routes across Eurasia, likely formed by longitude and latitude breeding ground shifts during the LGM-Holocene transition. Contemporary inter-route environmental divergence appears to maintain distinct migration routes. We found that the novel gene *ADCY8* was associated with population-level migratory distance differences. We elucidated its regulatory mechanism and found the most likely selective agent for this divergence was long-term memory. Global warming is predicted to influence migration strategies and diminish breeding ranges of Eurasian Arctic peregrines. Harnessing ecological interactions and evolutionary processes to study climate-driven changes in migration can facilitate the conservation of migratory birds.

Global climate change and anthropogenic development are expected to affect the annual adaptive movements of migratory Arctic birds[1-3], with potential fitness effects imposed by inhospitable routes and temporally mismatched breeding[2,4]. Next generation genome sequencing has facilitated studies of the interaction between genomic variation and environment in migratory birds[5]. However, to date there is no published research on the role of climate-driven genomic responses in shaping differences of migratory strategy among bird populations. Here, we combined satellite-tracking of 56 peregrine falcons from migratory Arctic populations[6] (**Fig. 1a**, **Extended Data Fig. 1**, **Supplementary Table 1**) with genome data to explore their demographic history and the spatiotemporal dynamics of their migration.

Migration patterns of Arctic peregrines

From 41 individuals, we identified 150 completed migration paths (**Supplementary Table 2**). Peregrines initiated autumn migration in September, travelled 2,280-11,002 km, in *ca*. 27 days (95% confidence interval (CI): 14-46) covering 213 km/day (49-420), and arrived at their wintering areas in October. Peregrines migrate solitarily, with those departing from different breeding grounds, except Kola and Kolguev, using different routes and wintering at widely distributed sites across four distinct regions (**Fig. 1a**, **Extended Data Fig. 2**). Individuals tracked for more than one year exhibited strong path repeatability during migration (n = 26; R_{rpt} = 0.45, P < 0.001), complete fidelity to wintering locations and limited breeding dispersal (5.37 km on average; **Fig. 1b**, **Supplementary Table 3**). All populations demonstrated a high degree of migratory connectivity (R^2 = 0.86, P < 0.001; **Fig. 1c**), suggesting strong selection for long-term memory.

Principal component analysis (PCA) identified two main groups with migratory distance being the most significant differentiation (**Figs. 1d, e**, **Extended Data Fig. 3**, **Supplementary Table 4**). The Eastern birds flew significantly farther than Western birds (6,134 km *vs* 3,680 km; P < 1E-6; **Fig. 1e**). We therefore classified them as long-distance (LD; Kolyma-Lena-Popigai-Yamal) and short-distance (SD; Kolguev-Kola) migrants.

Historical formation of migration routes

We sequenced the genomes of 35 peregrines, obtaining 6,328,655 high-quality SNPs (**Supplementary Table 5)**. Multiple analytical approaches consistently supported four distinct genetic clusters corresponding to the sequenced populations, with Yamal and Kolyma inferred to have diverged after the separation of their ancestors from that of Kola and Kolguev (**Fig. 2a**). Sequential Markovian Coalescent (SMC++) analysis revealed that the effective population size (*Ne*) of the ancestral lineage increased from ~100 thousands of years ago (kya) to a peak 20-30 kya (**Fig. 2b**), around the Last Glacial Maximum (LGM)[7]. To resolve uncertainties in the recent demographic history (**Fig. 2a**), we developed a new Approximate Bayesian Computation (ABC) approach. The ABC-random forest model choice (**Fig. 2c**, **Supplementary Figs. 1-4**, **Supplementary Tables 6**, **7**) confirmed the divergence pattern of four studied populations and ABC simulations further found that Eastern and Western populations started to separate during the LGM (23.03 kya; 95% CI: 17.67-32.94), followed by an eastern split between Yamal and Kolyma 11.30 kya (9.14 -14.29), and between Kola and Kolguev 10.53 kya (9.18-12.90) (**Fig. 2d**, **Supplementary Table 8**, **Extended Data Fig. 4**).

Ecological Niche Modelling (ENM) based on present and paleo-climate datasets showed that potential breeding distribution range positively correlated with *Ne* fluctuations (**Supplementary Figs. 5, 6**). There was a much larger area suitable for breeding in Siberia during the LGM than the last interglacial period (LIG; 120-140 kya)[8] or Mid-Holocene, coinciding with the largest *N*e estimate (**Fig. 2e**, **f**, **Supplementary Fig. 5**). Arctic-dwelling peregrines mainly occupy tundra habitat[9], and we found close coincidence between reconstructed tundra habitat and peregrine breeding distribution in the LGM (**Supplementary Fig. 7**), suggesting that enlargement of tundra habitat underpinned peregrine population expansion during the LGM. Conversely, population declines and gradual divergences after the LGM mirrored large-scale loss and northward contraction of tundra. Recent population declines after the Mid-Holocene (**Fig. 2b**) may have also resulted from anthropogenic factors since habitat distributions have remained relatively stable (**Fig. 2f, g**).

Interestingly, our ENM simulations suggest that peregrines had less potential western wintering area during the LGM, while eastern wintering areas remained stable (**Fig. 2e**, **Supplementary Fig. 5**). Thus, during the LGM peregrines likely migrated to a wintering area across India and Southeast Asia, a striking south-eastward migration (**Fig. 2e**), distinct from the current south-westerly migration route formed during the Mid-Holocene (**Fig. 2f, g**). Furthermore, since the Mid-Holocene breeding areas are inferred to have shifted northward compared with the LGM, resulting in a longer migratory route (**Fig. 2e, f**), we conclude that glacial cycles can regulate both migratory orientation and distance.

Present migration route separation

We used the Hausdorff distance (*Hd*)[10] to quantify the distance between individual migration paths (**Methods** and **Supplementary Information**). Mean *Hd* within populations (17.05 ± 7.20) was significantly lower than that between neighbouring populations (35.83 ± 14.24, $P < 0.01$). Cluster analysis largely supported five migratory routes with Kola and Kolguev using the same route and very few individuals interchanging between populations (**Extended Data Fig. 5a**).

The proposition that migration routes are genetically determined, is mostly based on migratory restlessness[11], displacement experiments[12] and correlations between genetic background and migration route[13]. We addressed this fundamental question by randomly selecting 93 putatively neutrally-evolved SNP loci and 75 loci under positive selection (**Methods**). With these markers, we genotyped nine and six individuals respectively for the Popigai and Lena populations, where we obtained insufficient DNA from shed feathers for genome resequencing. Combining the genotypes with those from the population genomic data, we measured genetic differentiation (F_{ST}) among five peregrine populations and tested their relationship with the mean *Hd* for each migration route. We found a non-significant correlation between route *Hd* and F_{ST} ($R^2 = 0.02$, $P = 0.69$) based on neutral loci, suggesting that demographic isolation is not the major factor in isolating migration routes. In contrast, we found a significant relationship for loci under selection (R^2 = 0.42, P = 0.03), suggesting a role for adaptive genomic regions in the maintenance of separated migration routes (**Fig. 2h**).

