

## RESEARCH PAPER

# Conservation genomics of the critically endangered Chinese pangolin

Shichao Wei<sup>1,2</sup>, Huizhong Fan<sup>3</sup>, Wenliang Zhou<sup>2</sup>, Guangping Huang<sup>1</sup>, Yan Hua<sup>4</sup>, Shibao Wu<sup>5</sup>, Xiao Wei<sup>6</sup>, Yiting Chen<sup>2</sup>, Xinyue Tan<sup>2</sup> & Fuwen Wei<sup>1,2,3\*</sup>

<sup>1</sup>Jiangxi Province Key Laboratory of Conservation Biology, College of Forestry, Jiangxi Agricultural University, Nanchang 330045, China;

<sup>2</sup>Center for Evolution and Conservation Biology, Southern Marine Science and Engineering Guangdong Laboratory (Guangzhou), Guangzhou 511458, China;

<sup>3</sup>Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China;

<sup>4</sup>Guangdong Provincial Key Laboratory of Silviculture, Protection and Utilization, Guangdong Academy of Forestry, Guangzhou 510520, China;

<sup>5</sup>School of Life Science, South China Normal University, Guangzhou 510631, China;

<sup>6</sup>Terrestrial Wildlife Rescue and Epidemic Diseases Surveillance Center of Guangxi, Nanning 530003, China

\*Corresponding author (email: [weifw@ioz.ac.cn](mailto:weifw@ioz.ac.cn))

Received 15 December 2023; Accepted 7 April 2024; Published online 4 July 2024

The Chinese pangolin (*Manis pentadactyla*, MP) has been extensively exploited and is now on the brink of extinction, but its population structure, evolutionary history, and adaptive potential are unclear. Here, we analyzed 94 genomes from three subspecies of the Chinese pangolin and identified three distinct genetic clusters (MPA, MPB, and MPC), with MPB further divided into MPB1 and MPB2 subpopulations. The divergence of these populations was driven by past climate change. For MPB2 and MPC, recent human activities have caused dramatic population decline and small population size as well as increased inbreeding, but not decrease in genomic variation and increase in genetic load probably due to strong gene flow; therefore, it is crucial to strengthen *in situ* habitat management for these two populations. By contrast, although human activities have a milder impact on MPA, it is at high risk of extinction due to long-term contraction and isolation, and genetic rescue is urgently needed. MPB1 exhibited a relatively healthy population status and can potentially serve as a source population. Overall, our findings provide novel insights into the conservation of the Chinese pangolin and biogeography of the mammals of eastern Asia.

*Manis pentadactyla* | conservation genomics | demographic history | human activities | conservation strategies

## INTRODUCTION

Genetic information is crucial for developing conservation management strategies for endangered populations and species (Wei et al., 2022a; Willi et al., 2022). Traditional population genetics provide limited genetic information based on a few genetic markers (e.g., mitochondrion and microsatellite) across individuals (Hohenlohe et al., 2021). Conservation genomics employs several markers to simultaneously genotype putatively neutral and adaptive loci, offering accurate information for the population structure, population history, and evolutionary potential for endangered species (Hu et al., 2023; Ralls et al., 2018; Ralls et al., 2018; Zhao et al., 2013).

Pangolins, members of the order Pholidota and family Manidae, represent the most heavily poached and trafficked wild mammals globally (Challender et al., 2015). This extensive exploitation is driven by a demand for their meat and scales, pushing them toward extinction (Challender and Hywood, 2011; Heinrich et al., 2016). All pangolin species are critically endangered on the International Union for Conservation of Nature (IUCN) Red List and are listed in CITES Appendix I. The Chinese pangolin (*Manis pentadactyla*, Linnaeus, 1758), distributed from the Eastern Himalayas to Southern China, has been classified into three subspecies: Taiwan (*M. p. pendactyla*, Linnaeus, 1758), Hainan (*M. p. pusilla*, Allen, 1906) and South

China (*M. p. aurita*, Hodgson, 1836) (Challender et al., 2019). These subspecies are mainly distributed in the islands of Taiwan and Hainan and mainland Asia (Challender et al., 2019). However, the status of these subspecies and population structure has not been confirmed through molecular evidence and necessitates further research (Challender et al., 2019). Effective conservation management strategies are urgently needed for the Chinese pangolin. It has been predicted a 50% habitat reduction and a 94% population decline from the 1960s to the early 2000s (Wu, 2004; Yang et al., 2018). In addition, a 90% decline has occurred over the last 20 years, and a further 90% decline has been predicted over the next twenty years (Challender et al., 2019).

Studies have investigated the genomic characteristics of the Chinese pangolin with samples from Yunnan, Guangdong, Taiwan and unknown origins (Hu et al., 2020a; Wang et al., 2022; Wei et al., 2022b). However, limited sampling range and a high proportion of unknown origin have hindered the population structure of the Chinese pangolin. It becomes particularly challenging when genetic evidence contradicts intraspecies taxonomy in endangered species, for instance, the red panda (*Ailurus fulgens*) (Hu et al., 2020b) and the takin (*Budorcas taxicolor*) (Yang et al., 2022). Furthermore, long runs of homozygosity that commonly represent recent inbreeding, were insufficiently detected as the reference genome was assembled

with short-read sequencing technologies (Edge and Bansal, 2019). Therefore, more extensive and accurate origin sampling must be performed for the population structure, as well as evaluation of population history, evolutionary potential, and genomic consequences of human activities in the Chinese pangolin.

Here, we analyzed whole-genome resequencing of 94 Chinese pangolins, including 56 newly sequenced individuals and 38 individuals from published databases. We aim to provide a comprehensive evaluation of their population structure and genomic consequences caused by past climate change and human activities. These findings will enable us to propose adaptive management strategies for populations of the Chinese pangolin.

## RESULTS

### Population structure of the Chinese pangolin

We used whole genome resequencing data from 94 samples for genomic analysis, which covers the main distribution range of the Chinese pangolin (Figure 1A). Among the samples' sequence data, 56 were newly sequenced and 38 were downloaded from other studies (Table S1 in Supporting Information). We first confirmed that all individuals were Chinese pangolins based on the phylogeny inference that was constructed by whole mitochondrial sequences (Figure S1 in Supporting Information). Genome coverage ranged from  $10.84\times$  to  $50.75\times$ , with an average coverage of  $(24.31\pm 10.48)\times$  (Table S1 in Supporting Information). We identified 3.39 million SNPs across the genome, of which 3.26 million were located on autosomes, 129.73 thousand found on the X-chromosome.

