






## Whole-genome resequencing of Chinese pangolins reveals a population structure and provides insights into their conservation

Qing Wang <sup>1,2,9</sup>, Tianming Lan <sup>2,3,4,9</sup>✉, Haimeng Li<sup>1,2,9</sup>, Sunil Kumar Sahu <sup>2</sup>, Minhui Shi<sup>1,2</sup>, Yixin Zhu<sup>1,2</sup>, Lei Han<sup>5</sup>, Shangchen Yang<sup>6</sup>, Qian Li<sup>1,2</sup>, Le Zhang<sup>5</sup>, Zhangwen Deng<sup>7</sup>, Huan Liu <sup>2,3,4</sup>✉ & Yan Hua <sup>8</sup>✉

Poaching and trafficking have a substantial negative impact on the population growth and range expansion of the Chinese pangolin (*Manis pentadactyla*). However, recently reported activities of Chinese pangolins in several sites of Guangdong province in China indicate a promising sign for the recovery of this threatened species. Here, we re-sequence genomes of 15 individuals and perform comprehensive population genomics analyses with previously published 22 individuals. These Chinese pangolins are found to be divided into three distinct populations. Multiple lines of evidence indicate the existence of a newly discovered population (CPA) comprises entirely of individuals from Guangdong province. The other two populations (CPB and CPC) have previously been documented. The genetic differentiation of the CPA and CPC is extremely large ( $F_{ST} = 0.541$ ), which is larger than many subspecies-level differentiations. Even for the closer CPA and CPB, their differentiation ( $F_{ST} = 0.101$ ) is still comparable with the population-level differentiation of many endangered species. Further analysis reveals that the CPA and CPB populations separate 2.5–4.0 thousand years ago (kya), and on the other hand, CPA and CPC diverge around 25–40 kya. The CPA population harbors more runs of homozygosity (ROHs) than the CPB and CPC populations, indicating that inbreeding is more prevalent in the CPA population. Although the CPC population has less mutational load than CPA and CPB populations, we predict that several Loss of Function (LoF) mutations will be translocated into the CPA or CPB populations by using the CPC as a donor population for genetic rescue. Our findings imply that the conservation of Chinese pangolins is challenging, and implementing genetic rescue among the three groups should be done with extreme caution.

<sup>1</sup> College of Life Sciences, University of Chinese Academy of Sciences, Beijing, China. <sup>2</sup> State Key Laboratory of Agricultural Genomics, BGI-Shenzhen, Shenzhen, China. <sup>3</sup> BGI Life Science Joint Research Center, Northeast Forestry University, Harbin, China. <sup>4</sup> Guangdong Provincial Key Laboratory of Genome Read and Write, BGI-Shenzhen, Shenzhen, China. <sup>5</sup> College of Wildlife and Protected Area, Northeast Forestry University, Harbin, China. <sup>6</sup> College of Life Sciences, Zhejiang University, Hangzhou, China. <sup>7</sup> Guangxi Forest Inventory and Planning Institute, Nanning, China. <sup>8</sup> Guangdong Provincial Key Laboratory of Silviculture, Protection and Utilization, Guangdong Academy of Forestry, Guangzhou, China. <sup>9</sup> These authors contributed equally: Qing Wang, Tianming Lan, Haimeng Li. ✉email: [lantianming@genomics.cn](mailto:lantianming@genomics.cn); [liuhuan@genomics.cn](mailto:liuhuan@genomics.cn); [wildlife530@hotmail.com](mailto:wildlife530@hotmail.com)

Small populations particularly tend to be crashed due to loss of genetic diversity, the accumulation of deleterious mutations, and changes in genetic make-up resulting from genetic drift and increased inbreeding<sup>1,2</sup>. Establishing gene flow between populations is an effective method to improve the fitness of isolated small populations by increasing genetic diversity and decreasing inbreeding, which is also known as the genetic rescue<sup>3–5</sup>. For example, gene rescue substantially improved the fitness of small inbred populations in African lions (*Panthera leo*)<sup>6</sup>, mountain pygmy possum (*Burramys parvus*)<sup>4</sup> and Florida panther (*Puma concolor coryi*)<sup>7</sup>. However, the making of suitable strategy for genetic rescue highly depends on comprehensive investigation of the genetic background of these small populations, including the genetic diversity, population differentiation, mutational load, gene flow, local adaptation, the extent of inbreeding and outbreeding, etc.<sup>2,8,9</sup>.

Molecular markers have ever been widely used in conservation genetics to guide the protection and conservation of threatened species, with the most promising markers of microsatellite and mitochondrial DNA<sup>10–12</sup>. However, limitations of these molecular markers are also obvious, including the limited information to explain local adaptation, inbreeding/outbreeding depression and adaptive variations<sup>13</sup>. Bias introduced into parameter estimations due to the limited genetic information is also needed to be emphasized for the use of neutral markers<sup>14,15</sup>, and the evaluation of genetic diversity of the giant panda (*Ailuropoda melanoleuca*) is a good example for this issue<sup>16</sup>. Fortunately, all the above-mentioned limitations could be much improved at the era of whole genome sequencing. With the rapid development of sequencing technology and the plummeting cost of re-sequencing, the conservation genetics is in transition to conservation genomics<sup>17</sup>. Recently, several genomic-based studies on investigation of endangered animals were reported, explaining many long-standing questions, such as the purging of deleterious mutations in Bengal tigers (*Panthera tigris tigris*)<sup>18</sup>, the local adaptation of giant panda<sup>19</sup>, the detailed inbreeding depression in kākāpō (*Strigops habroptilus*)<sup>20</sup>, which are unthinkable without application of genome-wide markers.

Pangolins belong to the placental mammal order of Pholidota, representing one of the most unusual orders of mammals due to their overlapping epidermal scales, myrmecophagous diet, lack of teeth, as well as their extraordinarily elongated tongue<sup>21,22</sup>. They play important roles in ecosystems, including predators of social insects, creators of burrows, hosts of endo- and ectoparasites, and also prey for other predators<sup>23–25</sup>. Humans, on the other hand, have extensively killed and exploited pangolins as a luxury delicacy and used their scales in traditional medicines<sup>26–28</sup>. Because of this overexploitation, all the eight extant Pholidota species have undergone severe population decline<sup>29</sup>. Of all eight pangolin species, the Chinese pangolin (*Manis pentadactyla*) is one of the most threatened species with a ~94% decline in its whole population between the 1960s and 1990s, largely due to the extensively illegal trade in Asia<sup>30,31</sup>. It has been classified as a critically endangered species in the IUCN Red List of Threatened Species and listed in ‘Appendix I’ of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES)<sup>32</sup>. This species was once distributed in the vast areas of the southern Yangtze River in China and some northern areas of Southeast Asia<sup>30,33,34</sup>. However, new burrows have not been seen for more than 20 years in some areas of the Dawuling Natural Reserve and Luofushan Natural Reserve in Guangdong province<sup>35,36</sup>. Moreover, the Chinese pangolin might have been extirpated from several areas, including Jiangsu, Shanghai, Henan, etc<sup>37,38</sup>.

A large majority of research on pangolins is restricted to ecological study or genetic study but with DNA markers limited at mitochondrial DNA and microsatellite fragment<sup>38–40</sup>. Recently, Hu et al.<sup>41</sup> reported the population identities of illegally traded

individuals, revealed population fluctuations, and an increase in inbreeding and mutation load in Chinese pangolin populations, providing valuable information for the conservation of this species. Further, great progress has been made towards the conservation of Chinese pangolin, with several Chinese pangolins being photographed by infrared cameras in the wild in at least six cities of Guangdong province since 2020<sup>42</sup>, which was a promising sign for the rescue and recovery of this species. Although, Hu et al.’s study included samples from Taiwan<sup>43</sup>, Yunnan province, and some unknown geographical locations, more samples from Guangdong province are needed for conjoint analysis to reveal more comprehensive genomic make-ups of this critically endangered species.

The planning of genetic rescue necessitates a thorough assessment of genetic background from a variety of perspectives, which is the prime objective of our research to aid in the continued conservation of Chinese pangolins. In this study, we intensively investigated the genomic characteristics of Chinese pangolins from Guangdong province. Together with previously published genomes, we revealed the population structure of the Chinese pangolins, and extensively explored the genetic diversity, population differentiation, mutational load, gene flow and local adaptations of each population. Our study provides a valuable resource and insights for the future conservation and conservation of this endangered species.

