The Genomic Footprints of the Fall and Recovery of the Crested Ibis

Feng, Shaohong; Fang, Qi; Barnett, Ross; Li, Cai; Han, Sojung; Kuhlwilm, Martin; Zhou, Long; Pan, Hailin; Deng, Yuan; Chen, Guangji; Gamauf, Anita; Woog, Friederike; Prys-Jones, Robert; Marques-Bonet, Tomas; Gilbert, M. Thomas P.; Zhang, Guojie

Published in:
Current Biology

DOI:
10.1016/j.cub.2018.12.008

Publication date:
2019

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY-NC-ND

Citation for published version (APA):
The Genomic Footprints of the Fall and Recovery of the Crested Ibis

Highlights
- Recent population decline of crested ibis in 10 kya was likely shaped by human activity
- Modern group has lost almost half of genetic variations in pre-bottleneck ancestors
- Modern group suffers from a high inbreeding coefficient and deleterious mutation load
- Ancestral balancing selection is outweighed by genetic drift in the modern group

Authors
Shaohong Feng, Qi Fang, Ross Barnett, ..., Tomas Marques-Bonet, M. Thomas P. Gilbert, Guojie Zhang

Correspondence
guojie.zhang@bio.ku.dk

In Brief
Feng et al. use whole-genome sequencing of contemporary and historic crested ibis, an iconic endangered bird species, to explore how their genetic diversity has changed through time. Their analyses reveal the roles of genetic drift and intensive inbreeding on the loss of genetic diversity in today’s population.
SUMMARY

Human-induced environmental change and habitat fragmentation pose major threats to biodiversity and require active conservation efforts to mitigate their consequences. Genetic rescue through translocation and the introduction of variation into imperiled populations has been argued as a powerful means to preserve, or even increase, the genetic diversity and evolutionary potential of endangered species [1–4]. However, factors such as outbreeding depression [5, 6] and a reduction in available genetic diversity render the success of such approaches uncertain. An improved evaluation of the consequence of genetic restoration requires knowledge of temporal changes to genetic diversity before and after the advent of management programs. To provide such information, a growing number of studies have included small numbers of genomic loci extracted from historic and even ancient specimens [7, 8]. We extend this approach to its natural conclusion, by characterizing the complete genomic sequences of modern and historic population samples of the crested ibis (Nipponia nippon), an endangered bird that is perhaps the most successful example of how conservation effort has brought a species back from the brink of extinction. Though its once tiny population has today recovered to >2,000 individuals [9], this process was accompanied by almost half of ancestral loss of genetic variation and high deleterious mutation load. We furthermore show how genetic drift coupled to inbreeding following the population bottleneck has largely purged the ancient polymorphisms from the current population. In conclusion, we demonstrate the unique promise of exploiting genomic information held within museum samples for conservation and ecological research.

RESULTS

Historical and Contemporary Population Structure of the Crested Ibis

We collected 57 historic crested ibis samples dating to between 1841 and 1922. The geographic origin of these samples broadly covers the historical distribution of the species, including east China, northwest China, northeast China, Korea, Japan, and Russia (Figure 1A). We initially generated an average of ~179 million (M) paired-end short sequencing reads for historic samples (49 base pairs [bp] for most samples and 100 bp for two; Data S1). After filtering to remove contaminants and other sequencing artifacts, we obtained final genomic coverage at ca. 7-fold for each sample. Analysis with mapDamage2.0 revealed signs of cytosine deamination at the ends of the sequences, which are normally expected
from the historic DNA sequences \( \text{Data S1} \). We combined these data with previously published sequencing data for eight contemporary individuals \[10\] and used ANGSD v0.615 \[11\] to identify 5,268,206 SNPs in total. On average, each historical sample contained 1,088,820 (±112,638) heterozygous sites, while each contemporary sample contained 788,162 (±35,747) heterozygous sites.

After excluding linked SNP loci that could potentially bias clustering results (see STAR Methods), we built a neighbor-joining (NJ) tree \[12\] using all 65 samples (Figure 1B). The NJ tree assigned the historical samples to three major groups, consistent with their original geographic provenance. A northeast Asian group (NE) exhibits the most diversity and can be further divided into three subgroups (NE-1, NE-2, and NE-3) according to their genetic distances. The remaining samples fall into two groups: east China (EA) and northwest China (NW). All contemporary samples (MC) are from northwest China and form a monophyletic group. These clustering results were also supported by principal component analysis (PCA) \[13\] (Figure 1C).

We ran STRUCTURE v2.3.4 \[14\] to explore the genetic composition of each group and subgroup after initially removing potential bias caused by the missing loci (Figure 1D; Data S2). Sample clusters were evaluated using the \( \Delta K \) statistic to determine the number of clusters (Figure 1D). The population origin of each individual is indicated on x axis. Each individual is represented by a bar that is segmented into colors based on the ancestry proportions given the assumption of \( K \) populations.

See also Data S1 and S2.

### Population Size Dynamics of the Crested Ibis

We used the historic genomic data to confirm the decline of the effective population size (\( \text{Ne} \)) reported in a previous study \[10\] (Figure S1). Because the power of PSMC (pairwise sequentially Markovian coalescent) to reconstruct demography is poor over recent history due to the limited number of recombination events \[16\], we used an alternate method based on approximate Bayesian computation, PopSizeABC v2.1 \[17\], to better characterize the \( \text{Ne} \) dynamics over the last 10,000 years, during which the climate has been relatively stable \[18\]. We found that all
groups declined at different rates over the last 10,000 years, and such decline was followed by a short period of relative population stability before the near extinction of the crested ibis in the 20th century (Figure 2A). The recent demographic pattern implies that the impact of human activity on the population decline may have already commenced up to 600 years ago (the starting point of decline in the NW group).

Environmental suitability is an important factor that influences population fluctuation. Therefore, we used ecological niche models (ENMs) [19] to reconstruct the suitable niche for crested ibis during the contemporary period (19th and 20th centuries), the last glacial maximum (LGM, approximately 22 kya), and the last interglacial (LIG, approximately 120-140 kya) periods. During the LIG period, the potential breeding range of the crested ibis covered broad areas across south China, with extensive hotspots of habitat (Figure 2D). During the LGM, their potential breeding range showed extensive contraction, corresponding to the climate change during this period [20] (Figure 2C). Today, our simulations predict that the suitable breeding range would largely resemble that for the LGM, with multiple distribution hotspots with suitable probability higher than 50% (Figure 2B). Nevertheless, our data shows that Ne continued to decline in the contemporary period (Table S1), consistent with the fact that the historical groups went extinct across most of the breeding range during the 19th and 20th centuries [9]. The reconstructed suitable niche pattern implies that the population collapse during the late-19th and early-20th centuries was more likely a result of human activity rather than climatic factors. This is consistent with previous analyses on the availability of habitat suitable for crested ibis, which demonstrated how radical land-cover changes (e.g., through urbanization) have impacted their population dynamics.