Environmental divergence was evident among the migration routes, with Köppen-Geiger climate zones being significantly different between adjacent routes ($P < 0.01$, x^2 test; **Fig. 2i, Extended Data Fig. 5b**) and during most breeding and wintering periods (**Supplementary Fig. 8**). We found a coincidence between abrupt changes of climate zone and population route boundaries (**Extended Data Fig. 5c, d**). Migration cost based on least-cost paths was significantly lower for birds following their population-specific route (P = 0.01, effective size = 0.45; **Extended Data Fig. 5e, f**). We conclude that current migration routes are mainly maintained by environmental constraints, with synergistic contribution from local adaptation.

Genes for migratory distance differences

We used window-based F_{ST} and extended haplotype homozygosity (XP-EHH) to detect selection signals across the genomes of SD and LD peregrines. The methods combined identified 149 selection sweeps (37 genes) between the two groups (**Supplementary Table 9**), and found the most significant outlier occurred at the *ADCY8* locus (**Fig. 2j**). We narrowed the signal down to a 1.8 kb region, containing 14 linked SNPs in the second intron of the gene (**Fig. 2k**). Haplotype frequency analysis demonstrated a positive selection signature of *ADCY8* in the LD group (**Fig. 2l**). The dominant haplotype (Hap2) was at a high frequency (34/38) in the LD group, whereas the SD group had six haplotypes with varying frequencies (Hap2 11/26, Hap1 6/26, with the other four haplotypes occurring 4, 2, 2, 1 times; **Fig. 2l**). We investigated the potential functional significance of Hap2 in LD peregrines and a randomly selected Hap1 sequence from SD peregrines by designing a dual luciferase reporter assay for functional analysis in our cultured chicken hippocampus primary cells (**Methods**). In contrast to Hap2 insert cells, the Hap1 insert cells showed a significantly lower luciferase activity ($P < 0.001$; **Fig. 2m**, **Supplementary Fig. 9**), suggesting that the peregrine Hap1 sequence has a suppressing effect.

Of the 14 loci identified within the *ADCY8* locus, the SNP (*C*/*T*) in the position 5,170,169 of peregrine chromosome 3 produced the largest XP-EHH value (**Fig. 2n**) and the allele *T* was 100% fixed in the LD group, suggesting that this standing variation is under strongest selection and may have a major role in functional differentiation. A search of the focal fragment against the motif database found that the ancestral SD group had a 5'-*CGTCA*-3' motif, a canonical half-site cAMP-responsive element (CRE) that is a binding site for the transcription factor CREB[14,15], while the fixed chr3-5170169*T changes the first nucleotide of the motif. We used ATAC-seq[16] to sequence both hippocampus and cerebral cortex tissues from a SD peregrine. Our ATAC-seq analysis detected a significant peak coinciding with the position of this motif (**Fig. 2n)**, experimentally supporting its existence as a functional element.

Previous studies suggest that CREB can regulate gene expression by binding to the CRE element through CREB basic region/leucine zipper domain (bZIP), which can be regulated by its DNA methylation level[17,18]. In peregrines, the identified substitution from *C* to *T* in *ADCY8* gene creates a novel transcriptional binding site, 5'-*TGTCA*-3', which potentially disrupts the DNA methylated site, CpG island, on the canonical motif. Moreover, we found that, *CREB1*, the transcription activator, expressed the bZIP domain in the peregrine brain, but the conditional repressor *CREB2*[19] did not (**Supplementary Fig. 10**). Our results suggest that the new CRE motif may be free from DNA methylation and facilitate the binding of CREB1 on *ADCY8*, and will consequently maintain a higher activity of ADCY8 in LD peregrines. Supporting evidence came from our comparison of expression levels of *ADCY8* in the peregrine brain with two genotypes of the focal SNP (*RPKM* = 81.11 in *CT vs* 73.96 in *CC*; $P = 3.73\text{E-}5$, hypergeometric test).

Empirical evidence indicates that *ADCY8* is involved in long term memory[20,21]. *ADCY8* encodes Adenylyl Cyclase type 8 that catalyses the conversion of ATP to cAMP, which acts as a secondary messenger and regulates downstream memory-related genes[22,23]. We found LD peregrines had a significantly higher mean migration path fidelity than SD peregrines (*P* < 0.001; **Supplementary Fig. 11**), requiring strong long-term memory. Together with numerical evidence that *CREB* genes determine the development of long-term memory[24], our work suggests that the *ADCY8* and *CREB* genes play a key role in influencing migratory flight distance via co-regulating the capacity of long term memory. For the *T* allele that is positively selected in the LD populations, the selection time was estimated to be 18.87 kya (**Supplementary Information**), after the separation of two migratory groups, strengthening our conclusion that the regulation of migratory distances is the result of natural selection.

Predicted effect of global warming

We used ENM simulations to project future (2070) breeding and wintering distributions for each Arctic peregrine population under Representative Concentration Pathway (RCP) 8.5. The breeding and wintering distributions of all populations would shift poleward by 2.08° (95% CI: 0.31-3.44) and 1.47° (0.11-11.06) latitude, respectively (**Extended Data Fig. 6**, **Supplementary Fig. 12**), which is consistent with observations for most Arctic shorebirds[25] and congruent with the climatic envelope corresponding to tundra habitats. Greatest reduction is predicted to occur in Kolguev and Kola, losing 100% and 93% of their suitable breeding habitats, respectively (**Fig. 3a**). We also found Western peregrines may have a much shorter migration route (655 km; 442-868), while Eastern peregrines may have a longer route (286 km; 56-515) (**Fig. 3a**, **Supplementary Fig. 13**). If the climate warms at the same rate as over recent decades, peregrines in Western Eurasia may stop migrating altogether, while Eastern peregrines may face greater risks since mortality is positively associated with migratory distance[26].

Recent population declines in migratory Arctic birds have been attributed to the amplification of global warming in the High Arctic[1,4,25]. Climate change may have already impacted peregrine populations, so we compared *Ne* changes of each population with local temperature during breeding periods (May-July) since 1840. Our *SNeP* analysis showed that each population has undergone declines during the past 25 generations (*ca*. 150 years). However, *Ne* slope (*NeS*) analysis further revealed some variation, with recovery detected during relatively cooler summers (**Fig. 3b**), and four populations showed the largest negative *NeS* 8~9 generations ago (1960s), coinciding widespread use of organochlorine pesticides[27]. Importantly our generalized linear modelling (GLM) analysis found that *Ne* change was negatively correlated with mean breeding season temperature in Kola ($R^2 = 0.46$, $P = 0.03$) and Yamal ($R^2 = 0.39$, $P = 0.02$), with mean breeding season temperature and duration of extreme warm days in Kolguev ($R^2 = 0.61$, $P = 0.01$), and with duration of extreme cold/warm days in Kolyma ($R^2 = 0.68$, $P = 0.02$; **Fig. 3c**, **Supplementary Table 10**). Our predicted *NeS* in the future (i.e. from 2020 to 2100) showed a continuing decrease trend, but the SD migrants in Kola and Kolguev will suffer the highest probability of population decline (**Fig. 3d**, **Supplementary Fig. 14**).

Discussion

Spatiotemporal changes in animal migration behaviour are thought to be related to climatic changes, anthropogenic impacts[28] and evolutionary responses of migrants. Since these dynamic processes can leave a footprint on the genome, by combining animal movement and population genomic data, we were able to identify a major role of climate in the formation and maintenance of peregrine migration patterns (**Fig. 2**, **Extended Data Fig. 5**).