## Results

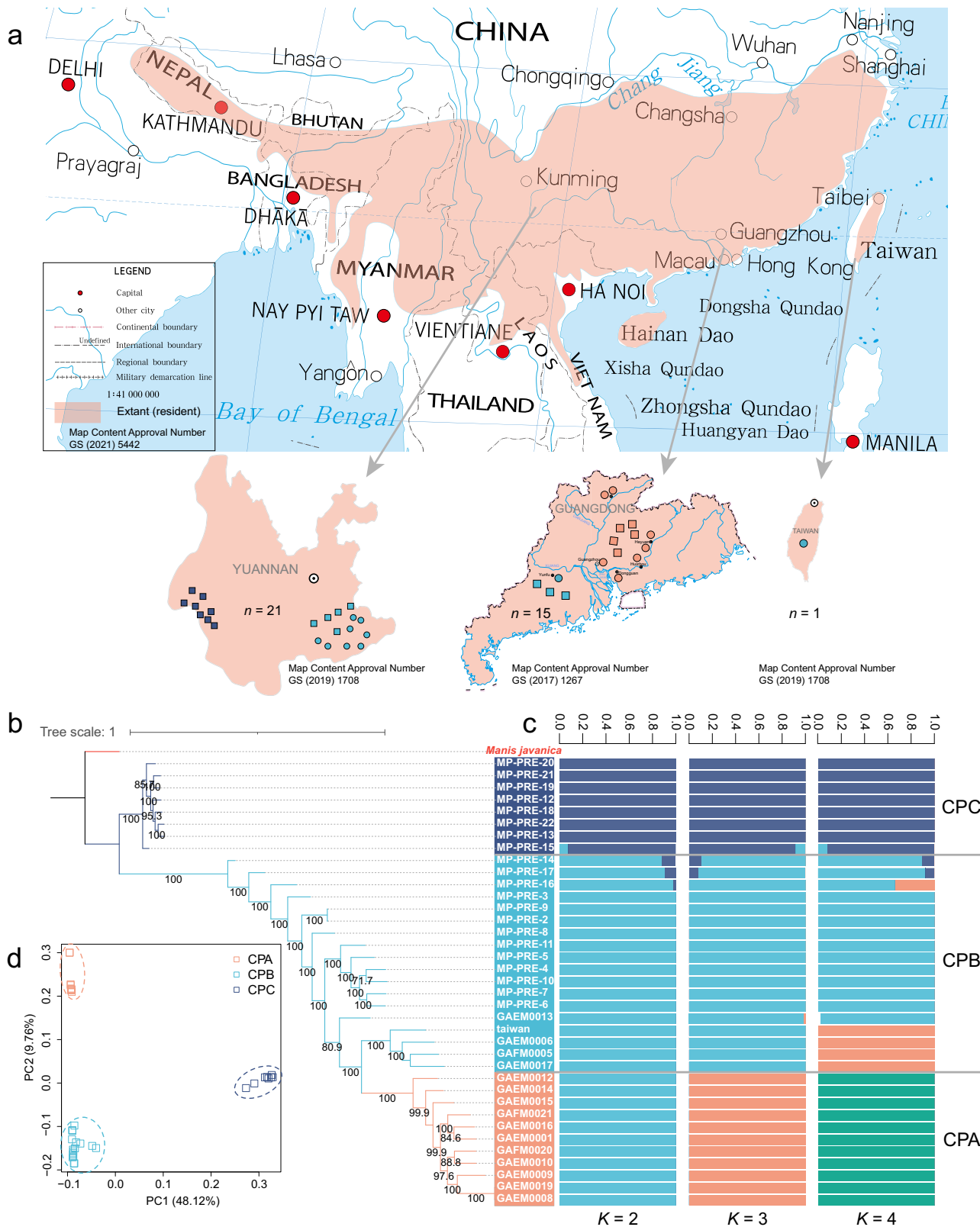
**Characteristics of sequencing data and variants.** The sequencing depth of 15 Chinese pangolin individuals ranged from 13.17 X to 22.85 X, with an average depth of  $17.03 \pm 2.43$  X (Supplementary Table 1). Together with the published genome sequencing data from other 22 individuals, we obtained 40,855,034 raw single-nucleotide polymorphisms (SNPs) for the whole pangolin population. After deep filtration, 35,023,399 SNPs were retained for further analysis. As expected, most of the SNPs were distributed in the intergenic regions. The distribution of SNPs against the minor allele frequency (MAF) showed that the proportion of SNPs with a minor allele frequency less than 5% (20,155,038 SNPs) was 49.3%, representing a large proportion of low-frequency variants in the pangolin population (Supplementary Fig. 1). In addition, we also observed significant variability in the SNP number of different populations ( $P = 7.5 \times 10^{-11} \sim 1.1 \times 10^{-12}$ ) (Supplementary Fig. 2).

## Genetic relationships among Chinese pangolin populations.

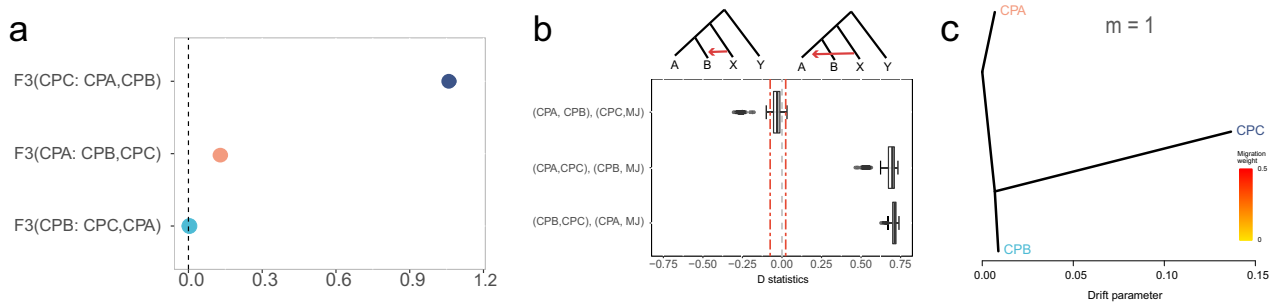
We first constructed phylogenetic trees with nuclear genomes and mitochondrial gene sequences and confirmed that these Guangdong individuals are Chinese pangolins (Supplementary Figs. 3–4). We then extensively explored the genetic relationships of 37 Chinese pangolins in the whole-genome level. In general, results from principal component analysis (PCA), admixture and phylogenetic tree supported that CPA, CPB and CPC populations were assigned to three distinct clusters (Fig. 1b–d). Fifteen individuals from Guangdong province were distributed into two clusters, with one formed by only Guangdong individuals (CPA:  $n = 11$ ) and the other one (CPB:  $n = 18$ ) consisting of individuals from Guangdong ( $n = 4$ ), Taiwan ( $n = 1$ ) and Yunnan ( $n = 13$ ). CPC population ( $n = 8$ ) was made up of 8 individuals who might have originated from Myanmar<sup>41</sup>.

## Genetic differentiation and Gene flow among populations.

Genome-wide population differentiation further revealed that the CPA was distinct from the CPC with an extremely high fixation index ( $F_{ST}$ ) value of 0.541. The  $F_{ST}$  between the closer CPA and CPB populations was still larger than 0.1 (Supplementary Fig. 5). The  $F_3$  statistics also supported that none of these three pangolin



**Fig. 1** Distribution and population structure of Chinese pangolins in southern China. **a** Distribution of extant Chinese pangolins and 37 individuals in the current study. The distribution range of extant Chinese pangolins was from IUCN<sup>29</sup>. Pink regions represent areas where Chinese pangolins have been recorded. Circles represent samples with known origins. Squares represent the samples of unknown geographical information. CPA ( $n = 11$ ), CPB ( $n = 18$ ) and CPC ( $n = 8$ ) was in orange, blue and purple, respectively. **b** Maximum likelihood phylogenetic tree of Chinese pangolin populations. Numbers at tree nodes represented the bootstrap supports (1000 replications). The Malayan pangolin, CPA, CPB and CPC are in red, orange, blue and purple, respectively. **c** Population structure inferred using ADMIXTURE analysis from  $K = 2$  to  $K = 4$ . Clusters were noted by different colors, and each bar represented an individual. **d** Principal component analysis of all 37 individuals.



**Fig. 2 Genetic differentiation and gene flow analysis among Chinese pangolins. a** F3 statistics for CPA, CPB and CPC. **b** D-statistics: A, B, X were among each of three Chinese pangolin populations and the Malayan pangolin (MJ) was set as the outgroup (Y). The red dashed lines represent the threshold value of D that is significantly deviating from 0. **c** TreeMix trees with no migration events ( $m = 1$ ).

populations was admixed by the other two populations with positive scores in each combination (Fig. 2a).