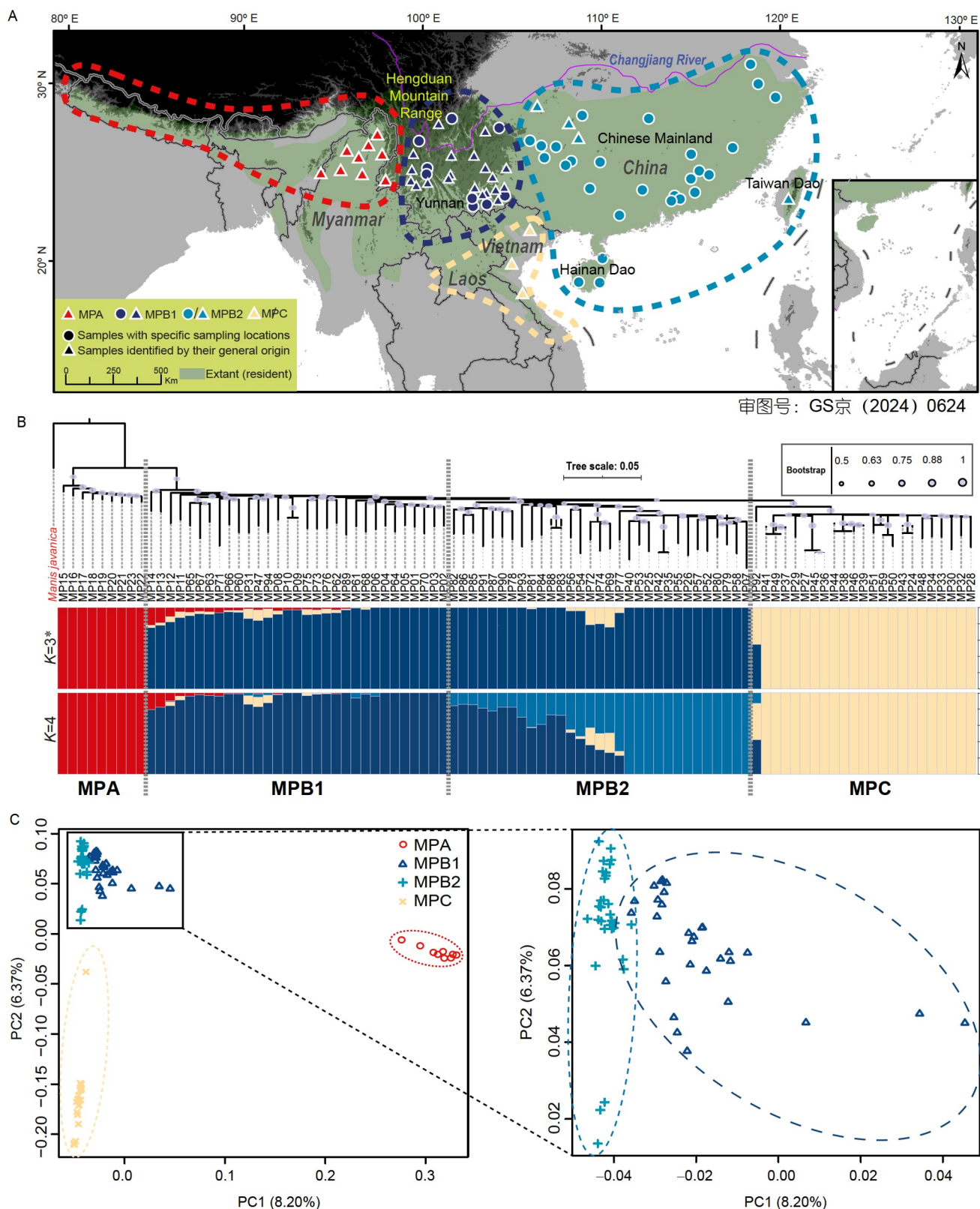
Population structure analysis based on autosomal SNPs, X-chromosome SNPs and mtGenome sequences revealed three distinct genetic clusters (MPA, MPB and MPC) (Figure 1B and C; Figures S2–S5 in Supporting Information). These clusters exhibited significant genetic divergence, with the  $F_{ST}$  values ranging from 0.10 to 0.14 between MPB and the other two genetic clusters, and the  $F_{ST}$  value between MPA and MPC was 0.19 (Table S2 in Supporting Information). Combining the results of sample origin and genetic structure, we infer that MPA is distributed in the Hengduan Mountain Range along with the southern Tibetan Plateau (Figure 1A). MPB is distributed in Southern China, encompassing the Chinese Mainland and islands of Taiwan and Hainan. MPC is distributed in the eastern Indo-China Peninsula, covering Vietnam and Laos. The Hengduan Mountain Range is most likely the boundary of MPA and MPB, whereas the demarcation line between tropics and subtropics likely serves as the boundary between MPB and MPC (Figure 1A).

Notably, we found that MPB has further diverged into two populations, MPB1 and MPB2, showing two genetic subclusters in the ADMIXTURE analysis, principal component analysis (PCA) and phylogenetic tree reconstruction based on autosomal and X-chromosome SNPs (Figure 1B and C; Figures S2 and S3 in Supporting Information). The  $F_{ST}$  value between MPB1 and MPB2 was estimated at 0.05 (Table S2 in Supporting Information), indicating moderate differentiation as originally described. Together, these findings indicate the presence of three distinct genetic clusters, which can be further divided into four populations.

### Population evolutionary history is affected by past climate change and human activities

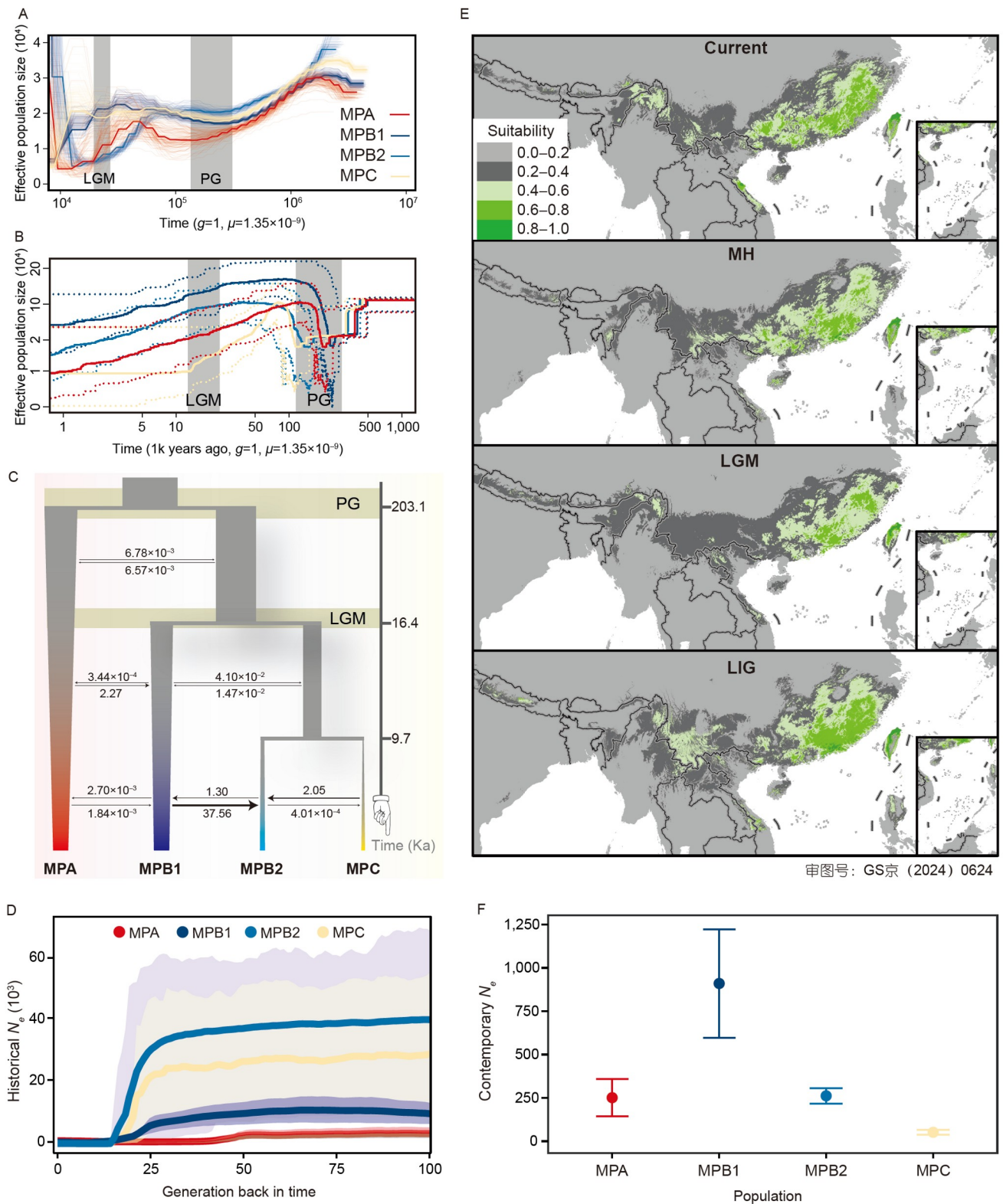
MPA first diverged from MPB1-MPB2-MPC at approximately 0.20 Ma, coinciding with the bottleneck event that occurred in all populations 0.15–0.4 Ma during the Penultimate Glaciation (PG; 0.13–0.3 million years ago (Ma)) (Zheng et al., 2002) (Figure 2A, B and E; Table S3 in Supporting Information). Later, MPA underwent significant population decline and habitat reduction and retreated northward during the Last Glacial Maximum (LGM; 16–26 thousand years ago (Ka)), leading to a long-term isolation in the Hengduan Mountain Range. Subsequently, MPB1 diverged from MPB2–MPC at 16.4 Ka, possibly driven by LGM (Lu et al., 2013), evidence by population decline and decrease in habitat suitability from the Last Interglacial (LIG; 120–140 Ka) to LGM (Figure 2A–C and E). Finally, MPB2 diverged from MPC at 9.7 Ka, near the Mid-Holocene (MH; 5–9 Ka). At the same period, MPB1 underwent contraction in the southern, resulting in a disconnection between MPB1 and MPC from LGM to MH (Figure 2C and E). Thus, climate change during MH drove the divergence of MPB1 and MPC; and niche isolation imposed between the tropics and subtropics maintained that differentiation. Furthermore, the geographical isolation between the populations of the Chinese Mainland and those of the islands of Taiwan and Hainan began during LGM and completed approximately 8 Ka (Figure S6 in Supporting Information). Interestingly, despite significant distinct suitable habitats, strong recent gene flow was detected among MPB1, MPB2, and MPC populations (Figure 2C; Table S4 in Supporting Information). Notably, our SDM simulations generally delineate four population boundaries in accordance with the inference of population structure analysis (Figures 1A and 2E). Overall, these results revealed that past climate change had driven the divergence of Chinese pangolin populations.