Previous studies have identified several candidate genes that may regulate migration[29,30]. The higher activity of ADCY8 we identified in LD peregrine migrants (**Fig. 2**) may increase their long-term memory. Our analysis reveals a unique mutation that facilitates the binding of its transcription factor CREB1, and fixation of this variation happened after the divergence of LD and SD populations. Our work thus not only reveals a novel causative gene to explain migratory differences, but also provides a mechanistic basis.

In a changing global climate, peregrines may move to new wintering areas and adjust their migration routes. However, our prediction of dramatic shrinkage in Arctic breeding areas, together with a predicted population collapse in the European Arctic, represents a clear threat to peregrines and possibly many other migratory Arctic species. Our study demonstrates the value of an integrated approach, combining satellite telemetry, population and functional genomics and modelling, to untangle intriguing scientific questions related to migration, laying a cornerstone for conserving migratory species in conjunction with ecological interactions and evolutionary processes.

References

1. McRae, L. et al. *Arctic Species Trend Index 2010. Tracking Trends in Arctic Wildlife*. *CAFF International Secretariat* (2010).
2. Lameris, T. K. et al. Potential for an Arctic-breeding migratory bird to adjust spring migration phenology to Arctic amplification. *Global Change Biology* **23**, 4058-4067 (2017).
3. Trautmann, S. Climate Change Impacts on Bird Species. In *Bird Species* (pp. 217-234). Springer, Cham (2018).
4. Zurell, D., Graham, C. H., Gallien, L., Thuiller, W. & Zimmermann, N. E. Long-distance migratory birds threatened by multiple independent risks from global change. *Nature Climate Change* **8**, 992 (2018).
5. Bay, R. A. et al. Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science* **359**, 83-86 (2018).
6. White, C. M., Cade, T. J. & Enderson, J. H. *Peregrine Falcons of the World* (Lynx Edicions, Barcelona, 2013).
7. Clark, P. U. et al. The last glacial maximum. *Science* **325**, 710-714 (2009).
8. Otto-Bliesner, B. L., Marshall, S. J., Overpeck, J. T., Miller, G. H. & Hu, A. Simulating Arctic climate warmth and icefield retreat in the last interglaciation. *Science* **311**, 1751-1753 (2006).

9. Brambilla, M., Rubolini, D. & Guidali, F. Factors affecting breeding habitat selection in a cliff-nesting peregrine *Falco peregrinus* population. *Journal of Ornithology* **147**, 428-435 (2006).
10. Hausdorff, F. Bemerkung über den Inhalt von Punktmengen. *Mathematische Annalen* **75**, 428-433 (1914).
11. Pulido, F. The genetics and evolution of avian migration. *Bioscience* **57**, 165-174 (2007).
12. Perdeck, A. C. An experiment on the ending of autumn migration in starlings. *Ardea* **52**, 133-139 (1964).
13. Delmore, K. E., Toews, D. P., Germain, R. R., Owens, G. L. & Irwin, D. E. The genetics of seasonal migration and plumage color. *Current Biology* **26**, 2167-2173 (2016).
14. Impey, S. et al. Stimulation of cAMP response element (CRE)-mediated transcription during contextual learning. *Nature Neuroscience* **1**, 595 (1998).
15. Bourtchuladze, R. et al. Deficient long-term memory in mice with a targeted mutation of the cAMP-responsive element-binding protein. *Cell* **79**, 59-68 (1994).
16. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods* **10**, 1213 (2013).
17. Mayr, B. & Montminy, M. Transcriptional regulation by the phosphorylation-dependent factor CREB. *Nature Reviews Molecular Cell Biology* **2**, 599 (2001).
18. Iguchi-Ariga, S. M. & Schaffner, W. CpG methylation of the cAMP-responsive enhancer/promoter sequence *TGACGTCA* abolishes specific factor binding as well as transcriptional activation. *Genes & Development* **3**, 612-619 (1989).
19. Bartsch, D. et al. Aplysia CREB2 represses long-term facilitation: relief of repression converts transient facilitation into long-term functional and structural change. *Cell* **83**, 979-992 (1995).
20. Wieczorek, L. et al. Absence of Ca^{2+}-stimulated adenylyl cyclases leads to reduced synaptic plasticity and impaired experience-dependent fear memory. *Translational Psychiatry* **2**, e126 (2012).

21. Rosenegger, D., Wright, C. & Lukowiak, K. A quantitative proteomic analysis of long-term memory. *Molecular Brain* **3**, 9 (2010).
22. Ferguson, G. D. & Storm, D. R. Why calcium-stimulated adenylyl cyclases? *Physiology* **19**, 271-276 (2004).
23. Zhang, M. et al. Ca-stimulated type 8 adenylyl cyclase is required for rapid acquisition of novel spatial information and for working/episodic-like memory. *Journal of Neuroscience* **28**, 4736-4744 (2008).
24. Yin, J. C. & Tully, T. CREB and the formation of long-term memory. *Current Opinion in Neurobiology* **6**, 264-268 (1996).
25. Wauchope, H. S. et al. Rapid climate-driven loss of breeding habitat for Arctic migratory birds. *Global Change Biology* **23**, 1085-1094 (2017).
26. Lok, T., Overdijk, O. & Piersma, T. The cost of migration: spoonbills suffer higher mortality during trans-Saharan spring migrations only. *Biology Letters* **11**, 20140944 (2015).
27. Brown, J. W. et al. Appraisal of the consequences of the DDT-induced bottleneck on the level and geographic distribution of neutral genetic variation in Canadian peregrine falcons, *Falco peregrinus*. *Molecular Ecology* **16**, 327-343 (2007).
28. Wilcove, D. S. & Wikelski, M. Going, going, gone: is animal migration disappearing. *PLoS Biology* **6**: e188 (2008).
29. Mueller, J. C., Pulido, F. & Kempenaers, B. Identification of a gene associated with avian migratory behaviour. *Proceedings of the Royal Society B: Biological Sciences* **278**, 2848-2856 (2011).
30. Peterson, M. P. et al. Variation in candidate genes *CLOCK* and *ADCYAP1* does not consistently predict differences in migratory behavior in the songbird genus Junco. *F1000Research* **2**, 115 (2013).

Figure Legends

Figure 1. Migration system. **a**, Five migration routes for 56 satellite tracked peregrine falcons in Old World. Only complete migration path is shown (n = 41). **b**, Individual migration path fidelity of one representative individual (Lena Delta) tracked for four years (2010-2014). **c**, Migration connectivity at the population level as shown in a linear regression analysis. Significance was calculated using F-test (P = 2.2E-16). **d**, PCA analysis of migratory strategy. Small and large dots represent individuals (n = 41) and centroids of the minimum convex polygon for each population (n = 6). PC1 is the mix of autumn departure and arrival dates and PC2 is the migration distance. **e**, Migratory distance comparison (P < 1E-6, two-sided t-tests, effect size = 1.43) between short-distance (SD) (n = 12 individuals; 3,680 km: 95% CI: 2,443-5,018) and long-distance (LD) groups (n = 32 individuals; 6,134 km: 95% CI: 3,282-8,828). In the box plots, the center line represents the median, whiskers represent maximum and minimum values, and box boundaries represent 75th and 25th percentiles. The P values for the comparisons for any two populations within SD and LD were not significant (ns; P > 0.05, two-sided t-tests). The tracking data of three additional Kola peregrines from a previous study (**Methods**) were also used for this analysis.