Figure 2. Inference of Recent Population Size Tendency and Prediction of Breeding Ranges for the Crested Ibis at the Contemporary, the Last Glacial Maximum, and the Last Interglacial Periods

(A) The recent effective population size (Ne) for each group is inferred using PopSizeABC. A 90% confidence interval is indicated by dotted lines. (B–D) Reconstruction of ecological niche models for the contemporary (19th and 20th centuries), B, last glacial maximum (LGM, approximately 22 kya), C and the last interglacial (LIG, approximately 120–140 kya, D) periods. Pictures show the point-wise mean of 10 replicates using 19 environmental layers. The colors indicate the predicted probability that conditions are suitable, with red indicating high and blue indicating low probabilities. The white square points in (B) indicate the geographic coordinates of the crested ibis samples.

See also Figure S1 and Table S1.
development, increased cultivation of land, and increased crop diversification) between the 1970s and 2000s reduced the swamp environments that used to be their feeding areas [21, 22].

Severe Genetic Diversity Loss in the Contemporary Population Caused by Inbreeding
Although the size of the contemporary population has been steadily increasing owing to recent conservation efforts, its nucleotide diversity (\( \pi \)) is considerably lower than that of the historic population (mean value is 0.112, while mean value of the historic samples is 0.208; Figure 3A). We also found accelerated fixation of polymorphisms in the contemporary population, with ca. 70% of the SNPs that were polymorphic in the historical groups having been fixed in the contemporary group (see STAR Methods). We further identified the identical by descent (IBD) regions, which means the alleles derive from the ancestor and maintain in descendant populations [23]. We obtained a total of 444.5 megabases (Mb) merged excess IBD regions in the contemporary group (Figure 3B), 31.8 Mb for the NW, 31.0 Mb for the NE, and 0 Mb for the EA groups.

Such dramatic loss of genetic diversity could be caused by the unavoidable effects of inbreeding. To assess the extent of inbreeding, we estimated \( F_{\text{IBD}} \) based on the correlation between uniting gametes following Wright [24]. A lower (i.e., negative) value indicates an excess of heterozygosity and little inbreeding effect, while a positive value indicates an excess of homozygosity and strong inbreeding effect. We subsequently calculated \( F_{\text{IBD}} \) and \( N_e \) for 14 animal species spanning different conservation statuses. A regression correlation including the crested ibis and other species clearly shows that the contemporary population is a significant outlier, highlighting the severity of its inbreeding (Figure 3C; Table S2; see STAR Methods). This finding is consistent with the fact that the gene pool of all contemporary individuals was inherited from only two breeding pairs.

Patterns of Deleterious Mutations across Populations Support \( N_e \) Collapse, Bottlenecks, and Inbreeding in the Contemporary Population
Recent conservation genetic studies have demonstrated that small populations are susceptible to allelic drift, leading to allele loss or fixation, which can be evaluated using the site-frequency-spectrum (SFS) [25]. In our analysis, all historic populations showed a general pattern in which most mutations occurred at low frequencies, while few mutations segregated at high frequencies (Figures S2A–S2C). However, in the contemporary population, we observed a reduced number of low-frequency mutations and a relatively higher fraction of mutations that segregate at medium or higher frequencies (Figure S2D). We hypothesize that these mutations occur at low rates and that relaxed purifying selection has probably led to random fixation, likely due to the high level of inbreeding in the contemporary population. An alternative possibility might be that the contemporary individuals sampled were from a sub-structured population; however, we find no such evidence in the results of the admixture analysis to support this hypothesis.

We further used the Grantham score [26] as an index to identify potential deleteriousness of missense mutations in the coding regions of historic and contemporary populations. The proportional excess of homozygous deleterious-derived mutations could be likely due to the varying efficacy of purifying selection, resulting from differences in the demographic histories of the populations [27, 28]. We found that the ratio of homozygous sites to homo- and heterozygous sites in the contemporary population is significantly higher than that in the historic populations (0.328 ± 0.041 in contemporary group, 0.207 ± 0.039 in historical group, Welch two-sample t test, 2.524e–05) (Figure S2E). A similar difference could also be seen when using loss of function (LOF) as an index (Figure S2F; see STAR Methods). Despite only ~100 years difference between the collection dates of historic and contemporary samples, the number of homozygous deleterious mutations has doubled in the contemporary population, showing the severity of the bottleneck.

Role of Natural Selection in Shaping the Genetic Diversity of the Crested Ibis
The effect of selection is considered to be limited in contrast to that of genetic drift in populations with a small \( N_e \) [29]. Immune genes are believed to be under pathogen-mediated balancing selection, thus maintaining high polymorphism levels in wild populations [30, 31]. However, we found that the genetic diversity of the major histocompatibility complex (MHC) has experienced drastic haplotype loss in the contemporary group. The overall allelic richness (AR) of the entire MHC region in the contemporary group is significantly lower than that shown in the historical groups (mean AR of the contemporary group is 2.44, mean AR of the historical group is 2.92. Welch two-sample t test, \( p < 2.2e^{-16} \); see STAR Methods). Furthermore, haplotype blocks in the contemporary group were much longer and exhibited lower diversity (Figures 4 and S3). This dramatically reduced diversity of the MHC in the contemporary population may confer a high risk of increasing the burden of transmissible pathogens when captive animals are released into the wild [32].

We subsequently performed a broader scan across the coding region to identify genomic loci that were previously under balancing selection but have been fixed in the current population (see STAR Methods). In total we found 53 genes with high Tajima’s D values in at least two of the historic groups, implying that they were once under balancing selection. Of these, nine genes exhibit a negative Tajima’s D value in the contemporary group that might have been substantially affected by bottleneck effects during the breeding program. Several of these genes are involved in processes relating to immunity, the nervous system, and reproduction. For example, polymorphism has been completely lost in half of the SNP loci in gene KPNA1, a member of the importin or karyopherin alpha family that has been reported to be involved in the development and/or repair processes of the nervous system [33].