The Chinese pangolin has been extensively exploited and its population has experienced a dramatic decline in recent times (Challender et al., 2019; Wu, 2004; Yang et al., 2018). We investigated the genomic consequences of human activities in the populations of the Chinese pangolin. We used GONE analysis to reconstruct changes in effective population size over the past century and found varying degrees of change in the four populations (Figure 2D). MPA, MPB1, MPB2, and MPC experienced declines of approximately 5-, 8-, 200-, and 150-fold, respectively. The decline in MPB2 and MPBC occurred 10–40 years ago (Figure 2D). Contemporary effective population size ( $N_e$ ) estimates based on linkage disequilibrium (LD) indicated that average  $N_e$  values for MPA, MPB1, MPB2, and MPC were 250, 900, 250, and 60, respectively (Figure 2F). Human disturbance statistics showed a rise in human population counts and intensified land use for building, cropland, and grazing over the past century across the regions of the four populations (Figure S7 in Supporting Information). The region for MPB2 had markedly higher population counts and intensified land use compared with other regions. Although, the region for MPC exhibited significantly lower human population counts and land use, there was a dramatic recent population decline, which could be attributed to other types of human activities, such as hunting. These results indicate that human activities have led to the recent population decline and small  $N_e$ , especially for MPB2 and MPC.



审图号: GS京 (2024) 0624

**Figure 1.** Sampling origin and population genetic structure based on autosomal SNPs of the Chinese pangolin. A, Ninety-four samples were used in this study. Among these, 74 with known origins are shown on the map. Circles represent samples with specific sampling locations; triangles represent samples with known origins limited to general areas, such as provinces or regions. The 94 Chinese pangolin (*Manis pentadactyla*, MP) samples were clustered into three genetic clusters (MPA, MPB and MPC), with MPB further divided into two populations (MPB1 and MPB2). MPA, MPB1, MPB2, and MPC populations are denoted by red, blue, watery blue, and yellow, respectively. The distribution range of extant Chinese pangolin was obtained from the IUCN. B, ADMIXTURE results with  $K$  values ranging from 3–4, and supported three ancestral populations ( $K=3$ ). C, Principal component analysis was performed on all 94 individuals. Different colors of individual data points correspond with their cluster identities.



审图号: GS京(2024) 0624

**Figure 2.** Ancient demographic history and species distribution model inferred for Chinese pangolin populations. A and B, Inference of ancient effective population size history using the PSMC algorithm and stairway plot, with  $1.35 \times 10^{-9}$  mutations per site per generation and 1 year as generation time. C, The divergence history of the four Chinese pangolin populations was reconstructed using fastsimcoal2. The time on the node indicates divergence time, and the number of migrants per year between adjacent populations is shown next to each arrow. D, Historical  $N_e$  estimates (thick lines) and 95% confidence intervals (light shaded regions) over the last >100 generations, estimated from patterns of LD among linked SNPs. E, Potential suitable areas in current and historical times (MH, 5–9 Ka; LGM, 16–26 Ka; and the LIG, 120–140 Ka) are depicted. Colors indicate the probability of occurrence. F, Estimates and 95% confidence intervals for contemporary effective population size were derived from patterns of LD between loci on separate chromosomes.

## Variations in linkage disequilibrium, genomic variation, and inbreeding levels across populations

Genomic variation analysis revealed an average heterozygosity ( $H_e$ ) of  $8.63 \times 10^{-4}$  across the four populations (Figure 3B), which is comparable with other studies (Hu et al., 2020a; Wang et al., 2022). Notably, using chromosomal-level genome assemblies by long-read sequencing technologies revealed that the average runs of homozygosity (ROHs) fraction in the genome ( $F_{ROH}$ ) was 0.30, that is significantly higher than 0.20 reported in other studies (Wang et al., 2022). The difference was mainly observed in medium (1,000–3,000 kb) and long (>3,000 kb)  $F_{ROH}$ , with our study identifying a fraction of 0.14 compared with 0.01 in other studies (Wang et al., 2022).

MPA exhibited the highest LD, the lowest genetic diversity, and highest  $F_{ROH}$  (Figure 3A–C; Table S5 in Supporting Information), indicating its lowest evolutionary potential compared with other populations. By contrast, MPB1 showed relatively lower LD, the highest genetic diversity, and lowest  $F_{ROH}$ , indicating that it has the highest evolutionary potential. MPB2 and MPC showed intermediate evolutionary potential (Figure 3A–C; Table S5 in Supporting Information). These results imply different demographic histories of populations and reflect the genetic impacts of long-term population isolation and contraction in MPA.

Interestingly, MPB2 and MPC have the highest levels of inbreeding at the medium (1,000–3,000 kb) and long-size range (>3,000 kb) (Figure 3C). Long-length ROH typically results from recent inbreeding (Kirin et al., 2010). Based on the formula proposed by Kardos et al. (2018) ( $\text{Generations} = 100 / (2 * \text{ROH length})$ ), recent inbreeding events (ROH length from 1 Mb to 3 Mb) in MPB2 and MPC were estimated to occur 17–50 years ago. This corresponds to the time of population decline in MPB2 and MPBC, which occurred 10–40 years ago (Figure 3D), as well as the time of a significant increase in human disturbance 20–60 years ago (Figure S7 in Supporting Information). Combining these findings with demographic history analysis, we conclude that human activities have caused a recent population decline, leading to high levels of recent inbreeding in MPB2 and MPC.

## Genetic load acceleration according to varying population demographic histories

Our demographic history analysis suggests long-term isolation and contraction for MPA and an extreme recent population decline for MPB2 and MPC. To assess how these demographic trajectories impacted fitness, we first examined putatively deleterious missense (DEL) and loss-of-function (LOF) mutations, which are proxies for genetic load. The dominance of all variants is assumed in two extreme scenarios with fully recessive ( $h=0$ ) or fully additive ( $h=0.5$ ). When deleterious mutations act in an additive manner, the genetic load is determined by counts of derived alleles per genome. In contrast, when deleterious mutations act in a recessive manner, the genetic load is determined by counts of homozygosity alleles per genome. Our analysis revealed that the number of derived and homozygosity alleles associated with LOF and DEL mutations was significantly higher in MPA than MPB1 (two-tailed MWU test  $P < 0.001$ ; Figure 4A and B). By contrast, these counts were similar or lower in MPB2 and MPC compared with MPB1 (Figure 4A and B). We further quantified genetic load using the ratio of homozygous

sites to homozygous and heterozygous sites (referred to as the ratio) for missense, LOF, and DEL mutations. These ratios followed a similar pattern, with MPA having the highest ratio, followed by MPB2, MPB1, and MPC (Figure 4C). Finally, we computed the relative accumulation of derived alleles ( $R_{XY}$ ) statistics to compare the expected number of derived alleles occurring only in one population between MPA, MPB2, MPC, and MPB1. LOF- and DEL-type mutations exhibited a relative accumulation of derived alleles in MPA (Figure 4D). By contrast, MPB2 and MPC showed a relative decrease or similarity in LOF- and DEL-type derived mutation alleles (Figure 4D).

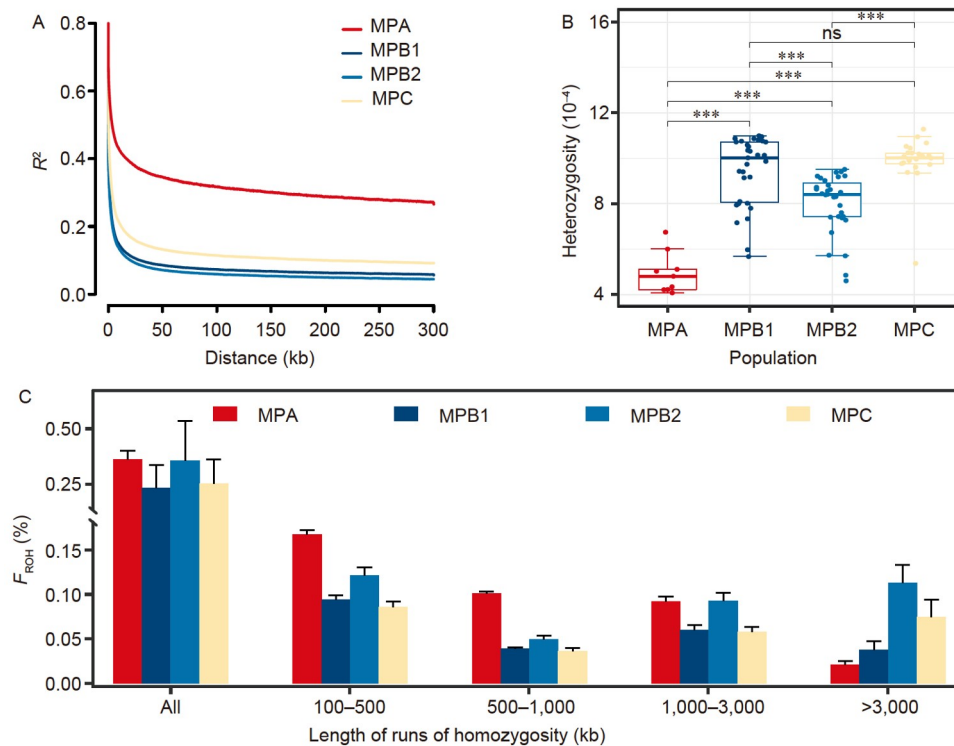
In summary, there is an increase in genetic load in MPA, whether deleterious mutations are recessive or additive. Notably, a significantly higher number of genes within homozygous deleterious mutations in ROHs was detected in MPA than in other populations (Figure S8 in Supporting Information), suggesting that the increased genetic load in MPA was likely due to inbreeding. In addition, the lowest level of genomic variation and highest levels of LD imply that MPA is at high risk of extinction. By contrast, human activities have a significant impact on MPB2 and MPC, but do not decrease genomic variation or increase genetic load.