Figure 2. Past formation, present maintenance of migration routes and genetic basis for migration distance differences. a, PCA, Neighbor-Joining tree and *frappe* showing the evolutionary relationship of the four genome-enabled populations. **b**, Demographic history reconstruction for each population using SMC++. **c**, Four candidate models for model choice in ABC. 84% of the total of 313 chunks (**Methods**) support Model 1. T_1, T_2, T_3 are divergence times. **d**, Posterior distribution of divergence time estimates for the Model-1 in ABC. **e-g**, Species distributions predicted during the LGM, Mid-Holocene and present. **h**, The relationship between the route distance, *Hd*, and neutral genetic distance (Left), $F_{ST}/(1-F_{ST})$, and genetic distance based on selected loci (Right). The dashed line is the linear regression line. Significance levels were calculated using F-test. **i**, Proportion of grids (0.083° × 0.083°) with different Köppen-Geiger climate zones within each migration route. Full names of climate zones can be seen in

Supplementary Information. **j**, Spline-window based F_{ST} and XP-EHH to detect selective sweeps. Red points indicate windows containing selected genes. **k**, *ADCY8* haplotype heatmap. The dashed rectangle marks the focal 1.8kb fragment in *ADCY8*. Red and light blue squares symbol different alleles in each column (SNP). **l**, Haplotype frequency in the identified segment. **m**, Results of dual-luciferase reporter assay in chicken hippocampus primary cells. Data are mean ± s.e.m. Significance levels were calculated using two-sided *t*-test (n = 14 replicates for each of the first three groups and n = 6 for the *pGL3*-basic group). **n**, XP-EHH results for every selected SNPs within 1.8 kb flanking regions (Upper) and ATAC-seq results confirming the existence of CRE-motif (Lower).

Figure 3. Shortened migration route and population decline in Europe populations due to global warming. a, Area changes in breeding and wintering areas (Upper) between present and future (2070; RCP 8.5) and migratory distance comparisons (Lower) in the six peregrine populations. N/A: no predicted future breeding areas. △d is the mean change of migration distances. In the box plots, the center line represents the median, whiskers represent maximum and minimum values, and box boundaries represent 75th and 25th percentiles. n = 200 for each comparison. Significance and effect size were calculated using two-sided *t*-tests and Cohen's *d*, respectively (Kola: P = 3.4E-9, effect size = 0.580; Yamal: P = 0.090, effect size = 0.170; Popigai: P = 0.676, effect size = 0.042; Lena: P = 0.868, effect size = 0.017; Kolyma: P = 0.015, effect size = 0.243). **b**, LD_{Ne} and *NeS* estimates in recent 25 generations. **c**, The linear regression between *NeS* and the most significant environmental variable identified in the GLM analysis. Data are median ± 95% CI. Dashed lines represent the linear regression lines. **d**, *NeS* changes of each population predicted in the future. Data are median ± 95% CI.

Methods

Tracking peregrine migration

We used satellite-received Argos Platform Transmitter Terminals (PTTs) and GSM-received GPS transmitters to track 56 peregrines from six breeding regions in Arctic Eurasia (**Supplementary Table 1**). From which, we obtained the data of at least one full migration for 41 peregrines: Kola Peninsula (*n* = 1 autumn/0 spring migration), Kolguev Island (*n* = 8/4), Yamal Peninsula (*n* = 9/5), Popigai River (*n* = 10/9), Lena Delta (*n* = 6/6) and Lower Kolyma River (*n* = 7/5) (**Supplementary Information**). Permits to trap, collect blood samples and deploy satellite transmitters on peregrines were provided by the relevant authorities in Russia. All lab experiment procedures were under the guidance of the Ethics Committee of the Institute of Zoology, Chinese Academy of Sciences (IoZ, CAS). Studies in this manuscript involving peregrine brain tissue collection and analyses were in full compliance with the Institutional Animal Care and Use Committee at the IoZ, CAS.

For tracking data processed by the Argos system, we removed duplicate timestamps and used the Douglas Argos Filter algorithm designed to retains points (**Supplementary Fig. 15**), which correspond to a realistic rate of movement (≤ 100 km/h) and which do not form tight angles between successive locations (≤ 15°)[31]. Migration strategy[32,33] was quantified using departure/arrival date, duration and migratory distance (migration path distance) for all individuals, where possible (**Supplementary Fig. 16**). We defined the start and end of migration as the day that birds moved more than 40 km from their breeding (natal) range, or arrived at wintering range.

Population migration connectivity and wintering distribution pattern

To quantify the degree of migratory connectivity[34], we extracted the longitudes of breeding and wintering sites for each tracked individual and used a linear regression to explore the correlation between breeding sites and wintering sites as used in a previous study[33]. The coefficient of determination (R^2) was used to proxy the migratory connectivity. For the spatial distribution pattern of wintering sites, we obtained the minimum convex polygon (MCP) of wintering sites and used a *G*-function from *R* package *spatstat*[35] to conduct a point pattern analysis. A greater empirical $\hat{G}(x)$ than the theoretical function suggests that the sites tend to be closer than expected, in contrast with a dispersed pattern. More details please see **Supplementary Information**.

Individual migration path fidelity

We estimated the path fidelity of each bird by assessing individual repeatability[36] of migration paths across multiple years ($n \geq 2$). To quantify the consistency of these migratory paths, we first calculated the track deviation from the great circle distance for each path in each year and then evaluated the repeatability of deviation after standardized measurement based on latitude. Specifically, we measured the total migratory distance (sum of all distances between successive positions along a migration path). Then, the straightness was calculated as the distance between the start and end locations of the path divided by total migratory distance[37]. The repeatability of path straightness was calculated using a linear mixed model implemented in the *R* package *rptR*[38], followed by a significance testing through a permutation test.

To compare the path fidelity between SD and LD migratory groups, we considered the latitudes and longitudes of all sites along each migration path as response and independent variables respectively, and conducted a linear regression analysis to estimate the regression coefficient α of an individual in different migration periods. We calculated the standard variation of α (α_{sd}), as a proxy of individual path fidelity, that is, a lower α_{sd} indicates that an individual uses a more similar path across migration periods. We compared differences in the mean α_{sd} estimates between SD and LD groups using a *t*-test.

Migratory strategy comparison

Principal component analysis was used to cluster individuals based on their migratory strategy (i.e. migratory distance, duration and departure/arrival date during autumn migration). Noted that we did not use the spring departure/arrival date because we did not obtain these data for the Kola population. In our study, only individuals that completed at least one migration route were included, and for individuals that were tracked for multiple years, we used mean migratory values to control pseudo-replication. We removed one of the variables if they were highly related ($|r| > 0.75$). For the comparison of migratory strategy between SD and LD groups, we used a random forest model in *R* and *t*-test to detect the most significantly different strategy parameter. We calculated the effect size (Cohen's *d*[39]) for the *t*-test. In the comparisons, as we obtained the data of only one complete autumn migration in the Kola population (**Supplementary Table 2**), we also used the tracking data of three Kola peregrines from a previous study[40].