To explore possible genetic exchanges among these Chinese pangolin populations, we performed ABBA-BABA test to quantify the shared derived alleles between populations. We set the Malayan pangolin (*Manis javanica*) as the Y (the outgroup). A, B and X were from all possible combinations of the CPA, CPB and CPC populations. Interestingly, when the X population was set to be the CPC population, we were unable to detect any significantly unbalanced derived allele sharing, supporting the large differentiation between CPC and other two populations (Fig. 2b). We also performed a TreeMix analysis to further quantify gene flow among populations. However, no migration events found among the three populations with tests from  $m = 1$  to  $m = 10$  (Fig. 2c).

Further, identity-by-descent (IBD) analysis showed no large segments ( $>1$  Mb) among populations. For medium-size ( $100 \text{ kb} < \text{IBD} < 1 \text{ Mb}$ ) segments, we identified 0.81 Mb shared between CPA and CPB, 0 Mb shared between CPA and CPC, and 4.44 Mb shared between CPB and CPC (Supplementary Fig. 6). However, much more and larger IBD segments were found within populations (Supplementary Fig. 7). The total length of IBD in CPA, CPB and CPC were 251.75 Mb (10.49%), 51.45 Mb (2.14%) and 48.38 Mb (2.02%), respectively.

**Population separation among CPA, CPB and CPC populations.** Considering multiple lines of evidence from population structure and gene flow, we inferred that the separations between CPA and CPB were more recent than that of CPC and CPA/CPB. As we expected, the relative cross coalescence rate (RCCR) estimation showed that the separation between CPA and CPB was 2.5 to 4.0 thousand years ago (kya), which was much more recent than that between CPC and CPA (25–40 kya), and CPC and CPB (25–40 kya) (Fig. 3a). Considering the possible phasing errors which could influence the accuracy of the MSMC2 method, we further performed SMC++ to validate results from MSMC2. Again, we found that the CPA and CPB populations were separated at  $\sim 3.4$  kya, CPC was separated from CPA and CPB at  $\sim 25$  kya (Fig. 3b), which supported results from MSMC2.

**Genetic diversity and inbreeding.** The richness of genetic diversity in a population reflects its evolutionary potentials to adapt to environmental changes<sup>44</sup>. The average heterozygosity ( $H_e$ ) among all the individuals was 0.18% calculated by all SNPs. This value decreased to 0.15% when low-frequency alleles were excluded (allele frequency (AF)  $< 5\%$ ). To better compare our results to the previous report, we also calculated the average  $H_e$  (0.11%) using SNPs with AF  $> 20\%$  (Supplementary Fig. 8a). Similar patterns were also found in nucleotide diversity ( $\pi$ ), with

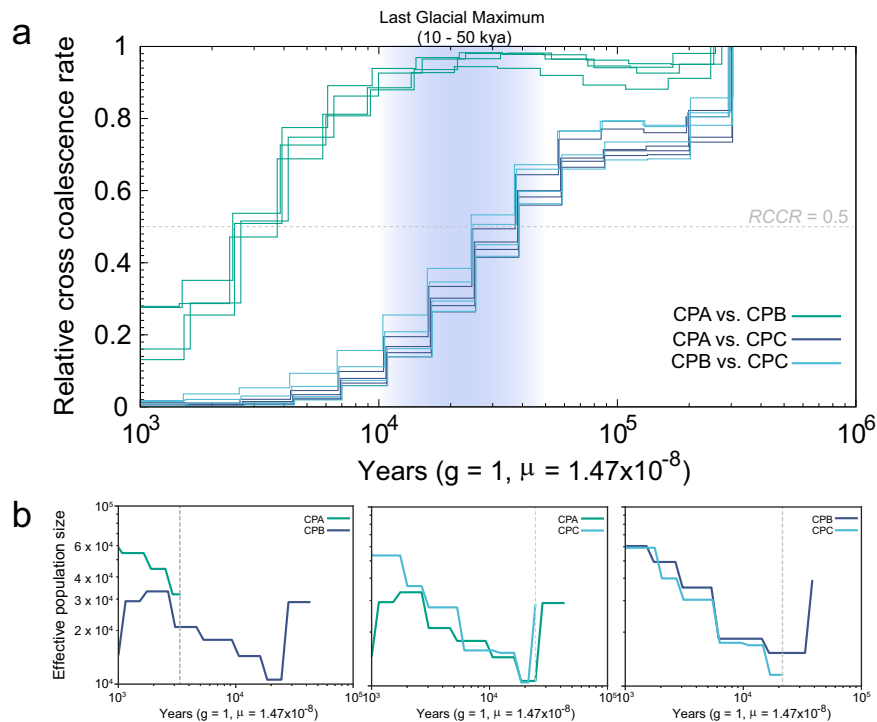
the CPB population having the highest level of genetic diversity (Supplementary Fig. 8b).

By using the same method from Hu et al.<sup>41</sup>, we found 77,251 runs of homozygosity (ROHs) ranging from 100.0 kb to 3155.8 kb in the whole Chinese pangolin population. Of these ROHs, 448.58 Mb ( $F_{\text{ROH}} = 18.69\%$ ) and 21.57 Mb ( $F_{\text{ROH}} = 0.90\%$ ) were assigned to medium-size (100 kb–1 Mb) and long ROH ( $> 1$  Mb) (Fig. 4c, d; Supplementary Figs. 8c, d). The medium-size ROHs in CPA, CPB, and CPC population was 469.38 Mb ( $F_{\text{ROH}} = 19.56\%$ ), 451.63 Mb ( $F_{\text{ROH}} = 18.82\%$ ), and 413.52 Mb ( $F_{\text{ROH}} = 17.23\%$ ), respectively. The long-size ROHs in CPA, CPB, and CPC population was 29.54 Mb ( $F_{\text{ROH}} = 1.23\%$ ), 22.28 Mb ( $F_{\text{ROH}} = 0.93\%$ ), and 9.11 Mb ( $F_{\text{ROH}} = 0.38\%$ ), respectively.

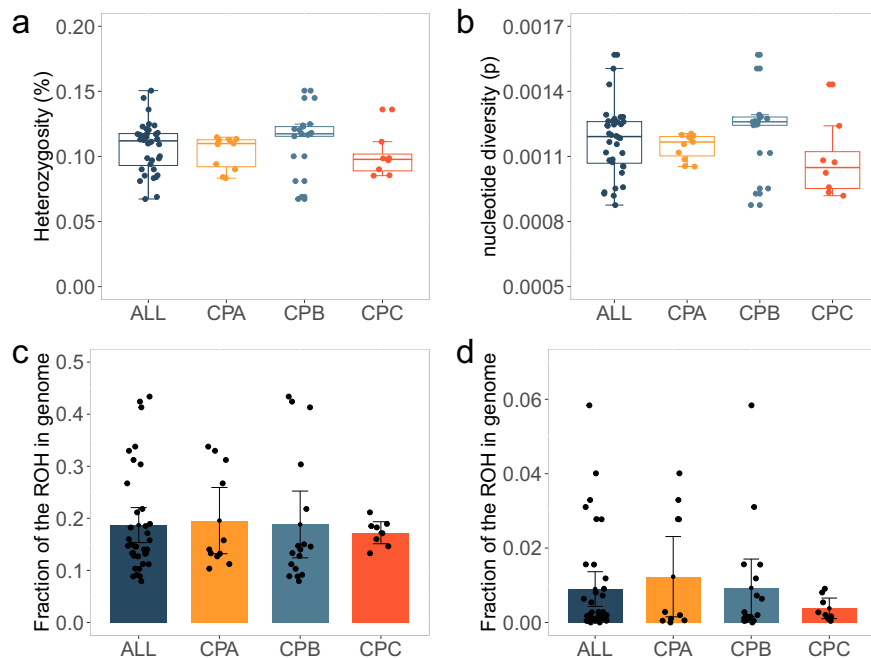
**Mutational load.** On average, we found  $266 \pm 55$  deleterious nonsynonymous SNPs (nsSNPs) across all three populations (Supplementary Fig. 9a). Individuals in the CPC population ( $169 \pm 15$ ) exhibited a significantly smaller number of deleterious nsSNPs when compared to CPA ( $287 \pm 15$ ;  $P = 1.7 \times 10^{-10}$ ) and CPB ( $298 \pm 20$ ;  $P = 8.5 \times 10^{-12}$ ) populations. However, there was no significant difference in the ratio of deleterious nsSNPs to total nsSNPs across all populations ( $P = 0.8\sim 0.9$ ; Supplementary Fig. 9b). We also considered the number of Loss of Function (LoF) variants in each population (from 785 to 1250; Supplementary Fig. 9c). The number of LoF variants in CPC population was significantly lower than CPA and CPB populations ( $P = 1.4 \times 10^{-7}$  and  $P = 5.5 \times 10^{-9}$ ). However, the CPC population had a significantly higher ratio of LoF variants to total nsSNPs than CPA and CPB populations ( $P = 8.5 \times 10^{-8}$  and  $P = 2.9 \times 10^{-8}$ ; Supplementary Fig. 9d).