DISCUSSION
Thanks to the development of paleogenomic sequencing techniques and the recovery of genetic data from museum samples collected up to 100 years before the introduction of the modern management program, our study has reconstructed a much-improved demographic history of the crested ibis over the last 10,000 years. The continued population decline of this species in this period therefore not only leads to the extreme bottleneck...
that started in the late-19th century but also provides new insights into the negative effect of human activities. Given that crested ibis did not appear to favor colder climates, the warmer climate of the past 10,000 years should have stopped the prior decline and incurred a recovery—something we see no evidence of. One explanation for this discrepancy might be that the...
Figure 4. Haplotype Structure and Length Distribution of the Class I Region in Historical and Contemporary Groups

(A) Haplotype blocks are detected based on the SNP datasets for historical and contemporary groups, respectively. Colored circles show the haplotype structure (each box indicates the start and end SNPs of one haplotype) and the diversity along the class I region (color indicates the allelic richness). The allelic richness of

(legend continued on next page)
crested ibis had entered into an evolutionary dead end from which they were unable recover, because of their small Ne and the low genetic diversity. However, given the success of the current management program in increasing their population size to ca. 2,600 individuals from only seven founding individuals, it is clear that the species still maintains a robust reproductive rate. An alternate explanation for their population decline might therefore relate to human interference. Human activities, like over-hunting and development of farm land, might have led to the fragmentation of niches for the crested ibis and ultimately increased the inbreeding coefficient within the isolated subpopulations. Consistently, recent research on crested ibis notes how high human-induced mortality risks play a significantly negative effect on the population growth and range expansion of the surviving population [34].

One factor that contributed to the crested ibis’s recent demography is the effect of inbreeding. There is little doubt that inbreeding reduces the fitness of a population [35–37], particularly for critically endangered species with small population sizes. However, previous studies that have argued that inbreeding depression can elevate extinction risk are based on simulations [38], and thus there is a need for actual evidence with which to verify whether the consequences of inbreeding, such as the accumulation of deleterious alleles, will lead to and even accelerate the loss of diversity. Our findings indicate that the contemporary crested ibis population has dramatically reduced the ancestral genetic diversity, undergone higher allelic fixation, and carried an elevated load of homozygous deleterious mutations that associated with a raised inbreeding coefficient. Importantly, the inclusion of samples spanning 1841–2011 CE enabled us to accurately quantify the extent to which the most recent demographic decline has led to a loss in their genetic variation as well as increased inbreeding levels and genetic load, all of which are regarded as useful genomic parameters for endangered species [39]. Our analyses showed how the overall nucleotide diversity of the contemporary group is only 53.85% of that which was present ca. 100 years ago. This has been accompanied by an almost 4-fold elevation in their inbreeding coefficient and a doubling of their homozygous deleterious mutation load.

The conservation program has had remarkable success in increasing the size of the crested ibis population, from the seven individuals during their near extinction to currently ca. 2,600 individuals. Nevertheless, the contemporary population suffers from dramatic diversity loss across the genome. This overall reduced genetic diversity represents a risk when they are introduced to the natural environment as part of the conservation program. We acknowledge, however that it may be, that the protected environment of the management program will help them to overcome the potential harmful effects from the reduced diversity, thus enabling them to survive into the future with low genetic diversity, as for example has been the case for the endangered island fox over the past millennia [40]. In that example, it is speculated that their isolated island habitat might provide a friendly environment, given its lack of competitors and predators, thus to some extent reducing the negative effect of genetic load.

Ultimately, recent successful re-introduction of the crested ibis into the wild has provided an important step forward in this regard [41]. Long-term monitoring of their genetic diversity in the wild will provide a crucial index for their long-term persistence and management. Recently, efforts have been commenced to establish a DNA identification profiling (DIP) platform using short tandem repeat markers to help avoid close crosses [42]. Given that our analyses identified several candidate genes that no longer exhibit high polymorphism, we suggest it may be useful to incorporate them into the existing DIP for monitoring and preserving genomic diversity. Lastly, our study highlights how integrating genomic information from the modern and historic samples can help to reveal the evolutionary history of endangered species and in doing so provide guiding information of benefit to future conservation programs.

### STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **CONTACT FOR RESOURCE SHARING**
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
- **METHOD DETAILS**
  - Whole-genome sequencing of museum samples
  - Read alignment and SNP calling
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - Data pre-processing
  - Population structure
  - Population demographic analysis
  - Ecological niche models analysis
  - Temporal population polymorphism
  - Correlation between inbreeding coefficient and effective population size (Ne)
  - Patterns of deleterious mutations across populations
  - MHC diversity analysis
  - Natural selection analysis
- **DATA AND SOFTWARE AVAILABILITY**

### SUPPLEMENTAL INFORMATION

Supplemental Information includes three figures, two tables, and two data files and can be found with this article online at https://doi.org/10.1016/j.cub.2018.12.008.

### ACKNOWLEDGMENTS

We are grateful to the National Museum of Natural History (USNM) (Gary Graves, Chris Milensky); Senckenberg Museum of Natural History (Gerald Mayr); Natural History Museum, Tring (Hein van Grouw); Department of Zoology, National Museums Liverpool (Tony Parker, Clemency Fisher); Center for Natural History/Zoological Museum Hamburg (Alexander Has); Naturalis...
Biodiversity Center Leiden, the Netherlands (Pepijn Kamminga); NRM Stockholm (Ulf Johansson); and Natural History Museum of Denmark (Jon Fjeldså) for providing access to museum specimens and Nigel Collar for providing background information critical to locating the samples. This work was supported by the Strategic Priority Research Program of Chinese Academy of Sciences (XDB31020000, XDB13000000), Carlsberg Foundation grant to G.Z. (CF16-0663), ERC Consolidator Grant 681396 ‘Extinction Genomics’ (M.T.P.G.), BFU2017-86471-P (MINECO/FEDER, UE) (T.M.-B.), U01 MH106874 grant (T.M.-B.), Howard Hughes International Early Career (T.M.-B.), Obra Social “La Caixa” and Secretaria d’Universitats i Recerca (T.M.-B.), ERC Research Programme of Department d’Economia i Coneixement de la Generalitat de Catalunya (T.M.-B.), and Deutsche Forschungsgemeinschaft (DFG) fellowship (KU 3467/1-1) (M.K.).