Sample collection and genomic DNA extraction and sequencing

Blood samples were collected for the genome resequencing of 35 peregrines across the Eurasian Arctic (10 from Kola, 5 from Kolguev, 11 from Yamal and 9 from Kolyma). We also obtained nine and six feather samples from Popigai and Lena, respectively. Genomic DNA was extracted from blood and feather samples using the Blood & Cell Culture DNA Midi Kit and Blood & Tissue Extraction Kit (Qiagen), respectively. Paired-end libraries with insert size of 170 bp for blood DNA were constructed and subjected to sequencing on an Illumina HiSeq 2000 platform in BGI, Shenzhen and Novogene, Beijing. The feather DNA was used for the following PCR experiments.

Sequencing data filtering and SNP calling

An average of 68.75 Gb clean data (55.79×) were generated for 35 individuals. We used FASTQC and trimmomatic[41] to remove reads with low quality as previously described[42]. We used the chromosome-level peregrine genome assembly[43] as the reference genome, where our original assembly[44] was upgraded to chromosomal fragments. We then used the Burrows-Wheeler Alignment[45] to map the filtered reads from each individual onto the autosomal reference genome with *Z*-chromosome fragments excluded (**Supplementary Information**). Finally, we used the pipeline in Genome Analysis Toolkit[46] (version 3.5) to call SNPs.

Population genomic analysis

Since close relatedness can bias population assignment[47], pairwise Identity-By-State scores among all individuals were estimated using PLINK[48] (version 1.9) and a threshold of 0.1 was applied, resulting in the removal of three closely related individuals for the following analysis (**Supplementary Fig. 17**). PCA was conducted on the autosomal biallelic SNPs for the remaining 32 individuals using our in-house scripts. To reconstruct the phylogenetic tree, we used PLINK to calculate genetic distances among the studied peregrines based on the identified SNPs with default settings. A neighbor joining unrooted tree was then obtained using the *R* package '*phangorn'* with upgma function[49]. Analysis of genetic structure was implemented in *frappe*[50] which employs an expectation maximization algorithm. The number of genetic clusters *K* was set to range from 2 to 5.

Demographic history reconstruction

We used SMC++ (version 1.10.0)[51] to model historical effective population sizes for each peregrine population with the mutation rate and generation time derived from our previous estimates[44]. To date the divergence among the peregrine populations, we developed a new ABC approach (details in **Supplementary Information**). Briefly, we established four candidate historical demographic models for model choice according to our phylogenetic results. The prior *Ne* distributions and divergence time parameters were set from a range of 1,000 to 100,000 and 1,000 to 10,000 generations ago, respectively, according to the SMC++ results. The peregrine genome was divided into chunks with a size of 2 Mb and gene number in each chunk was counted and ranked. We conducted 100,000 simulations for each candidate model using the rapid coalescent *scrm* simulator[52] and summarized them using 95 different summary statistics (e.g. the total number of segregating sites, summarized site frequency spectrum). A machine leaning tool, ABC random forest[53] was employed to conduct model choice. For the selected model with the highest approximated posterior probability (Model-1), we further simulated 1,000,000 datasets for parameter inferences. The neural network methods were applied for the inference in *R* package *abc*[54]. To evaluate the ABC performance, we applied the cross validation on model choice and parameter inference.

Ecological niche modeling

We used MaxEnt (version 3.3.3k), in the *R dismo*[55] package, to predict breeding and wintering distributions under present environment conditions. Based on the satellite tracking data, we randomly selected presence data within the MCPs of individuals' summer (June to August) and wintering areas (December to January). The 90% MCP was calculated for each individual using the *MCP* function of the *adehabitatHR*[56] package in *R*. For climate data, we downloaded 19 present and paleo bioclimatic variables (**Supplementary Table 11**) from WorldClim[57]. To reconstruct breeding and wintering distributions in the past, we projected the ENMs built under present climate to paleoclimates during the Mid-Holocene (*ca*. 6 kya), LGM (22 kya) and LIG (120-140 kya), respectively.

Paleo vegetation data analysis

To examine whether the predicted LGM breeding areas mostly consisted of tundra biome, we obtained Eurasian paleo pollen data from a previous study[58]. Paleo pollen data classified as tundra biome by the biome_2000 model were extracted for LGM and then mapped to our predicted breeding areas to estimate the overlapping extent.

Quantification of inter-route distances

Hausdorff distances were used to quantify the dissimilarity between migratory paths of individual pairs of peregrines. The approach measures how far apart two subsets of a space metric are from each other[10]. In our study, a migratory path was treated as positional distribution of bird movement points in time and space and the distance between the migration path *A* and *B* was obtained using '*hausdorff_dist*' function in *R* package *pracma*[59] (**Supplementary Information**). The mean *Hd* estimates were compared within and between neighboring populations using a *t*-test. The generated

Hd matrix was then used for the route clustering analysis using a '*hclust*' method in *R* package '*heatmap*'. The inter-route distance was finally calculated as the mean *Hd* between pairs of individual paths from different migration routes.

Maintenance mechanisms of present migration routes

To investigate maintenance mechanisms of present migration routes, we first checked the influence of neutral genetic and selective genetic distance on the route distances estimated above. For the estimation of genetic distance, we randomly selected 93 putatively neutrally-evolving and 75 selected SNP loci based on the selection analysis of 32 resequenced peregrine genomes described below. Then, we genotyped nine and six individuals (shed feathers) sampled in Popigai and Lena, respectively. The detailed PCR amplifications and sequencing of these feather DNA extracts are described in **Supplementary Information**. With the combined genotypes from blood and feather samples, we calculated the F_{ST} for each locus among five populations (Kolguev, Yamal, Popigai, Lena and Kolyma) using *vcftools*[60]. The genetic distance was then calculated as $F_{ST}/(1- F_{ST})$[5]. We fitted the relationship between migration route separation (*Hd*) and neutral and selective genetic distance, respectively, using a linear regression model.

To investigate the environmental divergence among migration routes, we randomly sampled 200 grids (0.083° × 0.083°) from the 90% MCP of each route and extracted the variable of climate zones referring to the Köppen-Geiger classification system[61] from each grid. Chi-squared tests were applied for the testing of differences in climate zones between adjacent migration routes (**Extended Data Fig. 5b**). To check the environmental boundaries, we further divided the Eurasian continent into geographic bands (width in 2° longitude based on the estimated migration distance per day in the studied peregrines) with direction parallel to the mean migration angle of all individuals and calculated the median value of climate zones of the grids (0.083° × 0.083°) at regular intervals (1° in latitude) from neighbouring bands for paired comparisons (illustrated in **Extended Data Fig. 5c**). Paired *t*-tests were used to check the abrupt change of climate zone (boundary) (**Extended Data

Fig. 5d) between adjacent bands. It is noted that during comparison, we used the same latitude range between the pairwise routes.

To test whether there is more benefit from being a conventional migrator (migrating within a population route) or an unconventional migrator (migrating across routes), we first proved that the tracked peregrines migrated in a least-cost manner (**Supplementary Fig. 18**). We then simulated scenarios that peregrines depart from their actual breeding sites, fly along least-cost paths, but winter in the actual wintering sites of neighboring routes (illustrated in **Extended Data Fig. 5e**). Taking account for migration path length, we estimated the relative least cost of cross-route migration (**Supplementary Information**). A *t*-test was used to compare the difference in the relative least cost between the actual within-route and simulated cross-route migration. For the *t*-test, the effective size *d* was calculated.