We further evaluated the level of mutational load by measuring the number of homozygous sites to the homozygous and heterozygous sites (hereafter ratio) for LoF mutation, missense mutation, and deleterious missense mutations. We found that this ratio of LoF mutation, missense mutation, and deleterious missense mutations in the 36 Chinese pangolins was 0.505, 0.482 and 0.432, respectively (Fig. 5a). The ratio of LoF mutation (0.545, 0.522), missense mutation (0.575, 0.540), and deleterious missense mutations (0.533, 0.485) in CPA and CPB population were very similar, but significantly higher than that in the CPC population (0.412 for LoF mutation, 0.232 for missense mutation, and 0.178 for deleterious missense mutations) (Fig. 5a, Supplementary Table 4).

We further evaluated the number of potentially introduced LoF mutations to the receiving population from the donor population. We found that CPB individuals would introduce 120–238 LoF mutations and CPC individuals would introduce 336–381 LoF mutations when CPA was regarded as the receiving population. The minimum LoF mutations (103–135) would be introduced to the CPB population from the CPA population. However, this



**Fig. 3 Population separation inferences in each pair of Chinese pangolin populations.** **a** The split time of CPA vs. CPB, CPA vs. CPC, and CPB vs. CPC populations was calculated by the MSMC2 software. Each line indicated a pair of separation inferences by randomly selecting two individuals. The grey dashed line represented that relative cross coalescence rate (RCCR) of 0.5. The Last Glacial Maximum period is indicated by purple vertical bars. **b** The split time between CPA vs. CPB, CPA vs. CPC, and CPB vs. CPC populations inferred by the SMC++ software. The lines represent the CPA (green), CPB (purple) and CPC (blue) populations. The grey dashed line denotes the divergence between two populations.

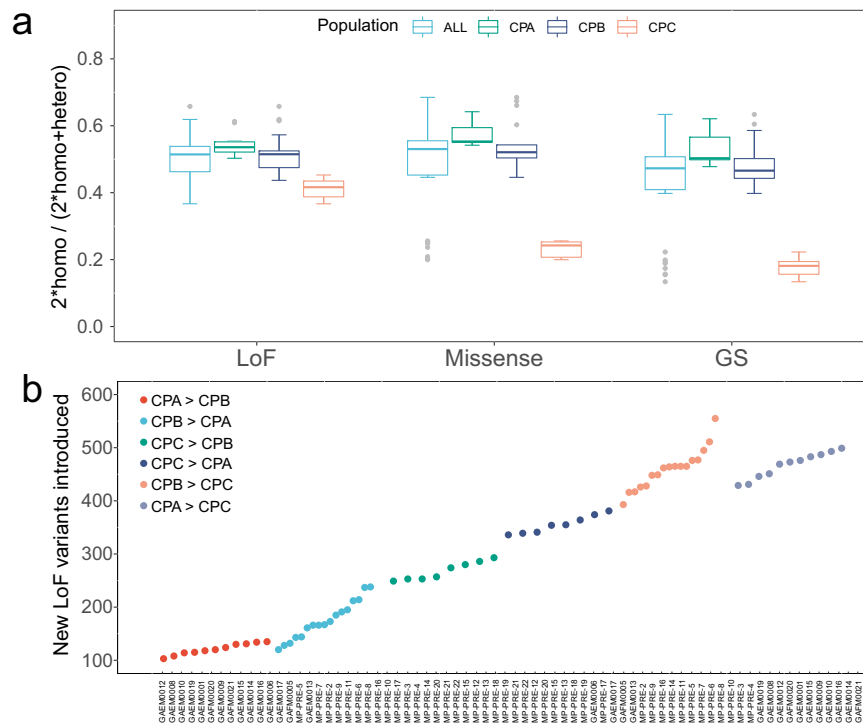


**Fig. 4 Genetic diversity and ROH distribution in the Chinese pangolin genome.** **a** Heterozygosity in each population. **b** Nucleotide diversity in each population. Error bars show range of values within 1.5 times the interquartile range. **c, d** Fraction of run of homozygosity (ROH) in the genome of each population, including the medium (**c**: 100 kb–1 Mb) and long (**d**: > 1 Mb) ROHs. Data are presented as mean  $\pm$  SD. ALL:  $n = 36$ . CPA:  $n = 11$ . CPB:  $n = 17$ . CPC:  $n = 8$ .

number rise to 393 to 555 by introducing CPA and CPB individuals to the CPC population (Fig. 5b).

We then analyzed the 2229, 1044, 1574 and 962 genes affected by the LoF mutations in the all 36 individuals, CPA, CPB and CPC populations. In general, significantly enriched Gene Ontology (GO)

categories and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways indicated that these genes were associated with male gamete generation (GO: 0048232;  $P = 3.8 \times 10^{-13}$ ), visual perception (GO:0007601;  $P = 8.8 \times 10^{-10}$ ) and protein digestion and absorption (KEGG: hsa04974;  $P = 3.0 \times 10^{-6}$ ) (Supplementary



**Fig. 5 Mutation load estimation for the pangolin population.** **a** The derived mutations in the coding region include Loss of Function (LoF) mutations, missense mutations, and deleterious missense mutations (GS). Error bars show range of values within 1.5 times the interquartile range. ALL:  $n = 36$ . CPA:  $n = 11$ . CPB:  $n = 17$ . CPC:  $n = 8$ . **b** The predicted number of new LoF variants introduced between individuals from different populations if translocation took place.

Tables 5–8). The KEGG pathway “protein digestion and absorption ( $P < 0.01$ )” was found to be significantly enriched in all three populations. In addition, the Insulin signaling pathway (KEGG: hsa04910;  $P = 6.4 \times 10^{-5}$ ), cellular carbohydrate metabolic process (GO: 0044262;  $P = 1.9 \times 10^{-4}$ ) and response to fructose (GO:0009750;  $P = 2.0 \times 10^{-4}$ ) were found in the CPB population. In particular, we did not find enriched gametogenesis-related pathways, such as male gamete generation, in the CPC population.

**Recent positive selection in Chinese pangolin populations.** A total of 2218 (CPA), 3154 (CPB) and 1878 (CPC) SNPs were identified under strong positive selection with  $iHS$  scores in the top 0.1% (Supplementary Fig. 10). By filtering out SNPs in the non-gene regions, 176, 299 and 277 genes were located in the positively selected regions for CPA, CPB and CPC populations, respectively. Eleven genes were shared among these three populations, but 130, 226 and 219 genes were unique to CPA, CPB and CPC populations, respectively. Enrichment analysis on these selected genes showed that 14, 13 and 16 GO categories or KEGG pathways were unique for CPA, CPB and CPC populations, respectively. These categories were correlated with various biological functions including cellular processes, biological regulation and metabolic processes (Supplementary Fig. 11).

## Discussion

Here we report the, to the best of our knowledge, first genetic survey at the whole genome scale for the Chinese pangolin from Guangdong province of China. We thoroughly examined the population structure of Chinese pangolins by combining Guangdong samples and published sequences, which presumably cover the majority of the distribution area of this species. We systematically investigated genomic make-ups of long-term declined Chinese pangolins, especially for the Guangdong population, including the genetic diversity, ROH and mutational load

assessment, which is expected to support the future conservation and genetic rescue of this species.

Although two pangolin individuals in Hu et al.’s study (MP09 and MP10) were mentioned to be possibly from Guangdong province, the detailed population structure is hard to be further described due to the ambiguous sampling site and small sample size<sup>41</sup>. Benefiting from the relatively large sample size collected from several sites in Guangdong province, we defined a new Chinese pangolin population, CPA, in our study from the perspective of genomics, with multiple lines of evidence (Fig. 1b–d). Negligible gene flow further strengthens the independent state of the CPA population (Fig. 2).