AUTHOR CONTRIBUTIONS
C.L., M.T.P.G. and G.Z. conceived the study; M.T.P.G. collected museum samples, which were prepared into sequencing libraries by R.B.; C.L. managed sample sequencing; C.L., Y.D., L.Z., and H.P. developed, tested, and implemented the detailed SNPs calling pipeline; S.F., Q.F., C.L., and G.C. carried out genetic analyses described in the manuscript; S.H., M.K. and T.M.-B. assisted with the mutation load analysis; S.F., Q.F., C.L., T.M.-B., M.T.P.G. and G.Z. wrote and edited the manuscript; All authors read and approved the final manuscript.

DECLARATION OF INTERESTS
The authors declare no competing interests.
Received: July 31, 2018
Revised: October 9, 2018
Accepted: December 6, 2018
Published: January 10, 2019

REFERENCES


# STAR METHODS

## KEY RESOURCES TABLE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Biological Samples</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27 <em>Nipponia nippon</em> tissue samples</td>
<td>American Museum of Natural History</td>
<td>Data S1</td>
</tr>
<tr>
<td>19 <em>Nipponia nippon</em> tissue samples</td>
<td>Natural History Museum (London)</td>
<td>Data S1</td>
</tr>
<tr>
<td>2 <em>Nipponia nippon</em> tissue samples</td>
<td>World Museum, Liverpool</td>
<td>Data S1</td>
</tr>
<tr>
<td>1 <em>Nipponia nippon</em> tissue sample</td>
<td>Naturalis Biodiversity Center Leiden, the Netherlands</td>
<td>Data S1</td>
</tr>
<tr>
<td>1 <em>Nipponia nippon</em> tissue sample</td>
<td>Senckenberg Natural History Museum</td>
<td>Data S1</td>
</tr>
<tr>
<td>2 <em>Nipponia nippon</em> tissue samples</td>
<td>Stuttgart State Museum of Natural History</td>
<td>Data S1</td>
</tr>
<tr>
<td>2 <em>Nipponia nippon</em> tissue samples</td>
<td>Smithsonian Institute National Museum of Natural History</td>
<td>Data S1</td>
</tr>
<tr>
<td>2 <em>Nipponia nippon</em> tissue samples</td>
<td>Natural History Museum Vienna</td>
<td>Data S1</td>
</tr>
<tr>
<td>1 <em>Nipponia nippon</em> tissue sample</td>
<td>Natural History Museum of Denmark</td>
<td>Data S1</td>
</tr>
<tr>
<td><strong>Deposited Data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Whole-genome sequencing data of 57 historical <em>Nipponia nippon</em> samples</td>
<td>This paper</td>
<td>NCBI Project ID: PRJNA481495; CNSA: CNP0000084</td>
</tr>
<tr>
<td>Whole-genome sequencing data of 8 contemporary <em>Nipponia nippon</em> samples</td>
<td>[10]</td>
<td>NCBI Project ID: PRJNA232572</td>
</tr>
<tr>
<td><em>Nipponia nippon</em> reference genome</td>
<td>[10]</td>
<td>NCBI Project ID: PRJNA232572</td>
</tr>
<tr>
<td>Whole-genome sequencing data of 13 <em>Pan paniscus</em></td>
<td>[43]</td>
<td>SRA: SRP018689</td>
</tr>
<tr>
<td>Whole-genome sequencing data of 5 <em>Pongo abelii</em></td>
<td>[43]</td>
<td>SRA: SRP018689</td>
</tr>
<tr>
<td>Whole-genome sequencing data of 5 <em>Pongo pygmaeus</em></td>
<td>[43]</td>
<td>SRA: SRP018689</td>
</tr>
<tr>
<td>Whole-genome sequencing data of 25 <em>Pan troglodytes</em></td>
<td>[43]</td>
<td>SRA: SRP018689</td>
</tr>
<tr>
<td>Whole-genome sequencing data of 31 <em>Gorilla gorilla</em></td>
<td>[43]</td>
<td>SRA: SRP018689</td>
</tr>
<tr>
<td>Whole-genome sequencing data of 10 <em>Vulpes vulpes</em></td>
<td>[44]</td>
<td>SRA: SRP100625</td>
</tr>
<tr>
<td>Whole-genome sequencing data of 18 <em>Ursus maritimus</em></td>
<td>[45]</td>
<td>SRA: SRA092289</td>
</tr>
<tr>
<td>Whole-genome sequencing data of 10 <em>Ursus arctos</em></td>
<td>[45]</td>
<td>SRA: SRA092289</td>
</tr>
<tr>
<td>Whole-genome sequencing data of 14 <em>Anas platyrhynchos</em></td>
<td>[46]</td>
<td>SRA: SRP125660</td>
</tr>
<tr>
<td>Whole-genome sequencing data of 6 Tibetan fowl from Qinghai</td>
<td>[47]</td>
<td>SRA: SRP040477</td>
</tr>
<tr>
<td>Whole-genome sequencing data of 5 <em>Platyspiza crassirostris</em></td>
<td>[48]</td>
<td>SRA: SRP048643</td>
</tr>
<tr>
<td>Whole-genome sequencing data of 20 <em>Ficedula albicollis</em></td>
<td>[49]</td>
<td>SRA: ERA362744, ERA362838, ERA391986, ERA391991</td>
</tr>
<tr>
<td>Whole-genome sequencing data of 34 <em>Ailuropoda melanoleuca</em></td>
<td>[50]</td>
<td>SRA: SRP013618</td>
</tr>
<tr>
<td>Whole-genome sequencing data of 6 <em>Ectopistes migratorius</em></td>
<td>[51]</td>
<td>SRA: SRP042357, SRP103042</td>
</tr>
<tr>
<td>MHC class 1 region sequences of <em>Nipponia nippon</em></td>
<td>NCBI</td>
<td>KP182409.1</td>
</tr>
<tr>
<td>MHC class 2 region sequences of <em>Nipponia nippon</em></td>
<td>NCBI</td>
<td>KP182408.1</td>
</tr>
</tbody>
</table>