## Potential risks associated with assisted gene flow among populations

Our findings indicate that MPA is at a high risk of extinction. Therefore, genetic rescue—introducing new individuals from other populations is urgently needed. However, considering the potential risks associated with genetic rescue programs is important (Kyriazis et al., 2021), particularly because we identified 3,107 genes with unique missense variants in the MPA population (Table S6 in Supporting Information). When MPA individuals were the introducing population 64–125 LOF mutations could be introduced (Figure S9 in Supporting Information); however, when MPA individuals were the receiving population, this number significantly decreased to 9–23.

## DISCUSSION

Through an enlarged sample collection across the main distribution range of Chinese pangolin, we elucidated that this species could be classified into four populations with varying degrees of genetic differentiation. The limited genetic differentiation observed between MPB1 and MPB2 may be attributed to the high level of gene flow. The observed divergence across MPA, MPB1, MPB2 and MPC ( $F_{ST} = 0.05–0.19$ ) is comparable to the genetic differentiation observed in the populations of other mammals that have a similar geographic distribution. For instance, takin (*Budorcas taxicolor*) populations have an  $F_{ST}$  value ranging from 0.07 to 0.20 (Yang et al., 2022), and tiger (*Panthera tigris*) populations have an  $F_{ST}$  value ranging from 0.11 to 0.59 (Sun et al., 2023). In addition, a clear distribution range and geographic boundary between four populations were revealed according to sample origin, which was further supported by the results of SDM analysis. These geographic boundaries also served as crucial population boundaries for other animals (Fu and Wen, 2023; Wei et al., 2020; Xing and Ree, 2017). Our findings provide significant insights into the biology and evolution, as well as effective conservation management strategies, of this species.



**Figure 3.** Linkage disequilibrium, genomic variation, inbreeding, recent demographic history and contemporary  $N_e$  in the four populations of the Chinese pangolin. A, LD of the four populations. B, Heterozygosity in each population. Error bars show range of values within 1.5 times the interquartile range. Significance levels: ns, not significant; \*\*\*,  $P < 0.001$ . C, Inbreeding estimated as the  $F_{ROH}$  in each population, including all, short (100–500 kb, 500–1,000 kb), medium (1,000–3,000 kb), and long (>3,000 kb) ROHs. Error bars represent the s.e.m.

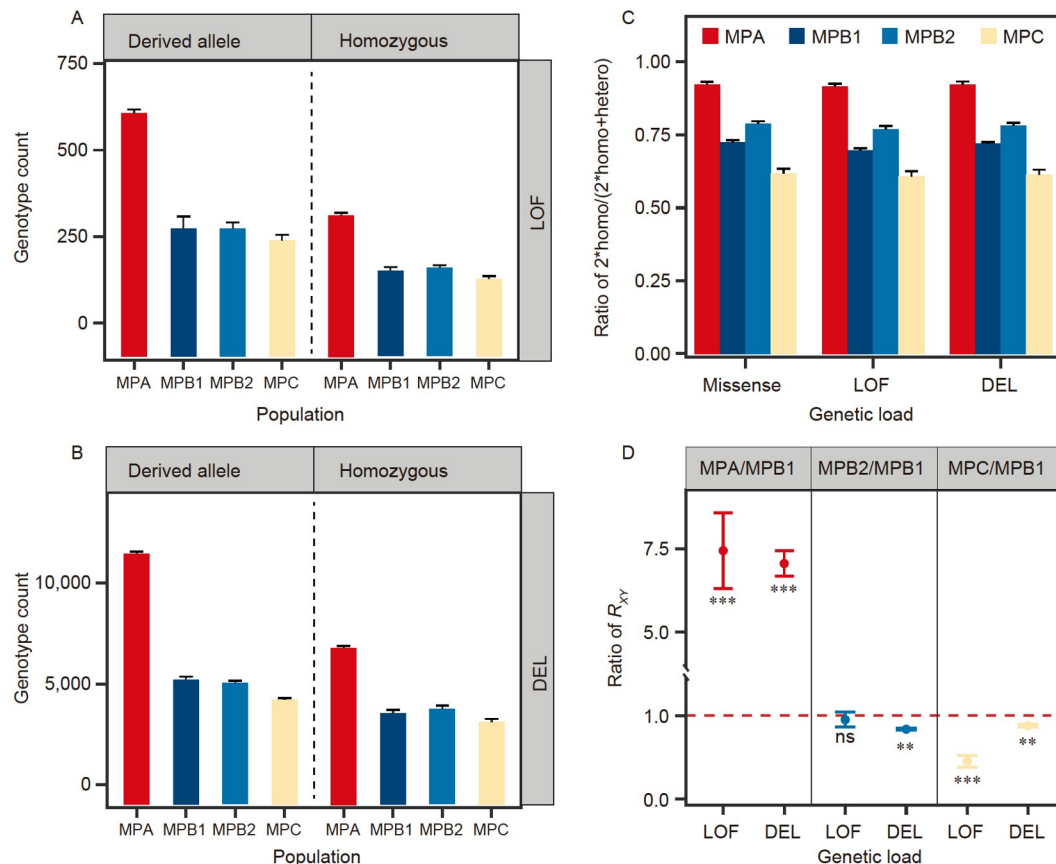
Our genomic evidence contradicts the previously proposed taxonomical viewpoints (Challender et al., 2019). Specifically, individuals from Chinese Mainland, Taiwan and Hainan formed a single genetic cluster (refer to MPB2), suggesting that *M. p. aurita* and *M. p. pusilla* might be defined as synonyms of *M. p. pendactyla*. However, caution should be exercised when drawing this conclusion due to the limited number of samples in the present study, such as three from Hainan and only one from Taiwan. Besides, the differing population demographic histories between Chinese Mainland, Taiwan, and Hainan should also be taken into account. Further efforts are encouraged to refine the intraspecies taxonomy, including the collection of more samples and phenotypic data.

The divergence among Chinese pangolin populations can primarily be attributed to past climate change. The Chinese pangolin, is highly sensitive to temperature, and a rapid drop in temperature can cause pneumonia and death in this species (Challender et al., 2019). The influence of past climate change on population or species divergence has been observed in other geographically overlapping mammals, such as the giant panda (Zhao et al., 2013), the takin (Yang et al., 2022), the red panda (Hu et al., 2020b) and the golden snub-nosed monkey (Kuang et al., 2019). Therefore, it appears to be a common pattern that past climate change has impacted the evolutionary history of mammals in eastern Asia (Fu and Wen, 2023).

For MPB2 and MPC, human activities have resulted in a dramatic recent population decline and increased levels of recent inbreeding. This result is in accordance with field-based population survey results, which have confirmed a 50% reduction in habitat range and a 94% decline in population in

the MPB2 location region from the 1960s to the early 2000s (Wu, 2004; Yang et al., 2018). However, the decline of the two populations was influenced by different factors related to human activities. Population count and intensified land use were found to affect MPB2, whereas hunting might have driven the decline of MPC, as Vietnam has been a major source for the illegal Chinese pangolin trade over the past fifty years (Challender and Hywood, 2011; Heinrich et al., 2016; Zhang et al., 2015). Despite these challenges, MPB2 and MPC do not exhibit a substantial decrease in genome variation or accumulation of genetic load. One possible explanation is that the strong recent gene flow observed among MPB1, MPB2, and MPC has helped mitigate the genetic consequences of recent decline (Robinson et al., 2023), considering gene flow as low as one effective migrant per generation may counteract genetic drift and reduce the frequency of deleterious variations (Mills and Allendorf, 1996). However, the genetic status of these populations remains alarming, with MPB2 and MPC having small and very small effective population sizes ( $N_e = 250$  and 60, respectively). If these populations do not fully recover and remain small, diminishing migration could increase genetic drift and cause weakly deleterious mutations to gradually drift to fixation. Hence, urgent protection measures should be taken with a primary focus on strengthening *in situ* habitat management and increasing the population size for MPB2 and MPC.