A very large genetic differentiation between this CPA and CPC ( $F_{ST} = 0.541$ ) population suggested a potential subspecies-level divergence when compared with other animals. For example, the Sichuan and Qinling giant panda subspecies was defined in 2005 by the skull and molar size<sup>45</sup>, and their genomic differentiation is much less than the CPA and CPC populations ( $F_{ST} = 0.140$ )<sup>16</sup>. The differentiation of tiger subspecies could be dating back to 110 kya<sup>46</sup>, however, the genomic differentiations among tiger subspecies are still much lower than these two Chinese pangolin populations, with the highest  $F_{ST}$  value between Amur tiger and Sumatran tigers is of 0.318<sup>47</sup>. This was not surprising, because the CPC population was significantly distinct from CPA and CPB from the perspective of genetics (Supplementary Fig. 5), which was also found in Hu et al.’s study with the  $F_{ST}$  of 0.495 between MPA and MPB<sup>41</sup>.

Interestingly, the genomic differentiation between CPA and CPB was moderate with a  $F_{ST}$  value of 0.101. Although this differentiation was much lower than that between CPA and CPC, this extent was comparable with differentiations between populations with extensive geographic separations, like the divergence of African leopard (*Panthera pardus*) populations with the average  $F_{ST}$  value of 0.104<sup>48</sup>, the two largest isolated populations

(Minshan and Qionglai populations) of the giant pandas<sup>16</sup>, and even comparable to that between African and Asian humans ( $F_{ST} = 0.120$ )<sup>49</sup> which are known to be separated before ~ 55 kya<sup>50</sup>. Taken together, we infer that the first and newly discovered CPA was a distinct Chinese pangolin population, and suggested an extra conservation unit that is parallel with CPB and CPC population. We believe the future large-scale population genomic analysis and corresponding ecological and morphological studies will further enrich and strengthen this claim.

We inferred the separation between CPA and CPB was much more recent than that between CPC and CPA/CPB (Fig. 3a), which is logically consistent with population structure analysis. The separation between CPC and CPA/CPB was coincided closely with the last glacial period (LGP, ~10–120 kya), especially the beginning of the last glacial maximum (LGM, ~10–50 kya)<sup>51,52</sup>. The extremely cold weather and the possibly continuous impact on the prey of Chinese pangolins during this period could be one of the reasons for population separation, considering that southern China was also severely influenced by the LGP<sup>53</sup>. However, climate changes may not be the main reason for population separation<sup>19,54,55</sup>. Human activity is always considered an important factor for shaping the evolution of animals<sup>9,36,56</sup>. The human expansion in East Asia occurred as early as 40 kya<sup>57</sup>, and frequent migration, admixture and replacement also occurred at the original distributed area of the Chinese pangolins<sup>41</sup>. Interestingly, the Han Chinese population began to expand and separated from the European human population at ~30 kya<sup>58,59</sup>. Moreover, the key period in human history was the onset of the Holocene since the development of more favorable climate conditions<sup>60</sup>. We concluded that human activity was most likely responsible for the separation of the CPA and CPB populations.

Regarding the genetic status of Chinese pangolin in this study, we observed that genome-wide *He* among the three populations are similar ( $0.107\% \pm 0.086\%$ ;  $P = 0.058\sim 0.755$ ), but lower than some other critically endangered species that are suffering from long-term population decline as the Chinese pangolin, such as Northern white rhinoceros (*Ceratotherium simum cottoni*) (0.110%) and Western lowland gorilla (*Gorilla gorilla gorilla*) (0.144%)<sup>2,61</sup>. Such a low level of genetic diversity suggested that Chinese pangolins likely had a relatively low adaptive potential<sup>62</sup>. The inbreeding often contributes a lot to the decrease of genetic diversity for endangered species, such as the ROH segments of highly inbred grey wolves (*Canis lupus*) population ranging from 2,695 bp to 95.8 Mb<sup>63</sup> and the modern Malay Peninsula population of Sumatran rhinoceros (*Dicerorhinus sumatrensis*) estimating 30% of the genomes contain longer ROH segments ( $\geq 2$  Mb)<sup>8</sup>. However, ROHs in the Chinese pangolins are found to be much shorter than the above-mentioned endangered species. We infer that the low genetic diversity in the Chinese pangolin may not from the extensively recent inbreeding due to the absence of large size ROHs<sup>64</sup>, but the long-term isolation without frequent gene flow between subpopulations could be an alternate explanation. Here we cannot exclude the bias introduced from the reference genome, because the accurate evaluation of ROH highly depends on the contiguity of the reference genome<sup>65</sup>, but no long-read assembled high-quality reference genomes are available now.

The decrease of fitness in a small population can accelerate by the accumulation of deleterious mutations<sup>66,67</sup>. When focusing on Guangdong populations, the ratio of deleterious mutations in CPA and CPB populations was higher than CPC population. LoF mutations enriched in pathways related to male gamete generation in the CPA and CPB populations, but not CPC population, further indicated more genetic burdens in Guangdong populations, considering the impaired gamete quality may reduce reproductive success and further affect species fitness and

survival<sup>68</sup>. Together with the overall distribution of deleterious mutations, we speculated that the CPC population has higher fitness than both CPA and CPB populations.

LoF mutations were also found significantly enriched in the protein digestion and absorption pathway ( $P < 0.01$ ) in all the three populations, revealing a potentially weak ability of digestion and absorption of protein in the whole Chinese pangolin population. In the CPB population, we also found enriched KEGG pathways related to carbohydrates metabolism, indicating an inferior performance in the energy supply. Considering that pangolins enjoy high protein, high fat, high calorie food<sup>69,70</sup>, the possibly weakened ability to use the protein and carbohydrates may further lower the fitness of the Chinese pangolins.

If a certain species has several isolated populations under different environments, local adaptation in different directions could occur in different populations<sup>71</sup>. Genetic rescue by populations with large differentiation shaped by natural selection tends to be disturbed by outbreeding depression<sup>72</sup>. To better identify the local adaptation in each population, we performed enrichment analysis of genes under recent natural selection in the three populations. A large proportion of enriched pathways were unique for the three populations (Supplementary Fig. 11), indicating an obvious adaptive difference in CPA, CPB and CPC populations. It was worth noting that much of the Yunnan province lies within the subtropical highland, while Guangdong province faces the South China Sea to the south with a humid subtropical climate<sup>73</sup>. We inferred that the local adaptation in a different direction might have occurred in these three populations. However, more research works are needed to explore how these genes under natural selection affect the survival of this species.

In addition, we predicted an average of 309 LoF mutations could be introduced into other populations as new deleterious alleles when conducted translocations among CPA, CPB and CPC populations (Fig. 5b). The number was several dozen times higher than that predicted among three Sumatran rhinoceros populations (an average of 10 new LoF variants)<sup>8</sup>. We also noticed that more LoF variants could be mutually introduced between CPA/CPB and the CPC population, which is consistent with the result from local adaptation, possibly due to the long-term isolation and adaptation to their local habitats (Yunnan and Guangdong). We cannot conclude that it is 100% not suitable to implement genetic rescue among these Chinese pangolin populations, however, genetic rescue between CPA and CPB should be safer from the perspective of outbreeding depression. The conclusion will be more convincing only if the exact effects of these LoF mutations on genes could be parsed in future works.

We should keep in mind that genetic rescue is a complicated process<sup>4,5,74</sup>, hence more reliable evidence is still needed to be put forward in further studies. Furthermore, as a distinct population separated from the CPB and CPC, the newly discovered CPA Chinese pangolin population in our study should be given high priority in future conservation works of the Chinese pangolin.

## Methods

**Samples and data collection, library preparation, and sequencing.** Approvals of all necessary research ethics and permits were granted by the Institutional Review Board of BGI (BGI-IRB E21056). 15 Chinese pangolin samples were collected from the Guangzhou Wildlife Rescue Center, Guangdong province, China for whole genome sequencing. All samples were taken from rescue individuals dead of natural causes or confiscated individuals from forestry police, and all individuals were identified as Chinese pangolins through distinct morphological characteristics. In addition, sequencing data of 22 Chinese pangolin individuals (BioProject: PRJNA529540 and PRJNA20331; Supplementary Table 2) and 72 Malayan pangolin individuals were downloaded from NCBI (BioProject: PRJNA529540; Supplementary Table 3) for downstream analysis in this study.

Genomic DNA was isolated using standard phenol/chloroform-isoamylalcohol extraction<sup>75</sup> and the precipitate was solved in 20–100  $\mu$ l distilled water. ~1  $\mu$ g

genomic DNA was used and sheared into fragments with 200 to 800 base pairs (bp) for paired-end (PE) DNA library construction with the insert size of ~350 bp following the manufacturer's instructions of BGISEQ platform (BGI, Shenzhen, China). DNA libraries were then subjected to the DNBSEQ-T1 sequencer for sequencing.