(Continued on next page)
<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>MHC extended region sequences of <em>Nipponia nippon</em></td>
<td>NCBI</td>
<td>KP182407.1</td>
</tr>
<tr>
<td>Locality data of crested ibis (1842.1.19.90) from Natural History Museum. Natural History Museum (London) Collection Specimens (published on 2015-10-10)</td>
<td>VertNet database</td>
<td><a href="http://data.nhm.ac.uk/object/7e36f664-1cf0-4f46-b30d-aa0715a5a53">http://data.nhm.ac.uk/object/7e36f664-1cf0-4f46-b30d-aa0715a5a53</a></td>
</tr>
<tr>
<td>Locality data of crested ibis (1852.2.5.12) from Natural History Museum. Natural History Museum (London) Collection Specimens (published on 2015-10-10)</td>
<td>VertNet database</td>
<td><a href="http://data.nhm.ac.uk/object/5a4a8a93-ead3-42ce-9229-a8d34a8d0a0e">http://data.nhm.ac.uk/object/5a4a8a93-ead3-42ce-9229-a8d34a8d0a0e</a></td>
</tr>
<tr>
<td>Locality data of crested ibis (1891.10.19.20) from Natural History Museum. Natural History Museum (London) Collection Specimens (published on 2015-10-10)</td>
<td>VertNet database</td>
<td><a href="http://data.nhm.ac.uk/object/b1a0a925-5132-40c7-a9c3-2a999d336660">http://data.nhm.ac.uk/object/b1a0a925-5132-40c7-a9c3-2a999d336660</a></td>
</tr>
<tr>
<td>Locality data of crested ibis (1892.4.2.492) from Natural History Museum. Natural History Museum (London) Collection Specimens (published on 2015-10-10)</td>
<td>VertNet database</td>
<td><a href="http://data.nhm.ac.uk/object/1a589acbc-266c-48c0-9365-6a7620d12e5d">http://data.nhm.ac.uk/object/1a589acbc-266c-48c0-9365-6a7620d12e5d</a></td>
</tr>
<tr>
<td>Locality data of crested ibis (1892.4.2.493) from Natural History Museum. Natural History Museum (London) Collection Specimens (published on 2015-10-10)</td>
<td>VertNet database</td>
<td><a href="http://data.nhm.ac.uk/object/7be93e6c-f56d-4922-95ce-31635a63c30">http://data.nhm.ac.uk/object/7be93e6c-f56d-4922-95ce-31635a63c30</a></td>
</tr>
<tr>
<td>Locality data of crested ibis (1897.10.30.2) from Natural History Museum. Natural History Museum (London) Collection Specimens (published on 2015-10-10)</td>
<td>VertNet database</td>
<td><a href="http://data.nhm.ac.uk/object/f7ca4de8-0772-4743-9c82-59f562631504d">http://data.nhm.ac.uk/object/f7ca4de8-0772-4743-9c82-59f562631504d</a></td>
</tr>
<tr>
<td>Locality data of crested ibis (1897.10.30.3) from Natural History Museum. Natural History Museum (London) Collection Specimens (published on 2015-10-10)</td>
<td>VertNet database</td>
<td><a href="http://data.nhm.ac.uk/object/f9b28058-f974-40c1-9300-0248271c1f28">http://data.nhm.ac.uk/object/f9b28058-f974-40c1-9300-0248271c1f28</a></td>
</tr>
<tr>
<td>Locality data of crested ibis (1897.10.30.4) from Natural History Museum. Natural History Museum (London) Collection Specimens (published on 2015-10-10)</td>
<td>VertNet database</td>
<td><a href="http://data.nhm.ac.uk/object/927fbd79d-4100-49a0-894a-3a8cfa644e1f">http://data.nhm.ac.uk/object/927fbd79d-4100-49a0-894a-3a8cfa644e1f</a></td>
</tr>
<tr>
<td>Locality data of crested ibis (1897.10.30.5) from Natural History Museum. Natural History Museum (London) Collection Specimens (published on 2015-10-10)</td>
<td>VertNet database</td>
<td><a href="http://data.nhm.ac.uk/object/218dc4cd-6aa4-4349-95b9-9bd253b1e036">http://data.nhm.ac.uk/object/218dc4cd-6aa4-4349-95b9-9bd253b1e036</a></td>
</tr>
<tr>
<td>Locality data of crested ibis (1897.10.30.6) from Natural History Museum. Natural History Museum (London) Collection Specimens (published on 2015-10-10)</td>
<td>VertNet database</td>
<td><a href="http://data.nhm.ac.uk/object/69828014-8f2c-4e67-8fc2-1ac1e623b63">http://data.nhm.ac.uk/object/69828014-8f2c-4e67-8fc2-1ac1e623b63</a></td>
</tr>
<tr>
<td>Locality data of crested ibis (1900.9.9.12) from Natural History Museum. Natural History Museum (London) Collection Specimens (published on 2015-10-10)</td>
<td>VertNet database</td>
<td><a href="http://data.nhm.ac.uk/object/e04fe7fe-46c3-43e2-89d4-cb106558181f">http://data.nhm.ac.uk/object/e04fe7fe-46c3-43e2-89d4-cb106558181f</a></td>
</tr>
<tr>
<td>Locality data of crested ibis (1900.9.9.13) from Natural History Museum. Natural History Museum (London) Collection Specimens (published on 2015-10-10)</td>
<td>VertNet database</td>
<td><a href="http://data.nhm.ac.uk/object/1e5fed44-06b0-4cf4-95b6-e1d3902947f3">http://data.nhm.ac.uk/object/1e5fed44-06b0-4cf4-95b6-e1d3902947f3</a></td>
</tr>
<tr>
<td>Locality data of crested ibis (1908.1.5.20) from Natural History Museum. Natural History Museum (London) Collection Specimens (published on 2015-10-10)</td>
<td>VertNet database</td>
<td><a href="http://data.nhm.ac.uk/object/3ff29518-a451-416a-9e56-4299299a304f">http://data.nhm.ac.uk/object/3ff29518-a451-416a-9e56-4299299a304f</a></td>
</tr>
</tbody>
</table>