MPA is at the highest risk of extinction, likely due to long-term isolation in the Hengduan Mountain Range, evidenced by older inbreeding and population decline and habitat reduction during LGM. Genetic rescue could be considered to enhance the evolutionary potential of MPA. However, caution is needed



**Figure 4.** Characterization of genetic load in the four populations of the Chinese pangolin. A and B, Derived genotypes and homozygous derived genotypes in two functional categories of variants, including deleterious nonsynonymous alleles (DEL) and LOF alleles. Error bars show range of values within 1.5 times the interquartile range. C, Derived mutations in the coding region comprise missense, LOF, and DEL mutations. Error bars represent the s.e.m. D,  $R_{xy}$  statistic in MPA (X) compared with MPB1 (Y), MPB2 (X) compared with MPB1 (Y), and MPC (X) compared with MPB1 (Y).  $R_{xy} > 1$  and  $R_{xy} < 1$  (dashed red line) indicate that deleterious alleles are more and less abundant in population X than in population Y, respectively. Solid points are point estimates and error bars represent the standard deviation based on jackknife distribution. We used a two-tailed Z score test without multiple testing adjustments. Significance levels: ns, not significant; \*\*,  $P < 0.01$ ; \*\*\*,  $P < 0.001$ .

when introducing individuals to genetic rescue programs (Kyriazis et al., 2021). Our estimates reveal that when individuals of the other three populations were introduced into MPA, comparable counts of new LOF variants were found which were comparable with those found in the Sumatran rhinoceros (von Seth et al., 2021) and Chinese alligator (Yang et al., 2023) populations. Therefore, it is feasible to introduce pangolins from other regions (e.g., from MPB1) to enhance the evolutionary potential of MPA.

MPB1 has the highest genomic variation and largest effective population size and has experienced a relatively mild population decline, without signs of recent inbreeding. This was likely due to lower human pressure compared with other regions. The Yunnan Plateau is recognized as a biodiversity hotspot, providing a stable and interconnected habitat for MPB1 due to its forested environment and continuous mountain range (Cheng et al., 2023; Myers et al., 2000). Therefore, we recommend that MPB1 could be a valuable source population for further genetic rescue efforts in other populations.

In conclusion, genomic information is crucial for the population structure, as well as understanding the evolutionary history and evolutionary potential of endangered species. Our study has clarified the population structure and showed that past climate change has driven the population divergence of the Chinese pangolin. Furthermore, we have revealed that human activities

have resulted in a dramatic recent population decline and increased levels of recent inbreeding in MPB2 and MPC. In addition, we have proposed adaptive management strategies for each population based on their demographic history and evolutionary potential. These findings have important implications for the conservation of the Chinese pangolin and shed light on how climate change and human activities have influenced mammal diversity in eastern Asia.

## MATERIALS AND METHODS

### Sample collection and sequencing

A total of 56 samples were obtained, of which 46 were confirmed to have accurate sampling location information and 10 lacked sampling location information (Table S1 in Supporting Information). Genome sequencing data from 38 Chinese pangolin samples were downloaded from NCBI (BioProject: PRJNA529540 and PRJNA20331) and CNGB Sequence Archive of China National GeneBank DataBase (CNGBdb) (accession number: CNP0001723) for analysis. We downloaded 72 Malayan pangolin (*Manis javanica*) genome sequencing data from NCBI (BioProject: PRJNA529540) for species identification.

Genome DNA extraction was conducted in a dedicated facility using the DNA extraction protocol for soft tissue museum

samples, according to Zhang et al. (2021). The ssDNA Library Prep Kit (Vazyme, ND620) was used to prepare genomic libraries. Libraries were sequenced using the Illumina NovaSeq 6000 platform, which generated paired/linked 150 bp Illumina short reads.

## Variant Calling and Filtering

SNP calling was processed using various bioinformatics tools based on whole-genome resequencing data of the 94 Chinese pangolins, including 56 newly sequenced individuals and 38 individuals from published databases (Table S1 in Supporting Information). Initially, FASTP v0.20.0 (Chen et al., 2018) was employed to filter out low-quality reads based on default parameters. The remaining high-quality sequence reads were aligned to the chromosomal-level reference genome achieved with long-read sequencing technologies of the Chinese pangolin from CNGBdb (BioProject: PRJCA020583) using BWA-MEM v0.7.17 (Li and Durbin, 2009) with default parameters. Subsequently, SAMTOOLS v1.9 (Li et al., 2009) was utilized to sort BAM files. To eliminate potential PCR duplications, the “MarkDuplicates” function in PICARD v2.21.6 was applied (<http://broadinstitute.github.io/picard>). All reads with low-quality bases ( $Q \leq 30$ ) or  $>3\%$  ambiguous nucleotides were removed. SNP calling was performed using the Genome Analysis Toolkit (GATK) v4.1.2.0 (McKenna et al., 2010), specifically employing the HaplotypeCaller analysis type with the “-ERC GVCF” option. Individual GVCF files were merged using the “CombineGVCFs” function in GATK. Candidate SNPs were identified and selected from the combined GVCF file using the “GenotypeGVCFs” and “SelectVariants” functions, respectively. To ensure the reliability of candidate SNPs, hard filters were applied to raw SNPs using GATK and VCFtools v0.1.17 (Danecek et al., 2011) according to the following criteria: qual by depth (QD)  $<2.0$ , RMS mapping quality (MQ)  $<40.0$ , fisher strand bias (FS)  $>60.0$ , strand odds ratio (SOR)  $>3.0$ , rank sum test for mapping quality (MQRankSum)  $<-12.5$ , rank sum test for read position (ReadPosRankSum)  $<-8.0$ , minimum read depth (minDP)  $<4$ , minor allele frequency (MAF)  $<0.05$  and max-missing  $>0.2$ . For our population genetic analyses, we applied rigorous data filters: (i) reads with depth  $<4\times$ , (ii) SNPs exhibiting a depth distribution of all sites  $<2.5\%$  or  $>97.5\%$ , (iii) SNPs with MAF  $<0.05$  and  $>20\%$  missing data across all individuals, and (iv) SNPs with excess heterozygosity (ExcHet  $<0.05$ ) and significantly deviating from Hardy-Weinberg Equilibrium (HWE  $<0.001$ ). After filtering, 3.39 Mb biallelic SNPs were retained.

## Species identification

Whole genome resequencing data from 94 Chinese pangolins and 72 Malayan pangolins were used to assemble the complete mitochondrial genome by employing MITOBIM v1.9.1 software (Hahn et al., 2013). The reference mitochondrial genome sequences of Chinese pangolin and Malayan pangolin were downloaded from the NCBI GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>, accession number NC\_016008 and NC\_026781, respectively). To determine the phylogenetic relationships, we employed the Neighbor-Joining (NJ) method and performed the analysis using MEGA (Kumar et al., 2018) with Kimura’s two-parameter model.

## Population genetic structure

We employed LD-based pruning to reduce the LD between SNPs in the genotype data. This was done using Plink v1.9 (Purcell et al., 2007) with the “-indep-pairwise 50 5 0.5” option for the autosomal SNP data set. We utilized the Genome-wide Complex Trait Analysis tool v1.93.0 (Yang et al., 2011) to perform principal component analysis (PCA). We also performed population genetic cluster analysis using ADMIXTURE v1.3.0 (Alexander et al., 2009) with default parameters. We gradually increased the assumed number of genetic clusters ( $K$ ) from 2 to 10. For each  $K$  value, we ran 10 independent replicates with different starting seeds and selected the replicates with the highest likelihood. To reconstruct phylogeny of the Chinese pangolin, we constructed NJ trees for autosomal SNPs, X-chromosomal SNPs, and mitochondrial genomes, with Malayan pangolin set as the outgroup by using MEGA with 1,000 bootstrap replicates. The resulting tree files in NWK format were visualized using the online tree illustration website (<https://itol.embl.de/>).

We utilized a sliding-window approach with a window size of 50 kb and an increment of 25 kb to assess the genome-wide distribution of  $F_{ST}$  values between populations by using VCFtools.