**Genome mapping, variants calling and filtering.** Sequencing reads were mapped to *M. pentadactyla* reference genome (YNU\_ManPten\_2.0, Genbank: GCA\_014570555.1) and *M. javanica* (YNU\_ManJav\_2.0, Genbank: GCA\_014570535.1)<sup>41</sup> by using the Burrows-Wheeler Aligner (BWA) with the *mem* algorithm<sup>76</sup> (version: 0.7.10-r789) using the default parameters. Sorting, reordering and reads deduplication were performed by the Picard tools (<http://picard.sourceforge.net>) (version: 2.1.1). Haplotype-Caller implanted in the Genome Analysis Toolkit (GATK, version: 3.3.0-g37228a)<sup>77</sup> was used for raw variants calling. Hard filtering was performed on the raw SNP variant set with parameters of “QUAL < 30.0 || QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0”. The final high-quality SNP set was used for further genetic analyses. Lastly, SNPs in scaffolds with length smaller than 100 kb were excluded.

SNPs were annotated by the software ANNOVAR<sup>78</sup> (version: 2015-12-14) with the genome of *M. pentadactyla* reference genome (YNU\_ManPten\_2.0). We sorted out the basic information of the resequencing data of each population (CPA, CPB and CPC) and grouped SNPs in different categories by results from ANNOVAR, including exotic, nonsynonymous, synonymous, UTR, intronic, intergenic, splicing, and non-coding RNA (ncRNA). In addition, we draw the distribution of MAF for all individuals. We prepared three SNP sets by controlling the filtration on the allele frequency for downstream analysis, including: 1) SNPs without any filtration of allele frequency; 2) SNPs with the allele frequency larger than 5%; 3) SNPs with the allele frequency larger than 20% following the method of Hu et al.<sup>41</sup>.

**Phylogenetic tree, PCA and admixture analysis.** To avoid closely-linked sites, PLINK<sup>79</sup> (version: 1.90b3.38) was used to produce a pruned subset of SNPs by Linkage disequilibrium (LD) values, resulting in a set of 171,405 SNPs. This SNP set was then converted into PHYLIP-format file by using an in-house Python script to construct the Maximum-Likelihood (ML) phylogenetic tree with 1000 replications by the PhyML<sup>80</sup> (version: 20151018) with all 37 samples. The Malayan pangolin (YNU\_ManJav\_2.0) was used as an outgroup.

For species identification, the Chinese pangolin genome (YNU\_ManPten\_2.0) and Malayan pangolin genome (YNU\_ManJav\_2.0) was regarded as the reference genome for 109 pangolins (37 Chinese pangolin individuals and 72 Malayan pangolin individuals), respectively. We then constructed two ML phylogenetic trees according to the above methods. We also extracted the cytochrome *b* (Cyt *b*) and cytochrome *c* oxidase subunit 1 (CO1) sequences from *Felis catus*: NC\_001700.1 and 8 species of *Manis* (*Manis tricuspis*: NC\_026780.1, *Manis tetradactyla*: MG196299.1, *Manis gigantea*: MG196303.1, *Manis temminckii*: KP306516.1, *M. pentadactyla*: MG196307.1, *Manis crassicaudata*: NC\_036433.1, *Manis culionensis*: NC\_036434.1, and *M. javanica*: KT445979.1), 37 Chinese pangolin individuals and 72 Malayan pangolin individuals. ML trees with 1000 replications based on mitochondrial gene sequences were constructed using Molecular Evolutionary Genetics Analysis (MEGA, version: 10) with the Kimura's two-parameter model<sup>81</sup>.

All SNPs without pruning and AF filtering were used as the initial input data for PCA and ADMIXTURE analysis. The Genome-wide Complex Trait Analysis<sup>82</sup> (GCTA, version: 1.92.2) software was used for PCA inference. The population genetic structure was then inferred by using the same dataset as PCA analysis with the ADMIXTURE<sup>83</sup> (version: 1.3) program. We predefined the number of genetic clusters (K) from 1 to 10 and ran the cross-validation error (CV) procedure to explore the best K, using the default parameters and settings.

**Population differentiation and gene flow.** We used Weir and Cockerham's  $F_{ST}$ <sup>84</sup> to estimate the population differentiation. All bi-allelic SNPs were used for the calculation of genome-wide  $F_{ST}$  between each pair of the populations using the VCFtools (version: 0.1.13) software<sup>85</sup>.

In order to explore whether one population was admixed by the other two populations, we performed the F3 test with combinations of CPA, CPB and CPC using the ‘qp3Pop’ in the ADMIXTOOLS<sup>86</sup> (version: 5.1). The EIGENSTRAT format input data was generated by CONVERTF program in the ADMIXTOOLS.

To examine the excess of shared derived alleles between different populations of Chinese pangolins, we applied the classic ABBA-BABA test (D statistics)<sup>87</sup> using the “-informative” command of the POPSTATS<sup>88</sup>. The four populations were set to be ((A, B), (X, Y)), and the A, B and X were among each of CPA, CPB, and CPC. The Y was Malayan pangolin which was the outgroup. We screened significant D values using the Z-score ( $|Z| > 3$ ) based on a block jackknife procedure.

The sharing of identity by descent (IBD) between individuals was calculated by the RefinedIBD<sup>89</sup> (version: 16May19. ad5) with analysis parameters (length = 0.1 and lod = 3.0). We compared different lengths of IBD (IBD > 1 Mb or 100 kb < IBD < 1 Mb) and their percentage on the genome among populations to evaluate the degree of gene flow.

We constructed maximum-likelihood population trees using TreeMix<sup>90</sup> (version: 1.13) to investigate the phylogenetic relationship in the presence of

admixture events among populations. TreeMix was run with the parameters -bootstrap 5000 -global and the migration event -m (from 0 to 10).

**Population separation among pangolin populations.** We first used MSMC2<sup>99</sup> (version: 2.1.2) to infer the separation among the three populations. MSMC2 was performed for four independent replications with two samples randomly selected from each population. Genotype phasing was using the Beagle<sup>91</sup> (version: 5.0) software with default parameters before the MSMC2 inference. Parameters for MSMC2 calculations were as follow: -skipAmbiguous -I 0-4, 0-5, 0-6, 0-7, 1-4, 1-5, 1-6, 1-7, 2-4, 2-5, 2-6, 2-7, 3-4, 3-5, 3-6, 3-7 -i 20 -t 6 -p ‘10\*1 + 15\*2’. To further validate the results inferred by the MSMC2, we performed the SMC++<sup>92</sup> (version: 1.5.2) to infer the split time of the two populations, because the SMC++ software does not depend on phasing data, which can avoid calculation bias introduced by switch errors during phasing analysis. The mutation rate and generation interval of the *M. pentadactyla* we used here was  $1.47 \times 10^{-8}$  41,43 per site per generation and one year<sup>41,93</sup>.

**Genetic diversity and ROH analysis.** VCFtools was used to estimate whole-genome genetic diversity, including heterozygosity (*He*) and nucleotide diversity ( $\pi$ )<sup>94</sup>. ROH was identified by PLINK following the method described by Dobrynin et al.<sup>95</sup>. Long ROH is often the indicator of recent inbreeding that occurred several decades ago. According to the formula reported by Kardos et al. (Generations =  $100/2 * ROH_{length}$ )<sup>41,63</sup>, we only counted ROHs that were larger than 100 kb in this study. If ROHs were longer than 1 Mb, we assumed that these ROHs were generated by more recent inbreeding (< 50 years). The genetic diversity, ROH and mutation load of Taiwan individuals have been studied in the study by Hu et al.<sup>41</sup>. Therefore, Taiwan individuals were excluded from this part and mutational load analysis.

**Mutational load.** We used the genotypes of the same alleles in the *M. javanica* to represent the ancestral state before identifying derived mutational loads. A deleterious mutation we used here means that an amino acid change in a protein was predicted to be harmful to the function, which becomes the main genetic basis of inbreeding depression<sup>96</sup>. The deleteriousness of derived mutations was diagnosed using the Grantham score (GS)<sup>97</sup>. Here, nsSNPs with GS  $\geq 150$  were defined as deleterious mutations<sup>97,98</sup>. We used “-aamatrixfile grantham matrix” parameter in the package ANNOVAR to print out GS for nonsynonymous variants. We counted the number of deleterious nsSNPs and the ratio of deleterious nsSNPs to total nsSNPs. Moreover, we selected derived mutations in coding regions of each pangolin individual for annotation by SnpEff<sup>99</sup> (version: 4.3). LoF variants here we used included *splice\_donor\_variant*, *splice\_acceptor\_variant* and *stop\_gained*. Numbers of LoF variants and ratios of LoF variants to total nsSNPs in Chinese pangolin populations were counted. Missense mutations were represented by *missense\_variant*. The ratio of homozygous (two per site) to (homozygous (two per site) plus heterozygous sites (one per site)) for all LoF, missense and deleterious variants were calculated for estimating the level of mutational load<sup>44</sup>.