(Continued on next page)
<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Locality data of crested ibis (1908.1.5.21) from Natural History Museum. Natural History Museum (London) Collection Specimens (published on 2015-10-10)</td>
<td>VertNet database</td>
<td><a href="http://data.nhm.ac.uk/object/1a3f8ba6-4d7f-45fe-b165-07d1e62607">http://data.nhm.ac.uk/object/1a3f8ba6-4d7f-45fe-b165-07d1e62607</a></td>
</tr>
<tr>
<td>Locality data of crested ibis (1908.1.5.22) from Natural History Museum. Natural History Museum (London) Collection Specimens (published on 2015-10-10)</td>
<td>VertNet database</td>
<td><a href="http://data.nhm.ac.uk/object/786144ad-774e-49d8-9ed9-b637c8b60cfe">http://data.nhm.ac.uk/object/786144ad-774e-49d8-9ed9-b637c8b60cfe</a></td>
</tr>
<tr>
<td>Locality data of crested ibis (1908.1.5.23) from Natural History Museum. Natural History Museum (London) Collection Specimens (published on 2015-10-10)</td>
<td>VertNet database</td>
<td><a href="http://data.nhm.ac.uk/object/7a9df17a-c7d3-4d73-8dd3-7c2c3ee6">http://data.nhm.ac.uk/object/7a9df17a-c7d3-4d73-8dd3-7c2c3ee6</a></td>
</tr>
<tr>
<td>Locality data of crested ibis (1908.1.5.24) from Natural History Museum. Natural History Museum (London) Collection Specimens (published on 2015-10-10)</td>
<td>VertNet database</td>
<td><a href="http://data.nhm.ac.uk/object/365367ee-fd32-4981-a62d-853df6b0a35">http://data.nhm.ac.uk/object/365367ee-fd32-4981-a62d-853df6b0a35</a></td>
</tr>
<tr>
<td>Locality data of crested ibis (1908.1.5.25) from Natural History Museum. Natural History Museum (London) Collection Specimens (published on 2015-10-10)</td>
<td>VertNet database</td>
<td><a href="http://data.nhm.ac.uk/object/35a1776c-756b-4876-b800-753f6b569ec">http://data.nhm.ac.uk/object/35a1776c-756b-4876-b800-753f6b569ec</a></td>
</tr>
<tr>
<td>Locality data of crested ibis (1908.1.5.26) from Natural History Museum. Natural History Museum (London) Collection Specimens (published on 2015-10-10)</td>
<td>VertNet database</td>
<td><a href="http://data.nhm.ac.uk/object/36508-9ca0-2e10ab18d3">http://data.nhm.ac.uk/object/36508-9ca0-2e10ab18d3</a></td>
</tr>
<tr>
<td>Locality data of crested ibis (1912.9.23.14) from Natural History Museum. Natural History Museum (London) Collection Specimens (published on 2015-10-10)</td>
<td>VertNet database</td>
<td><a href="http://data.nhm.ac.uk/object/6c07508-ceae-4688-880c-3d00855aaf">http://data.nhm.ac.uk/object/6c07508-ceae-4688-880c-3d00855aaf</a></td>
</tr>
</tbody>
</table>


Software and Algorithms

- PALEOMIX [53] https://geogenetics.ku.dk/publications/paleomix; RRID: SCR_015057
- Picard v1.92 N/A http://broadinstitute.github.io/picard/; RRID:SCR_006525
- RapidNJ v2.3.2 [54] http://birc.au.dk/software/rapidnj/
- STRUCTURE v2.3.4 [14] https://web.stanford.edu/group/pritchardiab/structure.html; RRID:SCR_002151
- Maxent v3.4.0 [55] https://github.com/mrmaxent/Maxent
- GRASS GIS v7.2.1 N/A https://grass.osgeo.org/grass7/
- R v3.2.2 [56] https://www.r-project.org

(Continued on next page)
CONTACT FOR RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Guojie Zhang (guojie.zhang@bio.ku.dk).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The current study compares whole genome sequencing data generated from 57 historic crested ibis (22 samples from Northeast Asia, 11 samples from East China, and 24 from Northeast China) against 8 previously published contemporary genomes, to characterize the temporal changes of genetic diversity for this iconic endangered species. All known information on the context and sequencing information of the samples is provided in Data S1. All researches are under the oversight of the institutional review board.

METHOD DETAILS

Whole-genome sequencing of museum samples
DNA was extracted at the Centre for GeoGenetics from preserved historic skin and toepad samples from 57 crested ibis. Tissue was macerated in extraction buffer and digested overnight according to a modified silica extraction protocol of Rohland [65]. Briefly, buffer (pH8.0, 0.5M EDTA, 1% SDS, 50mM DTT, 1mg/μl, ProteinaseK) was used with samples and negative controls (in a ratio of 1:16) for overnight digestion at 55°C. Samples and controls were then spun down and the supernatant was transferred to filter columns (Amicon, 30KDa cutoff). After concentration to sub 1ml volumes, concentrates were passed through MinElute columns (QIAGEN, Venlo, Netherlands) according to the manufacturer’s protocol, and finally eluted in 60μl EB buffer. This DNA was converted into Illumina sequencing libraries through blunt-end ligation using the NEB E6070 kit and the slightly modified version of the protocol of Vilstrup et al. [66], reported by Paijmans et al. [67]. Each DNA tissue extract was used to create one library. Each unique library included individual barcodes and were sequenced by Illumina HiSeq2000 using a TruSeq SBS sequencing kit version 3 (Illumina®, San Diego, CA) for 100 cycles from each end of the fragments. Reads were analyzed with Casava1.8.2 (Illumina, San Diego, CA). In summary, we obtained 10–28 Gb of sequencing for the 57 historical samples (Data S1).

Read alignment and SNP calling
We used the identical bioinformatical procedures for the raw reads of all 65 study individuals, including 57 museum individuals in this study and 8 contemporary samples obtained from the previous study (PRJNA308878) [10] (Data S1). All raw sequencing reads were put into PALEOMIX [53] for the adaptor removal, the crested ibis reference genome mapping (PRJNA308878) [10], duplicate filtering, initial BAM rescaled, etc. In order to improve the quality of the initial alignments, the PALEOMIX pipeline incorporates several useful steps: Picard tools for filtering, mapDamage2.0 for rescaling (only applied for the museum samples), and local re-alignment. After that, the mean sequencing depth estimated from the BAM files, ranged from 3.83-10.13x (museum samples) to 20.05-26.71x (contemporary samples), with an average depth of 7x over museum individuals and 21.64x over contemporary individuals (Data S1). Further to avoid any bias in SNP calling due to the different sequencing depths, we randomly extracted data to a coverage of only 7x from the original BAM files of the contemporary samples.
We identified SNPs using Analysis of Next Generation Sequencing Data (ANGSD) v0.615 [11] based on a BAM list containing 57 BAM files of the museum samples and 8 extracted BAM files of the contemporary samples with the following parameters:

\[ \text{doMajorMinor 1 -GL 2 -doMaf 2 -doGeno 7 -SNP_pval 1e-6 -doPost 1 -doCounts 1 -doglf 4 -dumpCounts 4 -HWE_pval 1 -minMaf 0.01} \]