## Ancient demographic inference and divergence history

We employed a combined strategy involving two complementary algorithms to infer the detailed historical changes in  $N_e$  of each Chinese pangolin population. Initially, we utilized the PSMC algorithm (Li and Durbin, 2011), based on individual genomes to reconstruct historical changes in effective population size with the following set of parameters: -N 30 -t 15 -r 5 -p 4+25\*2+4+6. The nucleotide mutation rate ( $\mu$ ) of the Chinese pangolin was estimated to be  $1.35 \times 10^{-9}$  mutations per site per generation, using Malayan pangolin as the reference species for comparison. The generation time for the Chinese pangolin was 1 year (Zhang et al., 2016). This estimation was derived from the formula:  $\mu = D \times g / 2T$  (Kondrashov and Crow, 1993), where  $D$  represents the observed frequency of pairwise differences between the two species,  $T$  is the estimated divergence time, and  $g$  is the estimated generation time for the Chinese pangolin.

Additionally, we performed stairway plot analysis (Liu and Fu, 2015), based on folded site frequency spectrum (SFS) data estimated using easySFS software (<https://github.com/isaacov-ercast/easySFS>). To mitigate the potential influence of individuals with hybrid signals, only individuals without hybrid signals from the population were retained (Table S7 in Supporting Information). The same filtering settings were applied for population genetic analyses, retaining only SNPs with no missing data for each population. The stairway plot and its variance were estimated from 200 bootstrap SFSs.

To further infer divergence scenarios and gene flow among populations, we employed fastsimcoal2 (Excoffier and Foll, 2011) with the following parameters: -u -n 100,000 -d -M 0.001 -l 100 -L 100 -q -C 10 -B 60 -c 10. Four alternative population divergence models were simulated to explore the divergence order of the four populations (Figure S10 in Supporting Information). The best model was selected based on the maximum likelihood value, and parametric bootstrap estimates were obtained from 100 simulated datasets.

We used  $D$ -statistics/ABBA-BABA statistics to test gene flow

among populations with AdmixTools v7.0 (Patterson et al., 2012). We performed multiple tests for (((P1, P2), P3), O), with Malayan pangolin as the outgroup (O). Significant gene flow occurred between P2 and P3 with  $D > 0$  and  $|Z_{\text{score}}| > 3$ , as well as between P1 and P3, with  $D < 0$  and  $|Z_{\text{score}}| > 3$  (Durand et al., 2011).

## Species distribution modeling

We collected 722 occurrence records for the Chinese pangolin, covering its entire range through field surveys, literature sources (Sharma et al., 2020; Wei et al., 2022b), and databases, such as the Global Biodiversity Information Network GBIF (<http://www.gbif.org>) and VertNet (<http://vertnet.org/>). Initially, records with obvious georeferencing errors (e.g., locations in water bodies) and duplicates were excluded. To prevent model overfitting, only one record was retained for multiple records within a spatial resolution of 10 km×10 km. The final dataset comprised 193 occurrence data points (Figure S11 in Supporting Information).

To explore the potential impact of climatic changes on Chinese pangolin distribution, we obtained 19 bioclimatic layers from four time periods: current, MH, LGM, and LIG, from the WorldClim database (<http://www.worldclim.org/>). For both MH and LGM climate scenarios, three general circulation models (GCMs) (CCSM4, MIROC-ESM, and MPI-ESM-P) were used, whereas only one CCSM was available for LIG. PCA was employed to select BIO (bioclimatic variables) parameters to avoid multicollinearity. The analysis revealed that five principal components (PCs) explained 90% variance in data, and one representative BIO parameter per PC (BIO4=temperature seasonality, BIO12=annual precipitation, BIO14=precipitation of driest month, BIO18=precipitation of warmest quarter, and BIO19=precipitation of coldest quarter, Table S8 in Supporting Information) was selected for SDM. To determine appropriate SDM model options and settings, various combinations of feature classes (determining the shape of the response curves) and regularization multipliers (determining the penalty for adding parameters to the model) were evaluated based on model scores using the Akaike Information Criterion (AIC). The best model, model H1, was determined by utilizing the R package of ENMevaluate (Muscarella et al., 2014), which resulted in the lowest AIC value (Table S9 in Supporting Information). Model H1 utilized the feature types “LQHP” (where L=linear, Q=quadratic, H=hinge, and P=product) and a regularization multiplier of 1 (Table S9 in Supporting Information). The model’s projections were generated using the maximum entropy algorithm in MaxEnt v.3.3.3e (Phillips et al., 2006). Model performance was assessed by randomly selecting 75% of the localities as a training set, with the remaining 25% reserved for model validation.

## Recent demography inference and human disturbance analysis

We utilized an LD-based method in GONE (Santiago et al., 2020) to estimate a time series of recent  $N_e$  for all populations. We employed autosomal SNP genotypes from all samples after filtering out genotypes with  $GQ < 20$ , missing genotypes at  $> 10\%$ , and a minimum minor allele count of 2 within each population. This filtering resulted in sample sizes of 9 (MPA), 9 (MPB1), 9 (MPB2), and 8 (MPC) for GONE analysis. We assumed a constant recombination rate of 1.1 cM Mb<sup>-1</sup>, which is typical

among mammals. The confidence intervals for historical  $N_e$  was repeated 500 times, with each repetition using a different randomly selected set of 10,000 SNPs per chromosome. Additionally, we estimated contemporary  $N_e$  based on LD patterns using NeEstimator v.2.1 (Do et al., 2014) and SNP data used for GONE analysis.

If human disturbance had driven the endangerment of the Chinese pangolin, we expect stronger human disturbance in regions where the population decline was dramatic. To test these hypotheses, gridded data on human population estimates, such as population count and total built-in area per grid cell (referred to as popc and uopp, respectively) and land use, including total area for cropland and grazing land per grid cell (referred to as cropland and grazing, respectively), were obtained at a five arc-minute spatial resolution from HYDE database 3.2 (Klein Goldewijk et al., 2017). Variations in these four categories were extracted based on the distribution regions of the four populations (Figure S12 in Supporting Information), divided according to genetic structure and SDM results from 11 time points spanning from the early 1900s to 2010. The mean values of these four categories for the four regions across time points were calculated using the Raster package in R (Hijmans, 2015).

## Estimation of genomic variation, LD, and inbreeding

To avoid the potential biases stemming from higher sequence depth, which could affect the estimation of genetic variation, LD, inbreeding, and genetic load, we subsampled individual sequence reads to a mean depth of 24×. We utilized VCFtools v0.1.17 (Danecek et al., 2011) to compute nucleotide diversity ( $\theta_\pi$ ) and Watterson’s estimator ( $\theta_w$ ) within each population with the parameters “-window-pi 50,000 -window-pi-step 25,000”. Additionally, the  $H_e$  of an individual was estimated using VCFtools. Genetic diversity for mitochondrial genomes was estimated using DnaSP v5.10.01 (Librado and Rozas, 2009).

Genome-level LD for each population was assessed using PopLDdecay (Zhang et al., 2019) with default settings. ROH indicative of inbreeding, were identified using Plink with the parameters “-homozyg-kb 200 -homozyg-snp 50” following the method described by Dobrynin et al., 2015. Here, ROHs were categorized into different length classes: short (100–500 kb, 500–1,000 kb), medium (1,000–3,000 kb), and long ( $> 3,000$  kb). Long ROHs, indicating recent inbreeding occurring several decades ago, were calculated using the formula of Kardos et al. (2018) ( $\text{Generations} = 100/2 \times \text{ROH length}$ ) (Hu et al., 2020a), ROHs  $> 1$  Mb were assumed to be the result of more recent inbreeding ( $< 50$  years).

## Quantifying genetic load across populations

We used three lines of evidence to quantify relative levels of genetic load variation in the four populations. We focused on mutations occurring within protein-coding regions, which are more likely to have a direct impact on fitness. Ancestral and derived alleles were determined by comparing the genomes of the Chinese pangolin genome to the Malayan pangolin using LastZ v1.03.54 (Harris, 2007). We normalized the differences in missing data among individuals by an average number of called genotypes using vcfr v.1.12.0 in the R package (Knaus and Grünwald, 2017). We identified derived alleles within two mutation types: DEL and LOF. For the identification of deleterious

missense and LOF mutations, we annotated filtered VCF files using SnpEff v4.3 (Cingolani et al., 2012). The coding sequence variants were classified into two categories: LOF and missense variants. LOF variants are predicted to eliminate or significantly impair gene function, including splice acceptor, splice donor, start lost, and stop gain mutations. Missense variants were further classified to DEL mutations (SIFT score<0.05) based on phylogenetic constraints using SIFT4G (Vaser et al., 2016). Considering the limited understanding of variant dominance in natural populations, and recognizing that dominant variants ( $0.5 < h \leq 1$ ) as segregating deleterious variants are unlikely to be dominant, we explored two extreme scenarios: (i) recessive dominance ( $h=0$ ), where fitness is reduced only in homozygous derived genotypes, and (ii) additive dominance ( $h=0.5$ ), where fitness decreases linearly with the number of derived alleles.