GO and KEGG functional enrichment of genes affected by LoF mutations was performed using Metascape website (Last modified January 1, 2022)<sup>100</sup>. The GO terms and KEGG pathways with an enrichment factor > 2 and a multi-test adjusted *P*-value < 0.05 were considered to be significantly enriched. *P*-values and multi-test adjusted *P*-values were transformed with log base 10.

**Recently natural selection in populations.** We computed the iHS<sup>101</sup> (version: 1.3) to identify genomic signatures of positive selection in the CPA, CPB and CPC populations. The iHS calculations were performed independently in each population. As the genetic distance between adjacent SNPs was needed for the calculation, a chromosome segment of 1 Mb was straightly converted as 1 centiMorgan (cM).

For the identification of candidate genes, SNPs within the top 0.1% iHS scores were assigned as candidate sites. Based on candidate sites, we then used three methods to screen candidate regions described as Voight et al.<sup>101</sup>: a) regions of consecutive 50 SNPs; b) regions of 100 kb with 50 kb step size; c) 5 kb flanking regions away from candidate sites. We then calculated the sum of the iHS scores (siHS score) of all candidate sites in each genome region, and candidate regions were selected with the top 10% siHS score. The intersection of candidate genes obtained by three region-selected methods were used for the next analysis<sup>101</sup>. For the candidate genes, GO and KEGG functional enrichment was performed using Metascape website<sup>100</sup>. The GO terms and KEGG pathways with an enrichment factor > 2 and a multi-test adjusted *P*-value < 0.05 were regarded as significantly enriched.

**Statistics and reproducibility.** To test the significant difference of ROH and mutation load between different populations, two-sided Welch two-sample *t*-tests were performed in R<sup>102</sup> (version: 4.0.2). *P*-value less than 0.05 was considered to be significant. All statistics was done using available packages and reproducibility can be accomplished using parameters we mentioned in Methods.



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

## Data availability

All data generated or analyzed during this study are included in this published article and its supplementary files. Source data underlying Figures (1bcd, 2, 3ab, 4 and 5ab) in this article were provided in Supplementary Data 1. The data that support the findings of this study have been deposited into CNGB Sequence Archive (CNSA)<sup>103</sup> of China National GeneBank DataBase (CNGBdb)<sup>104</sup> with accession number CNP0001723 (<https://db.cngb.org/>).

Received: 15 November 2021; Accepted: 22 July 2022;

Published online: 25 August 2022

## References

- Purvis, A., Gittleman, J. L., Cowlshaw, G. & Mace, G. M. Predicting extinction risk in declining species. *Proc. R. Soc. Lond. Ser. B: Biol. Sci.* **267**, 1947–1952 (2000).
- Xue, Y. et al. Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science* **348**, 242–245 (2015).
- Frankham, R. et al. *A practical guide for genetic management of fragmented animal and plant populations.* (Oxford University Press, 2019).
- Weeks, A. R. et al. Genetic rescue increases fitness and aids rapid recovery of an endangered marsupial population. *Nat. Commun.* **8**, 1–6 (2017).
- Frankham, R. Genetic rescue of small inbred populations: Meta-analysis reveals large and consistent benefits of gene flow. *Mol. Ecol.* **24**, 2610–2618 (2015).
- Trinkel, M. et al. Translocating lions into an inbred lion population in the Hluhluwe-iMfolozi Park, South Africa. *Anim. Conserv.* **11**, 138–143 (2008).
- Pimm, S. L., Dollar, L. & Bass, O. L. Jr The genetic rescue of the Florida panther. *Anim. Conserv.* **9**, 115–122 (2006).
- Von Seth, J. et al. Genomic insights into the conservation status of the world's last remaining Sumatran rhinoceros populations. *Nat. Commun.* **12**, 1–11 (2021).
- Wang, P. et al. Genomic Consequences of Long-Term Population Decline in Brown Eared Pheasant. *Mol. Biol. Evol.* **38**, 263–273 (2021).
- Hinten, G., Harriss, F., Rossetto, M. & Braverstock, P. Genetic variation and island biogeography: microsatellite and mitochondrial DNA variation in island populations of the Australian bush rat, *Rattus fuscipes greyii*. *Conserv. Genet.* **4**, 759–778 (2003).
- Sharma, R. et al. Genetic diversity and relationship of Indian cattle inferred from microsatellite and mitochondrial DNA markers. *BMC Genet.* **16**, 1–12 (2015).
- Jensen-Seaman, M. & Kidd, K. Mitochondrial DNA variation and biogeography of eastern gorillas. *Mol. Ecol.* **10**, 2241–2247 (2001).
- Allendorf, F. W., Hohenlohe, P. A. & Luikart, G. Genomics and the future of conservation genetics. *Nat. Rev. Genet.* **11**, 697–709 (2010).
- Ryynänen, H. J., Tonteri, A., Vasemägi, A. & Primmer, C. R. A comparison of biallelic markers and microsatellites for the estimation of population and conservation genetic parameters in Atlantic salmon (*Salmo salar*). *J. Heredity* **98**, 692–704 (2007).
- Allendorf, F. & Seeb, L. Concordance of genetic divergence among sockeye salmon populations at allozyme, nuclear DNA, and mitochondrial DNA markers. *Evolution* **54**, 640–651 (2000).
- Guang, X. et al. Chromosome-scale genomes provide new insights into subspecies divergence and evolutionary characteristics of the giant panda. *Sci. Bull.* **66**, 2002–2013 (2021).
- Ouborg, N. J., Pertoldi, C., Loeschke, V., Bijlsma, R. K. & Hedrick, P. W. Conservation genetics in transition to conservation genomics. *Trends Genet.* **26**, 177–187 (2010).
- Khan, A. et al. Genomic evidence for inbreeding depression and purging of deleterious genetic variation in Indian tigers. *Proc. Natl Acad. Sci.* **118**, e2023018118 (2021).
- Zhao, S. et al. Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. *Nat. Genet.* **45**, 67–71 (2013).
- Dussex, N. et al. Population genomics of the critically endangered kakāpō. *Cell Genomics* **1**, 100002 (2021).
- Clark, L., Van Thai, N. & Phuong, T. Q. in *Workshop on trade and conservation of pangolins native to south and southeast Asia*. 111.
- Gaudin, T. J., Emry, R. J. & Wible, J. R. The phylogeny of living and extinct pangolins (Mammalia, Pholidota) and associated taxa: a morphology based analysis. *J. Mamm. evolution* **16**, 235–305 (2009).
- Dorji, D. Distribution, habitat use, threats and conservation of the critically endangered Chinese pangolin (*Manis pentadactyla*) in Samtse District, Bhutan. *Unpublished. Rufford Small Grants, UK* (2017).
- Del Toro, I., Ribbons, R. R. & Pelini, S. L. The little things that run the world revisited: a review of ant-mediated ecosystem services and disservices (Hymenoptera: Formicidae). *Myrmecological N.* **17**, 133–146 (2012).
- Li, H.-F., Lin, J.-S., Lan, Y.-C., Pei, K. J.-C. & Su, N.-Y. Survey of the termites (Isoptera: Kalotermitidae, Rhinotermitidae, Termitidae) in a Formosan pangolin habitat. *Fla. Entomologist* **94**, 534–538 (2011).
- Zhou, Z.-M., Zhou, Y., Newman, C. & Macdonald, D. W. Scaling up pangolin protection in China. *Front. Ecol. Environ.* **12**, 97–98 (2014).
- Zhang, H. et al. Molecular tracing of confiscated pangolin scales for conservation and illegal trade monitoring in Southeast Asia. *Glob. Ecol. Conserv.* **4**, 414–422 (2015).
- Luczon, A. U., Ong, P. S., Quilang, J. P. & Fontanilla, I. K. C. Determining species identity from confiscated pangolin remains using DNA barcoding. *Mitochondrial DNA Part B* **1**, 763–766 (2016).
- IUCN. *The IUCN Red List of Threatened Species.* <https://www.iucnredlist.org/search?query=pangolin&searchType=species> (2021).
- Wu, S., Liu, N., Zhang, Y. & Ma, G. Assessment of threatened status of Chinese Pangolin (*Manis pentadactyla*). *Chin. J. Appl. Environ. Biol.* **10**, 456–461 (2004).
- Yue, Z. in *Proceedings of the workshop on trade and conservation of pangolins native to South and Southeast Asia*.
- Heinrich, S. et al. Where did all the pangolins go? International CITES trade in pangolin species. *Glob. Ecol. Conserv.* **8**, 241–253 (2016).
- Dongliang, Z. Present Situation and Countermeasures of the Protection and Management of *Manis pentadactyla* in Fujian Province. *J. Fujian Forestry Sci. Technol.* **23**, 85–88 (1996).
- Jiang, Z. et al. Red list of China's vertebrates. *Biodivers. Sci.* **24**, 500 (2016).
- Wu, S. et al. The population and density of pangolin in dawuling natural reserve and the number of pangolin resource in Guangdong province. *Acta Theriol. Sin.* **22**, 270–276 (2002).
- Yang, L. et al. Historical data for conservation: reconstructing range changes of Chinese pangolin (*Manis pentadactyla*) in eastern China (1970–2016). *Proc. R. Soc. B* **285**, 20181084 (2018).
- Wu, S., Ma, G., Liao, Q. & Lu, K. (China Forestry Publishing House, Beijing, 2005).
- Zhang, F. et al. Observations of Chinese pangolins (*Manis pentadactyla*) in mainland China. *Glob. Ecol. Conserv.* **26**, e01460 (2021).
- Nash, H. C., Wong, M. H. & Turvey, S. T. Using local ecological knowledge to determine status and threats of the Critically Endangered Chinese pangolin (*Manis pentadactyla*) in Hainan, China. *Biol. Conserv.* **196**, 189–195 (2016).
- Hassanin, A., Hugot, J.-P. & van Vuuren, B. J. Comparison of mitochondrial genome sequences of pangolins (Mammalia, Pholidota). *Comptes rendus biologies* **338**, 260–265 (2015).
- Hu, J.-Y. et al. Genomic consequences of population decline in critically endangered pangolins and their demographic histories. *Natl Sci. Rev.* **7**, 798–814 (2020).
- Chinanews. *For the first time in Guangdong during the day, the Chinese pangolin came out of the cave*, <https://www.tellerreport.com/life/2020-07-10-for-the-first-time-in-guangdong-during-the-day-the-chinese-pangolin-came-out-of-the-cave.HkzCOdjByv.html> (2020).
- Choo, S. W. et al. Pangolin genomes and the evolution of mammalian scales and immunity. *Genome Res.* **26**, 1312–1322 (2016).
- Robinson, J. A. et al. Genomic flatlining in the endangered island fox. *Curr. Biol.* **26**, 1183–1189 (2016).
- Wan, Q.-H., Wu, H. & Fang, S.-G. A new subspecies of giant panda (*Ailuropoda melanoleuca*) from Shaanxi, China. *J. Mammal.* **86**, 397–402 (2005).
- Liu, Y.-C. et al. Genome-wide evolutionary analysis of natural history and adaptation in the world's tigers. *Curr. Biol.* **28**, 3840–3849. e3846 (2018).
- Armstrong, E. E. et al. Recent evolutionary history of tigers highlights contrasting roles of genetic drift and selection. *Mol. Biol. evolution* **38**, 2366–2379 (2021).
- Pečnerová, P. et al. High genetic diversity and low differentiation reflect the ecological versatility of the African leopard. *Curr. Biol.* **31**, 1862–1871. e1865 (2021).
- Altshuler, D., Donnelly, P. & Consortium, I. H. A haplotype map of the human genome. *Nature* **437**, nature04226 (2005).
- Nei, M. & Roychoudhury, A. K. Evolutionary relationships of human populations on a global scale. *Mol. Biol. evolution* **10**, 927–943 (1993).
- Hoffecker, J. F. *Desolate landscapes: Ice-age settlement in Eastern Europe.* (Rutgers University Press, 2002).
- Clark, P. U. et al. The last glacial maximum. *Science* **325**, 710–714 (2009).
- He, K. & Jiang, X. Sky islands of southwest China. I: an overview of phylogeographic patterns. *Chin. Sci. Bull.* **59**, 585–597 (2014).
- Zhou, X. et al. Population genomics reveals low genetic diversity and adaptation to hypoxia in snub-nosed monkeys. *Mol. Biol. Evolution* **33**, 2670–2681 (2016).