Considering the low depth of sequencing data caused by the features of museum samples, we further added two steps to reduce the false positives:

1. ANGSD provides the genotypes for each individual. We removed the singleton SNPs.
2. If the genotype of one individual is assigned as major-major, we marked its genotype as NN when the depth of its major allele is less than 2. If genotype of one individual is assigned as minor-minor, we will mark its genotype to NN when the depth of its minor allele is less than 2. If the genotype of one individual is assigned as major/minor, we will mark its genotype to NN when neither the depth of its major nor minor allele is less than 2.

Post-processing, a total of 5,268,206 SNPs were retained for further analysis.

**QUANTIFICATION AND STATISTICAL ANALYSIS**

**Data pre-processing**

Since background linkage disequilibrium (LD) generated by genetic drift could generate spurious clustering [14, 68], we first need to avoid such bias. Given that the high missing ratio of the loci will influence the accuracy of the LD value estimation, in the following analyses we only used the loci for which the missing ratio is less than 20%. We first ran a pilot test to select the appropriate distance cutoff for pairwise LD calculation using PLINK v1.07 [59]. In our dataset, when the distance of neighboring SNPs is beyond 500k, 97.2% LD values are less than 0.2 (very weak LD values), which suggests there are very few localized SNP clusters on this scale. Furthermore, following published guidance [68], we set the cutoff for the high LD value to 0.5. We used the following steps to filter the data. On each chromosome/scaffold, we started at the first SNP locus, measured the pairwise LD successively, until we reached the first SNP for which LD was less than 0.5. We then retained the first and last SNP loci, then started this process again from the last SNP until all SNP loci had been scanned. We finally obtained a dataset of 3,246,559 SNPs to be used in the population structure analyses.

**Population structure**

We used the processed SNP set to construct a phylogenetic tree for all 65 sequenced samples using RapidNJ v2.3.2 [54] based on the neighbor-joining method [12]. The core of this method is to calculate the distance matrix of $D_{ij}$ between each pair of individuals (i and j) as the following formula:

$$D_{ij} = \frac{\sum_{m=1}^{M} d_{ij}/L}{M}$$

$M$ is the number of segregating sites in i and j

$L$ is the length of regions

$d_{ij}$ is the distance between individuals i and j at site

$d_{ij} = 0$, when individuals i and j are both homozygous for the same allele (AA/AA)

$d_{ij} = 0.5$, when at least one of the genotypes of an individual i or j is heterozygous (Aa/AA, AA/Aa or Aa/Aa)

$d_{ij} = 1$, when individuals i and j are both homozygous but for different alleles (AA/aa or aa/AA)

We also identified the groups/subgroup structure of all 65 individuals by principal component analysis (PCA) of biallelic SNPs using the R package ‘pcadapt’ v3.0.4 [13] in R v3.2.2 [56]. We chose 4 as the optimal K from the PCA by scree plot, because the eigenvalue departs from the straight line. PCA result is consistent with that observed in the NJ tree.

The individual ancestries of all 65 samples were inferred by the Bayesian inference program STRUCTURE v2.3.4 [14] after removing the missing loci. Analyses were run five times for each level of K (from 2 through 10), with a 40,000 Burn-in period and 15,000 MCMC repetitions using the admixture model without prior information about populations. The ad hoc statistic $D_k$, based on the rate of change in the log probability of data between successive K values, was used to find the optimal number of subpopulations [15] (Data S2).

**Population demographic analysis**

To gain insights into the demographic history of the historical and contemporary populations, we first performed the PSMC analysis [16]. We used the heterozygous SNP loci with MAF $\geq 0.2$ and sequencing depth $>2X$ and $<30X$ on autosomal sequences. According to our previous experience [59], we set the parameters for the PSMC to be “N30 -t5 -r5 -p 4+30*2+4+6+10,” and performed bootstraps (100 times) to represent the variance of results. To further reconstruct recent demographic history (from the sampling time points to 10 kya), we ran an approximate Bayesian computation approach named PopSizeABC v2.1 [17] using the same input
of the PSMC analysis, with the following parameters: mac (minor allele count threshold for AFS and IBS statistics computation) is 0; mac ld (minor allele count threshold for LD statistics computation) are 2,3,5,9 separately; L (size of each segment, in bp) is 4000000; nb_rep (number of simulated datasets) is 500; nb_seg (number of independent segments in each dataset) is 30; all others as default. \( N_e \) of the three groups (EA, NW and contemporary group) at the sampling time point was estimated by NeEstimator v2.1 [57], and according to the NJ tree, the \( N_e \) of three NE subgroups were calculated respectively using the same method (Table S1).

**Ecological niche models analysis**

There are 35 occurrences of crested ibis assigned to clear locality in the 19th and 20th centuries in the VertNet database, which can be used for the breeding ranges prediction. We put the latitude and longitude to these 35 occurrences based on their identified localities into this analysis. Subsequently these geographic coordinates were used to construct ecological niche models (ENMs) for the contemporary period (19th-20th centuries), the Last Glacial Maximum period (LGM, approximately 22 kya) and the Last Interglacial period (LIG, approximately 120-140 kya). To build the ENMs, we also used the bioclimatic variables from WorldClim v1.4 [52] as the environmental layers in Maxent v3.4.0 [55]. The spatial resolution for all bioclimatic variables was unified to 2.5 arc-minutes (approximately 4 km x 4 km). For the LIG period, we aggregated the original 30 arc-second data to above unified resolution by function r.resamp.stats in GRASS GIS v7.2.1 with new cell size 0.041666666667 and average aggregation method. For all 19 bioclimatic variables, the predictions were performed in 10 replicates using Maxent, with extrapolation disabled and the other parameters set to default.