First, we used two-tailed MWU tests to determine if the normalized count of derived alleles and homozygotes significantly varied across the four populations within the derived and homozygous mutation types. The counts of all DEL- and LOF-derived alleles are considered a proxy for additive genetic load, whereas the counts of DEL- and LOF-derived homozygote alleles represented a proxy for recessive load (Nigenda-Morales et al., 2023).

Second, we estimated genetic load by calculating the ratio of homozygous sites to homozygous and heterozygous sites (referred to as the ratio) for missense mutations, LOF mutations, and DELs for each individual.

Third, we calculated the relative accumulation of mutation ( $R_{XY}$ ) for LOF and DEL mutation types, as described (Nigenda-Morales et al., 2023). In this analysis, MPA, MPB2, and MPC were identified as population X and MPB1 as population Y. MPB1 served as the benchmark population because it exhibited a relatively healthier population status, characterized by the highest genomic variation, lowest linkage disequilibrium and inbreeding, and the highest contemporary effective population size. We estimated the standard errors of  $R_{XY}$  using a weighted-block jackknife method, as described (Do et al., 2015). If selection was equally effective, mutation rates remain consistent across both populations, and  $R_{XY}$  statistics are anticipated to be one. We evaluated whether the observed values of  $R_{XY}$  significantly deviated from the null expectation using the Z score test with  $P$  value<0.05.

## Data availability

The raw whole genome resequencing reads have been deposited in the CNGB Sequence Archive of China National GeneBank DataBase (CNGBdb) (accession number: CNP0001723).

## Compliance and ethics

The author(s) declare that they have no conflict of interest. F.Wei. conceived and supervised the project. S.Wei., H.Fan., W.Zhou., G.Huang., Y.Hua., S.Wu., X.Wei. and Y.Chen. collected the samples. X.Tan. carried out the DNA extraction and library construction. S.Wei. performed the bioinformatics analysis. S.Wei. wrote the manuscript. F.Wei., H.Fan., W.Zhou., S.Wei. revised the manuscript. All authors contributed to interpreting the data and approved the final manuscript.

## Acknowledgement

This work was supported by the Major Program of the Natural Science Foundation of Jiangxi Province of China (20233ACB209001), Guangdong Basic and Applied Basic Research Foundation (2022A1515111016), Science and Technology Department of Guangdong Province (2021QN02H103), Natural Resources Affairs Management-Ecological Forestry Construction Special Project of Forestry Administration of Guangdong Province (SLYJ2023B4002,

SLYJ2023B4003, SLYJ2023B4005). We express sincere thanks to the Department of Wildlife Conservation of Forestry Administration of Guangdong Province for their support during the project implementation. We are also grateful to Xianjing Peng, Qihai Zhou, Penglai Fan, Hua Wu, Baowei Zhang, Jiang Zhou and Wei Wei for their kind assistance in the sample collection. Special thanks are given to the following people who provided important help during various stages of the project: Lin Yang, Xin Du, Yisi Hu, Zhiwei Zhang, Xin Huang, Kai Wang, Hongliang Dou, Haiyang Gao, Fuyu An, Song Sun and Tao Luo.

## Supporting information

The supporting information is available online at <https://doi.org/10.1007/s11427-023-2540-y>. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

## References

- Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19, 1655–1664.
- Challender, D.W.S., Harrop, S.R., and MacMillan, D.C. (2015). Understanding markets to conserve trade-threatened species in CITES. *Biol Conservation* 187, 249–259.
- Challender, D.W., and Hywood, L. (2011). Asian pangolins: increasing affluence driving hunting pressure. *Traffic Bull* 23, 92–93.
- Challender, D.W., Nash, H.C., and Waterman, C. (2019). Pangolins: Science, Society and Conservation. London: Academic Press.
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890.
- Cheng, F., Tian, J., He, J., He, H., Liu, G., Zhang, Z., and Zhou, L. (2023). The spatial and temporal distribution of China's forest carbon. *Front Ecol Evol* 11, 1110594.
- Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms. SnpEff. *Fly* 6, 80–92.
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158.
- Do, C., Waples, R.S., Peel, D., Macbeth, G.M., Tillett, B.J., and Ovenden, J.R. (2014). NEESTIMATOR v2: re-implementation of software for the estimation of contemporary effective population size ( $N_e$ ) from genetic data. *Mol Ecol Resources* 14, 209–214.
- Do, R., Balick, D., Li, H., Adzhubei, I., Sunyaev, S., and Reich, D. (2015). No evidence that selection has been less effective at removing deleterious mutations in Europeans than in Africans. *Nat Genet* 47, 126–131.
- Dobrynin, P., Liu, S., Tamazian, G., Xiong, Z., Yurchenko, A.A., Krashenninnikova, K., Kliver, S., Schmidt-Küntzel, A., Koepfli, K.P., Johnson, W., et al. (2015). Genomic legacy of the African cheetah, *Acinonyx jubatus*. *Genome Biol* 16, 1–20.
- Durand, E.Y., Patterson, N., Reich, D., and Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Mol Biol Evol* 28, 2239–2252.
- Edge, P., and Bansal, V. (2019). Longshot enables accurate variant calling in diploid genomes from single-molecule long read sequencing. *Nat Commun* 10, 4660.
- Excoffier, L., and Foll, M. (2011). Fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. *Bioinformatics* 27, 1332–1334.
- Fu, J., and Wen, L. (2023). Impacts of Quaternary glaciation, geological history and geography on animal species history in continental East Asia: a phylogeographic review. *Mol Ecol* 32, 4497–4514.
- Harris, R.S. (2007). Improved pairwise alignment of genomic DNA. State College: The Pennsylvania State University.
- Hahn, C., Bachmann, L., and Chevreaux, B. (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucleic Acids Res* 41, e129.
- Hijmans, R.J. (2015). Raster: Geographic data analysis and modeling. R Package Version 2.4-15.
- Heinrich, S., Wittmann, T.A., Prowse, T.A.A., Ross, J.V., Delean, S., Shepherd, C.R., and Cassey, P. (2016). Where did all the pangolins go? International CITES trade in pangolin species. *Glob Ecol Conserv* 8, 241–253.
- Hohenlohe, P.A., Funk, W.C., and Rajora, O.P. (2021). Population genomics for wildlife conservation and management. *Mol Ecol* 30, 62–82.
- Hu, J.Y., Hao, Z.Q., Frantz, L., Wu, S.F., Chen, W., Jiang, Y.F., Wu, H., Kuang, W.M., Li, H., Zhang, Y.P., et al. (2020a). Genomic consequences of population decline in critically endangered pangolins and their demographic histories. *Natl Sci Rev* 7, 798–814.
- Hu, Y., Thapa, A., Fan, H., Ma, T., Wu, Q., Ma, S., Zhang, D., Wang, B., Li, M., Yan, L., et al. (2020b). Genomic evidence for two phylogenetic species and long-term population bottlenecks in red pandas. *Sci Adv* 6, eaax5751.
- Hu, Y., Wang, X., Xu, Y., Yang, H., Tong, Z., Tian, R., Xu, S., Yu, L., Guo, Y., Shi, P., et al. (2023). Molecular mechanisms of adaptive evolution in wild animals and plants. *Sci China Life Sci* 66, 453–495.