Identification of selective sweeps and detection of selected SNPs between SD and LD peregrines

We used two methods, a window-based F_{ST} and XP-EHH[62] to identify selective sweeps between the SD and LD groups. The F_{ST} of each locus was calculated using *vcftools*. A smoothed spline technique in *R* package *GenWin*[63] (version 1.0.1) was implemented to determine the window boundary and *w*-statistic was used as a proxy of windowed F_{ST}. The XP-EHH value of each locus was calculated using *selscan*[64] (version 1.1) with BEAGLE-phased[65] SNPs (*n iterations*=100). Outlier regions (top 1%) detected by both methods were considered as selective sweeps. For the sweeps identified on the focal gene, we calculated the nucleotide diversity (θ_π) of each locus using *vcftools* and narrowed the sweep down to a specific region. In addition, to verify the phased haplotype of *ADCY8* gene, we chose three individuals from each population for 10x genomics sequencing, followed by linked-reads phasing using the Long Ranger[66] (version 2.2.2) (**Supplementary Table 12**).

We then used an integrative method[67] to detect the selected SNPs between the pairwise groups (Kolguev *vs* LD and Yamal *vs* Kolyma), which integrated the results obtained from three selection tests: the *FLK*[68] based on comparing different patterns of allele frequencies among populations to the values expected under a scenario of neutral evolution[69], latent factor mixed models (LFMM) for which the environment is used as a fixed effect and latent factors used to infer environmental associations[70], and *pcadapt* (version 3.03) analysis based on Bayesian factor model. Detailed settings were described in our previous work[71]. The adjusted *z*-scores were calculated for each of three above tests, and the calibrated *P* values were obtained as previously reported[70]. The candidate SNP loci were ultimately determined using Benjamini-Hochberg FDR (false discovery rate) control. The level of FDR was set to 0.05.

Luciferase reporter assay for the focal *ADCY8* haplotype

To investigate the potential functional significance of different *ADCY8* haplotypes, a randomly selected Hap1 in SD peregrines and the dominant Haplotype (Hap2) in LD peregrines (**Fig. 2l**) were fully synthesized in SinoGenoMax Co., Ltd and inserted into the *pGL3*-promoter backbone according to KpnI/XhoI restriction sites. Positive (*pGL3*-promoter) and negative (*pGL3*-basic) controls were also constructed. The activity of Hap1 or Hap2 was examined in primary cells cultured from chicken embryonic hippocampus tissues. We isolated and cultured neurons from fertilized chicken eggs (Boehringer Ingelheim; details in **Supplementary Information**) for lipofection and luciferase reporter assays referring to previous studies[72]. *pGL3*-Hap1 or *pGL3*-Hap2 were co-transfected with pRL-TK into the neurons using Lipofectamine 2000 (Invitrogen). After 48 h incubation, the dual-luciferase activity was measured using Dual-Glo® Luciferase Assay kit (Promega). A *t*-test was used to compare the fluorescence intensity among experimental groups. At least three independent experiments of each assay were performed with a minimum of six replicates. To predict the motif in this focal region, we extracted the sequences (10 bp) around each SNP and searched the sequences against the TFBS motif database[73].

ATAC-seq and RNA-seq analysis

Hippocampus and cortex tissues were collected from a peregrine that died of natural causes in the Chongqing Zoo. For the ATAC experiment (**Supplementary Information**), the samples were prepared according to the manual[74] in Shanghai Jiayin Biotechnology Ltd, followed by library construction and subjected to sequencing on an Illumina NovaSeq 6000 (Novogene). The raw reads generated were further quality controlled. The clean reads were mapped to our reference peregrine genome and sequencing depth around the target region was evaluated.

For the RNA-seq, brain tissues were collected from two humanely euthanized peregrines from the Beijing Raptor Rescue Center. Total RNA was extracted from the samples of two peregrines (chr3-5170169SNP genotype *CC* and *CT*) using the TRIzol reagent according to the user guide (Invitrogen). For each brain RNA sample, library with insert-size of 350 bp was constructed and then sequenced on a HiSeq 2500 (Novogene). *Reads Per Kilobases per Million reads* (*RPKM*) of *ADCY8* with different genotypes were calculated by mapping RNA-seq reads to the peregrine gene set[44] using SOAP[75] (version 2.22). Expression difference between these genotypes was compared using a hypergeometric test.

Detailed descriptions of the ATAC-seq please see **Supplementary Information**.

Global warming impacts and prediction of migratory distance in the future

Based on the inferred present peregrine distribution, we predicted future (2070) potential breeding and wintering distributions under RCP 8.5, a scenario where emissions continue to rise through the 21st century. Occurrence probabilities were transformed into binary maps using true skill statistic-maximizing values as thresholds. Differences between present and future distributions were investigated using two parameters: area change and latitude shift. We counted the grids (0.083° × 0.083°) within the non-overlap regions between present and future, and the shifted latitude was represented as the ratio of area and longitude (per degree). To compare the migratory distance between the present and future, we randomly selected 50 sites per population within breeding or wintering ranges and calculated great circle distances between corresponding sites. The confidential interval of these distance estimates were calculated as the minimum and maximum of great circle distances among all the sites in each of breeding and wintering ranges. For comparison, we conducted the same analyses under RCP 4.5, a mitigation scenario where emissions peak around 2040 (**Supplementary Information**).

Prediction of effective population size changes in the future

To study how effective population size will change under future climate change, we initially investigated the association between most recent changes of *Ne* and climate variables since the industrial revolution (1840). We used a linkage disequilibrium based method *SNeP*[76] (version 1.11) to reconstruct recent *Ne* changes in generations for each population. For this analysis, we randomly selected 5,000 loci per chromosome and used a *NeS* to investigate the rate of *Ne* changes[77]. It is noted that *SNeP* could reliably examine the changes or trends of *N*e, rather than the actual *N*e[76].

For the climate data, the monthly temperature of the Community Climate System Model (version 4) output was downloaded from Coupled Model Intercomparison Project 5. The yearly mean temperature was calculated as the average of monthly temperature in the breeding season (May to July), a vital period for the breeding success of peregrines[78,79]. To quantify the historical weather extremes, we downloaded the gridded daily minimum and maximum temperature data from Berkeley Earth Surface Temperature. The number of days exceeding the 95% upper threshold of average temperature was considered as extreme hot days and below the 5% lower threshold as extreme cold days. The simulated future daily temperature (2020 to 2100) was downloaded from the NASA Earth Exchange Global Daily Downscaled Projections. The inferring method for future weather extremes was the same as historical extremes.

We constructed a GLM to model the relationship between *NeS* and changes of climate variables (**Supplementary Information**) and finally predicted the future *NeS* under future climate using the fitted GLM model.

Statistical analysis

All reported *P* values were from student *t*-tests (two-sided) unless otherwise specified. All assays were performed in at least three independent experiments with a minimum of six replicates. In the analysis of individual route repeatability, *P* values were calculated using a permutation test. The comparisons of climate zones between migration routes were calculated using an x^2 test. The *P* values of band comparisons of climate zones were calculated using a paired *t*-test with latitude differences controlled. For the *t*-test, Cohen's *d* is determined by calculating the mean difference between two groups, and then dividing the result by the pooled standard deviation. The expression difference between different alleles of the focal *ADCY8* SNP was compared using a hypergeometric test.