**Temporal population polymorphism**

In order to comprehensively characterize the population polymorphism, we used three methods. First, we estimated nucleotide diversity \( (\pi) \) for all 65 individuals in each 5Mb window across the chromosomes using the nuc.div function within the R package ‘pegas’ v0.10 [61]. Second, we measured the extent of the polymorphism fixation process in the contemporary group, when restricting the number of individuals with missing genotypes to be fewer than 5 in the historical groups, and no missing individuals in the contemporary group. Third, we identified the Identity by Descent (IBD) regions using GERMLINE v1.5.1 [60] on each chromosome in the historical and contemporary groups respectively.

**Correlation between inbreeding coefficient and effective population size \( (N_e) \)**

As we know that populations with smaller \( N_e \) tend to have proportionally more individuals with higher inbreeding coefficients than populations with larger \( N_e \), we explored whether the historical groups with the relative larger \( N_e \) had also suffered from inbreeding, and whether the inbreeding coefficient of the contemporary group is significantly higher than that of other species who have a similar \( N_e \). We therefore collected 14 published population SNP datasets [43–51], estimated their inbreeding coefficients and \( N_e \), and used these data to build the correlation relationship. The inbreeding coefficient \( F_{\text{UNI}} \) was estimated based on the correlation between uniting gametes following Wright [24] using Genome-wide Complex Trait Analysis (GCTA) v1.91.4 [58]. The \( N_e \) was estimated by NeEstimator v2.1 [57]. All results can be found in Table S2. Furthermore, the linear regression function (lm) in R v3.2.2 was applied to build and examine the correlation between \( F_{\text{UNI}} \) and \( N_e \), and was also used to determine whether the values of the contemporary group deviated significantly from the regression equation. The F-statistic of the coefficient indicated that \( F_{\text{UNI}} \) has a significant correlation with \( N_e \) (p value = 0.0000697), and the adjusted R-squared (0.7891) also supported the high credibility of the regression equation. By using this regression equation, we found that when \( N_e \) was set to 10.7 (\( N_e \) of the contemporary group), the \( F_{\text{UNI}} \) fit was 0.07 (95% CIs: 0.05-0.10). However, the \( F_{\text{UNI}} \) of the contemporary group was 0.17, verifying its significant deviation from the regression equation.

**Patterns of deleterious mutations across populations**

We used two different measures for estimating the deleterious load: deleterious missense mutations, and loss of function (LOF) mutations. Since these two measures need the ancestral state for the crested ibis for comparison, we used the genotypes of major alleles in the historical groups to represent the ancestral state. In each measure, we compared the counts of 1) only homozygous sites (2 per each site) and 2) both homozygous and heterozygous sites (2 per homo- and 1 per heterozygous site).

The first measure we used to diagnose the deleteriousness of missense mutations is the Grantham Score [26], a measure of physical/chemical properties of amino acid changes. The score ranges from 5 to 215, with scores bigger than 150 designated as radical or deleterious [70]. We further categorized coding mutations into 1) deleterious, 2) benign, and 3) synonymous mutations as follows:

1. deleterious: missense mutations with Grantham Scores equal to or more than 150.
2. benign: missense mutations with Grantham Scores less than 150.
3. synonymous: synonymous mutations, which do not cause amino acid sequence change.

The second measure is the LOF mutations inferred for the coding region SNPs of each sample using the SnpEff v4.3 algorithm [62]. We built a database from the crested ibis annotation and the reference crested ibis genome sequence, and used the input files in VCF format to annotate SNPs and assigned a SnpEff category to the input SNPs per individual. Briefly, two categories, stop codon (nonsense) and splice site disrupting single-nucleotide variants (SNVs), are considered as the LOF mutations. The ratio of homo- to
homo- and heterozygous LOF sites in the contemporary population is significantly higher than that in the historic populations (0.336 ± 0.042 in contemporary group, 0.253 ± 0.038 in historical group, Welch Two Sample t test, p = 0.0005386, Figure S2F).

**MHC diversity analysis**

A previous study exploring the MHC genomic organization and structure showed that the crested ibis contains 54 genes within three regions spanning 500 kb [71]. We downloaded the following three region sequences from NCBI (Class I 150 kb KP182409.1, Class II 85 kb KP182408.1, and Extended Region 265 kb KP182407.1). We produced the MHC SNPs dataset for all 65 individuals by mapping the raw sequencing data against three regions using the same method as hereinbefore described. Since the genome is organized into blocks of haplotypes [72, 73], we compared the extent of the diversity changes in MHC region based on the haplotype block composition. The software GERBIL v1.1, a tool in the GEVALT software suite [63], was used to capitalize on the greater resolution of genomic markers, and to aggregate SNPs into discrete haplotype blocks. Haplotype blocks were detected based on the MHC SNPs dataset of 57 historical samples after filtering the loci with more than half missing individuals. The historical MHC SNPs dataset was first converted to the format described in the GERBIL software manual. Briefly, individual tab-delimited files were produced for each of the scaffolds, with columns representing SNPs and every two consecutive rows corresponding to one sample. SNPs were represented in 0, 1 format or a “?” if missing. Re-formatted files were analyzed in an iterative fashion using default software parameters (sample code: gerbil.exe < input filename > <output filename > ). We also identified the haplotype blocks based on the SNPs dataset of 8 contemporary samples after filtering the loci with more than half missing individuals using the same way. In this way, haplotype structures in two groups were independent, and we used the length and the allelic richness of haplotype blocks as the indexes to feature the haplotype structures and the genetic diversity, respectively. Since the different sample size should be taken into account when comparing the genetic diversity between the historical and contemporary groups, we picked up 8 samples from the historical groups 10 times randomly, counted the number of alleles for each time and used the average values to represent the genetic diversity of the historical groups (Figure 4, Figure S3).

**Natural selection analysis**

To measure the selection across the coding regions, we used the R package ‘PopGenome’ v2.6.0 [64] to calculate Tajima’s D value [74], since a positive value is indicative of balancing selection or population contraction. We divided all SNPs into the 3 historical groups (NW, NE and EA) and the contemporary group, and only chose the loci without missing data in each group as input data. To reduce the influence of demographic history when interpreting the Tajima’s D results, for each group we selected genes with the top 5% highest Tajima’s D values, and considered those genes that were present in at least two of the historic populations as candidate genes that were under balancing selection at least 100 years ago.

**DATA AND SOFTWARE AVAILABILITY**

Raw sequence read data have been deposited into the study accession number PRJNA481495 and are also available in the CNGB Nucleotide Sequence Archive under accession number CNSA: CNP0000084.