- Kardos, M., Åkesson, M., Fountain, T., Flagstad, Ø., Liberg, O., Olason, P., Sand, H., Wabakken, P., Wikenros, C., and Ellegren, H. (2018). Genomic consequences of intensive inbreeding in an isolated wolf population. *Nat Ecol Evol* 2, 124–131.
- Kirin, M., McQuillan, R., Franklin, C.S., Campbell, H., McKeigue, P.M., and Wilson, J. F. (2010). Genomic runs of homozygosity record population history and consanguinity. *PLoS ONE* 5, e13996.
- Klein Goldewijk, K., Beusen, A., Doelman, J., and Stehfest, E. (2017). Anthropogenic land use estimates for the Holocene-HYDE 3.2. *Earth Syst Sci Data* 9, 927–953.
- Knaus, B.J., and Grünwald, N.J. (2017). *vcfR*: a package to manipulate and visualize variant call format data in R. *Mol Ecol Resources* 17, 44–53.
- Kondrashov, A.S., and Crow, J.F. (1993). A molecular approach to estimating the human deleterious mutation rate. *Hum Mutat* 2, 229–234.
- Kuang, W., Ming, C., Li, H., Wu, H., Frantz, L., Roos, C., Zhang, Y., Zhang, C., Jia, T., and Yang, J.Y. (2019). The origin and population history of the endangered golden snub-nosed monkey (*Rhinopithecus roxellana*). *Mol Biol Evol* 36, 487–499.
- Kumar, S., Stecher, G., Li, M., Nkayaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 35, 1547.
- Kyriazis, C.C., Wayne, R.K., and Lohmueller, K.E. (2021). Strongly deleterious mutations are a primary determinant of extinction risk due to inbreeding depression. *Evol Lett* 5, 33–47.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.
- Li, H., and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature* 475, 493–496.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Librado, P., and Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452.
- Liu, X., and Fu, Y.X. (2015). Exploring population size changes using SNP frequency spectra. *Nat Genet* 47, 555–559.
- Lu, H.Y., Yi, S.W., Xu, Z.W., Zhou, Y.L., Zeng, L., Zhu, F.Y., Feng, H., Dong, L.N., Zhuo, H.X., Yu, K.F., et al. (2013). Chinese deserts and sand fields in Last Glacial Maximum and Holocene Optimum. *Chin Sci Bull* 58, 2775–2783.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytzky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20, 1297–1303.
- Mills, L.S., and Allendorf, F.W. (1996). The one-migrant-per-generation rule in conservation and management. *Conserv Biol* 10, 1509–1518.
- Muscarella, R., Galante, P.J., Soley-Guardia, M., Boria, R.A., Kass, J.M., Uriarte, M., and Anderson, R.P. (2014). ENM eval: an R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods Ecol Evol* 5, 1198–1205.
- Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B., and Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature* 403, 853–858.
- Nigenda-Morales, S.F., Lin, M., Nuñez-Valencia, P.G., Kyriazis, C.C., Beichman, A.C., Robinson, J.A., Ragsdale, A.P., Urbán R., J., Archer, F.I., Vilorio-Gómora, L., et al. (2023). The genomic footprint of whaling and isolation in fin whale populations. *Nat Commun* 14, 5465.
- Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. *Genetics* 192, 1065–1093.
- Phillips, S.J., Anderson, R.P., and Schapire, R.E. (2006). Maximum entropy modeling of species geographic distributions. *Ecol Model* 190, 231–259.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81, 559–575.
- Ralls, K., Ballou, J.D., Dudash, M.R., Eldridge, M.D.B., Fenster, C.B., Lacy, R.C., Sunnucks, P., and Frankham, R. (2018). Call for a paradigm shift in the genetic management of fragmented populations. *Conserv Lett* 11, e12412.
- Robinson, J., Kyriazis, C.C., Yuan, S.C., and Lohmueller, K.E. (2023). Deleterious variation in natural populations and implications for conservation genetics. *Annu Rev Anim Biosci* 11, 93–114.
- Santiago, E., Novo, I., Pardiñas, A.F., Saura, M., Wang, J., and Caballero, A. (2020). Recent demographic history inferred by high-resolution analysis of linkage disequilibrium. *Mol Biol Evol* 37, 3642–3653.
- Sharma, H.P., Rimal, B., Zhang, M., Sharma, S., Poudyal, L.P., Maharjan, S., Kunwar, R., Kaspal, P., Bhandari, N., Baral, L., et al. (2020). Potential distribution of the critically endangered Chinese Pangolin (*Manis pentadactyla*) in different land covers of Nepal: implications for conservation. *Sustainability* 12, 1282.
- Sun, X., Liu, Y.C., Tiunov, M.P., Gimranov, D.O., Zhuang, Y., Han, Y., Driscoll, C.A., Pang, Y., Li, C., Pan, Y., et al. (2023). Ancient DNA reveals genetic admixture in China during tiger evolution. *Nat Ecol Evol* 7, 1914–1929.
- Vaser, R., Adusumalli, S., Leng, S.N., Sikic, M., and Ng, P.C. (2016). SIFT missense predictions for genomes. *Nat Protoc* 11, 1–9.
- von Seth, J., Dussex, N., Díez-del-Molino, D., van der Valk, T., Kutschera, V.E., Kierczak, M., Steiner, C.C., Liu, S., Gilbert, M.T.P., Sinding, M.H.S., et al. (2021). Genomic insights into the conservation status of the world's last remaining Sumatran rhinoceros populations. *Nat Commun* 12, 2393.
- Wang, Q., Lan, T., Li, H., Sahu, S.K., Shi, M., Zhu, Y., Han, L., Yang, S., Li, Q., Zhang, L., et al. (2022). Whole-genome resequencing of Chinese pangolins reveals a population structure and provides insights into their conservation. *Commun Biol* 5, 821.
- Wei, S., Li, Z., Momigliano, P., Fu, C., Wu, H., and Merilä, J. (2020). The roles of climate, geography and natural selection as drivers of genetic and phenotypic differentiation in a widespread amphibian *Hyla annectans* (Anura: Hylidae). *Mol Ecol* 29, 3667–3683.
- Wei, F., Huang, G., Guan, D., Fan, H., Zhou, W., Wang, D., and Hu, Y. (2022a). Digital Noah's Ark: last chance to save the endangered species. *Sci China Life Sci* 65, 2325–2327.
- Wei, S., Sun, S., Dou, H., An, F., Gao, H., Guo, C., and Hua, Y. (2022b). Influence of Pleistocene climate fluctuations on the demographic history and distribution of the critically endangered Chinese pangolin (*Manis pentadactyla*). *BMC Zool* 7, 1.
- Willi, Y., Kristensen, T.N., Sgrò, C.M., Weeks, A.R., Ørsted, M., and Hoffmann, A.A. (2022). Conservation genetics as a management tool: the five best-supported paradigms to assist the management of threatened species. *Proc Natl Acad Sci USA* 119, e2105076119.
- Wu, S.B. (2004). Assessment of threatened status of Chinese Pangolin (*Manis pentadactyla*). *Chin J Appl Environ Biol* 10, 456–461.
- Xing, Y., and Ree, R.H. (2017). Uplift-driven diversification in the Hengduan Mountains, a temperate biodiversity hotspot. *Proc Natl Acad Sci USA* 114, E3444–E3451.
- Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 88, 76–82.
- Yang, L., Chen, M., Challender, D.W.S., Waterman, C., Zhang, C., Huo, Z., Liu, H., and Luan, X. (2018). Historical data for conservation: reconstructing range changes of Chinese pangolin (*Manis pentadactyla*) in eastern China (1970–2016). *Proc R Soc B* 285, 20181084.
- Yang, L., Wei, F., Zhan, X., Fan, H., Zhao, P., Huang, G., Chang, J., Lei, Y., and Hu, Y. (2022). Evolutionary conservation genomics reveals recent speciation and local adaptation in threatened takins. *Mol Biol Evol* 39, msac111.
- Yang, S., Lan, T., Zhang, Y., Wang, Q., Li, H., Dussex, N., Sahu, S.K., Shi, M., Hu, M., Zhu, Y., et al. (2023). Genomic investigation of the Chinese alligator reveals wild-extinct genetic diversity and genomic consequences of their continuous decline. *Mol Ecol Resour* 23, 294–311.
- Zhang, C., Dong, S., Xu, J., He, W., and Yang, T. (2019). PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35, 1786–1788.
- Zhang, F., Wu, S., Zou, C., Wang, Q., Li, S., and Sun, R. (2016). A note on captive breeding and reproductive parameters of the Chinese pangolin, *Manis pentadactyla* Linnaeus, 1758. *ZooKeys* 618, 129–144.
- Zhang, H., Miller, M.P., Yang, F., Chan, H.K., Gaubert, P., Ades, G., and Fischer, G.A. (2015). Molecular tracing of confiscated pangolin scales for conservation and illegal trade monitoring in Southeast Asia. *Glob Ecol Conserv* 4, 414–422.
- Zhang, M., Cao, P., Dai, Q.Y., Wang, Y.Q., Feng, X.T., Wang, H.R., Wu, H., Min-Shan Ko, A., Mao, X.W., Liu, Y.C., et al. (2021). Comparative analysis of DNA extraction protocols for ancient soft tissue museum samples. *Zoological Res* 42, 280–286.
- Zhao, S., Zheng, P., Dong, S., Zhan, X., Wu, Q., Guo, X., Hu, Y., He, W., Zhang, S., Fan, W., et al. (2013). Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. *Nat Genet* 45, 67–71.
- Zheng, B., Xu, Q., and Shen, Y. (2002). The relationship between climate change and Quaternary glacial cycles on the Qinghai-Tibetan Plateau: review and speculation. *Quat Int* 97-98, 93–101.