31. Douglas, D. C., Weinzierl, R. C., Davidson, S., Kays, R., Wikelski, M. & Bohrer, G. Moderating Argos location errors in animal tracking data. *Methods in Ecology and Evolution* **3**, 999-1007 (2012).
32. Mueller, T., O'Hara, R. B., Converse, S. J., Urbanek, R. P. & Fagan, W. F. Social learning of migratory performance. *Science* **341**, 999-1002 (2013).
33. Trierweiler, C. et al. Migratory connectivity and population-specific migration routes in a long-distance migratory bird. *Proceedings of the Royal Society B: Biological Sciences* **281,** 20132897 (2014).
34. Ambrosini, R., Møller, A. P. & Saino, N. A quantitative measure of migratory connectivity. *Journal of Theoretical Biology* **257**, 203-211 (2009).
35. Baddeley, A., Rubak, E. & Turner. R. *Spatial Point Patterns: Methodology and Applications with R*. Chapman and Hall/CRC Press (2015).

36. López-López, D. P., García-Ripollés, C. & Urios, V. Individual repeatability in timing and spatial flexibility of migration routes of trans-Saharan migratory raptors. *Current Zoology* **60**, 642-652 (2014).

37. Benhamou, S. How to reliably estimate the tortuosity of an animal's path: straightness, sinuosity, or fractal dimension? *Journal of Theoretical Biology*, **229**, 209-220 (2004).

38. Stoffel, M. A., Nakagawa, S. & Schielzeth, H. rptR: Repeatability estimation and variance decomposition by generalized linear mixed-effects models. *Methods in Ecology and Evolution* **8**, 1639-1644 (2017).

39. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. New York, Routledge Academic (1988).

40. Ganusevich, S. A. et al. Autumn migration and wintering areas of peregrine falcons *Falco peregrinus* nesting on the Kola Peninsula, northern Russia. *Ibis* **146**, 291-297 (2004).

41. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120 (2014).

42. Zhao, S. et al. Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. *Nature Genetics* **45**, 67-71 (2013).

43. Damas, J. et al. Upgrading short-read animal genome assemblies to chromosome level using comparative genomics and a universal probe set. *Genome Research* **27**, 875-884 (2017).

44. Zhan, X. et al. Peregrine and saker falcon genome sequences provide insights into evolution of a predatory lifestyle. *Nature Genetics* **45,** 563-566 (2013).

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

46. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics* **43**, 491-498 (2011).

47. Rodriguez-Ramilo, S. T. & Wang, J. The effect of close relatives on unsupervised Bayesian clustering algorithms in population genetic structure analysis. *Molecular Ecology Resources* **12**, 873-884 (2012).

48. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* **81**, 559-575 (2007).

49. Schliep, K. P. phangorn: phylogenetic analysis in R. *Bioinformatics* **27**, 592-593 (2010).

50. Tang, H. et al. Genetic structure, self-identified race/ethnicity, and confounding in case-control association studies. *The American Journal of Human Genetics* **76**, 268-275 (2005).

51. Terhorst, J., Kamm, J. A. & Song, Y. S. Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature Genetics* **49**, 303-309 (2017).

52. Staab, P. R., Zhu, S., Metzler, D. & Lunter, G. scrm: Efficiently simulating long sequences using the approximated coalescent with recombination. *Bioinformatics* **31**, 1680-1682 (2015).

53. Pudlo, P., Marin, J. M., Estoup, A., Cornuet, J. M., Gautier, M., & Robert, C. P. Reliable ABC model choice via random forests. *Bioinformatics*, **32**, 859-866 (2016).

54. Csilléry, K., François, O. & Blum, M. G. abc: an R package for approximate Bayesian computation (ABC). *Methods in Ecology and Evolution* **3**, 475-479 (2012).

55. Hijmans, R. J., Phillips, S., Leathwick, J., Elith, J. & Hijmans, M. R. J. Package ‘dismo’. *Circles* **9**, 1-68 (2017).

56. Calenge, C. Home range estimation in R: the adehabitatHR package. *Office national de la classe et de la faune sauvage: Saint Benoist, Auffargis, France* (2011).

57. Fick, S. E. & Hijmans, R. J. WorldClim2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37**, 4302-4315 (2017).

58. Cao, X., Tian, F., Dallmeyer, A. & Herzschuh, U. Northern Hemisphere biome changes (> 30° N) since 40 cal ka BP and their driving factors inferred from model-data comparisons. *Quaternary Science Reviews* **220**, 291-309 (2019).

59. Borchers, H. W. & Borchers, M. H. W. R Package ‘pracma’. (2015).

60. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156-2158 (2011).

61. Beck, H. E. et al. Present and future Köppen-Geiger climate classification maps at 1-km resolution. *Scientific Data* **5**, 180214 (2018)

62. Sabeti, P. C. et al. Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913-918 (2007).

63. Beissinger, T. M., Rosa, G. J., Kaeppler, S. M., Gianola, D. & De Leon, N. Defining window-boundaries for genomic analyses using smoothing spline techniques. *Genetics Selection Evolution* **47**, 30 (2015).

64. Szpiech, Z. A. & Hernandez, R. D. selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Molecular Biology and Evolution* **31**, 2824-2827 (2014).

65. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *The American Journal of Human Genetics* **81**, 1084-1097 (2007).

66. Zheng, G. X., Lau, B. T., Levin, S. M., Jarosz, M., Bell, J. M., Hindson, C. M., Kyriazopoulou-Panagiotopoulou, S., Masquelier, D. A., Merrill, L., Terry, J. M. Haplotyping germline and cancer genomes with high-throughput linked-read sequencing. *Nature Biotechnology*, **34**, 303 (2016).

67. François, O., Martins, H., Caye, K. & Schoville, S. D. Controlling false discoveries in genome scans for selection. *Molecular Ecology*, **252**, 454-469 (2016).

68. Fariello, M. I., Boitard, S., Naya, H., SanCristobal, M. & Servin, B. (2013). Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics*, **193**, 929-941 (2013).

69. Bonhomme, M., Chevalet, C., Servin, B., Boitard, S., Abdallah, J., Blott, S. & SanCristobal, M. (2010). Detecting selection in population trees: the Lewontin and Krakauer test extended. *Genetics*, **186**, 241-262 (2010).

70. Frichot, E. & François, O. LEA: An R package for landscape and ecological association studies. *Methods in Ecology and Evolution*, **6,** 925-929 (2015).

71. Pan, S. et al. Population transcriptomes reveal synergistic responses of DNA polymorphism and RNA expression to extreme environments on the Qinghai-Tibetan Plateau in a predatory bird. *Molecular Ecology*, **26**, 2993-3010 (2017).

72. Zhang, Y. et al. An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *Journal of Neuroscience*, **34**, 11929-11947 (2014).

73. Yang, L. et al. TFBSshape: a motif database for DNA shape features of transcription factor binding sites. *Nucleic Acids Research* **42**, D148-D155 (2013).

74. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Current Protocols in Molecular Biology* **109**, 21-29 (2015).

75. Li, R. et al. *De novo* assembly of human genomes with massively parallel short read sequencing. *Genome Research* **20**, 265-272 (2010).

76. Barbato, M., Orozco-terWengel, P., Tapio, M. & Bruford, M. W. SNeP: a tool to estimate trends in recent effective population size trajectories using genome-wide SNP data. *Frontiers in Genetics* **6**, 109 (2015).

77. Pitt, D. et al. Demography and rapid local adaptation shape Creole cattle genome diversity in the tropics. *Evolutionary Applications* **12**, 105-122 (2019).

78. Carlzon, L., Karlsson, A., Falk, K., Liess, A. & Møller, S. Extreme weather affects peregrine falcon (*Falco peregrinus tundrius*) breeding success in South Greenland. *Ornis Hungarica* **26**, 38-50 (2018).

79. Franke, A. et al. Status and trends of circumpolar peregrine falcon and gyrfalcon populations. *Ambio* **49**, 762-783 (2020).

Acknowledgements

This study was supported by National Natural Science Foundation of China (No. 31821001, 31930013, 31911530186, 91740201), Strategic Priority Program of Chinese Academy of Sciences (XDB31000000), the National Key Programme of Research and Development, Ministry of Science and Technology (2016YFC0503200),Youth Innovation Promotion Association of Chinese Academy of Sciences (2020086) to S.P., Second Tibetan Plateau Scientific Expedition and Research Program (STEP) (No. 2019QZKK0501), Water Ecological Security Assessment, the Major Research Strategy for Middle and Lower Yangtze River (Grant No. ZDRW-ZS-2017-3), Biodiversity Survey, Monitoring and Assessment Project (2019-2023) of Ministry of Ecology and Environment, China, the Royal Society to X.Z. and M.W.B., and the CAS President's International Fellowship Initiative for Visiting Scientists to M.W.B. Funding for satellite tracking, sampling and partial genome resequencing was provided by the Environment Agency-Abu Dhabi. Additional tracking and sampling from Kolguev and Kola was funded by the Max Planck Institute of Animal Behavior to I.P. and Earthspan (www.earthspan.foundation). We thank Dr. J.M. Graves for her suggestions, M. Al Bowardi and M. Al Mansouri for their support, Drs. X. Li, Q. Dai, X. Liu, Y. Chen, J. Jiao, G. Wang and M. Barbato for their help in data analysis, W. Wu, X. Hou, Y. Wang and Y. Zhan for their advice on figure drawing, and O. Kulikova, V. Pozdnyakov, S. Troyev, L. Zhou, C. Dai, Y. Yin, C. Li and N. C. Fox during fieldwork.

Author contributions

X.Z led the project. X.Z. and A.D. conceived and designed the study. A.D., S.G., V.S., A.S., I.P., J.L. and Z.L. conducted the fieldwork and sample collection. X.Z. and A.D. examined migration paths, migration connectivity and genetic structure of peregrines across Eurasia. X.Z. and M.W.B. supervised the population genomic research. Z.G., S.P., L.H., J.C., and X.D. performed the data analyses. Z.L., Y.X., M.S., H.S. and F. J. conducted the molecular experiments. X.Z. and Z.G. wrote the manuscript with contributions from M.W.B., S.K. and A.D.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information is available in the online version of the paper.

Correspondence and requests for materials should be addressed to X.Z.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All of the sequenced genome data have been deposited in the GenBank under accession number PRJNA686418. The tracking data are included in the Arctic Animal Movement Archive and in Movebank under ID numbers 103426553 and 934079034.

Extended Data Figure Legends

Extended Data Fig. 1. Sampling sites for tracking peregrines in the Arctic. The sample size, visit years for each places and the peregrines equipped with Argos satellite transmitters are shown.

Extended Data Fig. 2. The broad-front migration pattern of peregrines. a, Four main wintering regions identified in the cluster analysis. **b,** Migration paths with the centroids of breeding and wintering MCP for each bird and the MCP of wintering ranges for all birds (dashed line) are shown. **c**, *G*-function results in the point pattern analysis showing a broad-front wintering distribution. The solid and dashed line mean the observed and theoretical value of *G*, respectively. The 95% CI of theoretical *G* value is shadowed. The *P* value was calculated for the statistic of maximum absolute deviation using Monte Carlo simulations (n = 100). **d**, The distance from each winter centroid to its nearest neighbour centroid (Nearest neighbour distance) is shown (n = 40).

Extended Data Fig. 3. The migration strategy comparisons between SD and LD. a, Variable importance estimated by random forest modelling. **b**, Migratory strategy comparisons between the short-distance (SD) and long-distance (LD) peregrine groups. Significance was determined by a two-sided *t*-test and sample size (n) for each comparison is shown. In the box plots, the center line represents the median, whiskers represent maximum and minimum values, and box boundaries represent 75th and 25th percentiles.

Extended Data Fig. 4. ABC simulation and parameter inference. a, Linear discriminant summary statistics values of the simulated datasets and the observations given four ABC candidate models. Based on the three statistics (LD1-3), Model-1 is best supported because the targets (dark) fit simulated data (shadow) well. **b**, Distribution of divergence times estimated using the chunks supporting the Model-1. One column represents one chunk and we only show 100 chunks. The density bar symbols posterior distribution of inferred divergence time in each chunk.

Extended Data Fig. 5 Maintenance mechanisms of present migration routes. a, Route cluster analysis based on *Hd*. **b**, x^2 testing results of climate zones between adjacent migration routes at the whole route level. **c**, The schematic diagram of environment comparisons between neighboring geographic bands. Each route was divided into geographic bands parallel to the main migration direction. Grids at regular intervals were chosen from neighboring bands for comparisons. **d**, Environmental boundaries coinciding with migration route boundaries. The Eurasia continent was divided into geographic bands (at 2° longitude). The *P* values of paired *t*-tests between compared bands are shown and the dashed line equals to 0.05 (Upper). The bar is scaled as the number of space between two targeted bands in a paired comparison. The MCPs (90%) of five migration routes are shaded (Lower). Arrows point the coincidence between environmental and migration route boundaries. Noted that the distinct environment difference within the Popigai route may result from the inclusion of large "barrier islands" of unsuitable region in comparison. **e**, Illustration of the model simulating the least-cost migration path. For a typical migration route, we simulated the potential migration path (dashed lines) that a peregrine depart from its actual breeding site (e.g. B1 in Route1), fly along a least-cost path, but winter in a wintering site of the neighboring route (e.g. W2 in Route2). B1-3 means breeding areas and W1-3 means wintering areas. Solid lines are the actual tracked migration path. **f**, Migration cost comparison between within-route and across-route paths ($P = 0.01$, $t = -2.58$, d$f = 101.68$), respectively. Significance was calculated using a two-sided *t*-test ($n = 45$ and 64 for within- and cross-route, respectively). In the box plots, the center line represents the median, whiskers represent maximum and minimum values, and box boundaries represent 75th and 25th percentiles.

Extended Data Fig. 6 Differences in breeding and wintering areas ($\triangle_{\text{Future-Present}}$) between present and future (2070). Predicted breeding (Upper) and wintering (Lower) area changes under RCP 8.5 scenario, and zoomed in Kola and Europe.