55. Kozma, R., Melsted, P., Magnússon, K. P. & Höglund, J. Looking into the past—the reaction of three grouse species to climate change over the last million years using whole genome sequences. *Mol. Ecol.* **25**, 570–580 (2016).
56. Dong, F. et al. Population genomic, climatic and anthropogenic evidence suggest the role of human forces in endangerment of green peafowl (*Pavo muticus*). *Proc. R. Soc. B* **288**, 20210073 (2021).
57. Wu, X. On the origin of modern humans in China. *Quat. Int.* **117**, 131–140 (2004).
58. Bergström, A. et al. Insights into human genetic variation and population history from 929 diverse genomes. *Science* **367**, eaay5012 (2020).
59. Schiffels, S. & Durbin, R. Inferring human population size and separation history from multiple genomic sequences. *Nat. Genet.* **46**, 919–925 (2014).
60. Schroeter, N. et al. Biomolecular evidence of early human occupation of a high-altitude site in Western Central Asia during the Holocene. *Front. Earth Sci.* **8**, 20 (2020).
61. Tunstall, T. et al. Evaluating recovery potential of the northern white rhinoceros from cryopreserved somatic cells. *Genome Res.* **28**, 780–788 (2018).
62. Spielman, D., Brook, B. W. & Frankham, R. Most species are not driven to extinction before genetic factors impact them. *Proc. Natl Acad. Sci.* **101**, 15261–15264 (2004).
63. Kardos, M. et al. Genetic consequences of intensive inbreeding in an isolated wolf population. *Nat. Ecol. evolution* **2**, 124–131 (2018).
64. van der Valk, T., Diez-del-Molino, D., Marques-Bonet, T., Guschanski, K. & Dalén, L. Historical genomes reveal the genomic consequences of recent population decline in eastern gorillas. *Curr. Biol.* **29**, 165–170. e166 (2019).
65. Meyermans, R., Gorssen, W., Buys, N. & Janssens, S. How to study runs of homozygosity using PLINK? A guide for analyzing medium density SNP data in livestock and pet species. *BMC genomics* **21**, 1–14 (2020).
66. DeRose, M. A. & Roff, D. A. A comparison of inbreeding depression in life-history and morphological traits in animals. *Evolution* **53**, 1288–1292 (1999).
67. Keller, L. F. & Waller, D. M. Inbreeding effects in wild populations. *Trends Ecol. evolution* **17**, 230–241 (2002).
68. Gallo, A., Boni, R. & Tosti, E. Gamete quality in a multistressor environment. *Environ. Int.* **138**, 105627 (2020).
69. Shibao, W., Qian, L., Ganxin, F. & Yayong, K. Preliminary study on food nutrient contents of Chinese pangolin (*Manis pentadactyla*). *J. Zhanjiang Norm. Coll.* **20**, 74–76 (1999).
70. Ma, J.-E. et al. Transcriptomic analysis identifies genes and pathways related to myrmecophagy in the Malayan pangolin (*Manis javanica*). *PeerJ* **5**, e4140 (2017).
71. Kawecki, T. J. & Ebert, D. Conceptual issues in local adaptation. *Ecol. Lett.* **7**, 1225–1241 (2004).
72. Edmands, S. Between a rock and a hard place: evaluating the relative risks of inbreeding and outbreeding for conservation and management. *Mol. Ecol.* **16**, 463–475 (2007).
73. WorldData.info. *The climate in China*, <https://www.worlddata.info/asia/china/climate.php> (2022).
74. Ralls, K., Sunnucks, P., Lacy, R. C. & Frankham, R. Genetic rescue: A critique of the evidence supports maximizing genetic diversity rather than minimizing the introduction of putatively harmful genetic variation. *Biol. Conserv.* **251**, 108784 (2020).
75. Barker, K. Phenol-Chloroform Isoamyl Alcohol (PCI) DNA extraction. *At the Bench* (1998).
76. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
77. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491 (2011).
78. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids Res.* **38**, e164–e164 (2010).
79. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
80. Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
81. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. evolution* **35**, 1547 (2018).
82. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
83. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
84. Weir, B. S. & Cockerham, C. C. Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358–1370 (1984).
85. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
86. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
87. Durand, E. Y., Patterson, N., Reich, D. & Slatkin, M. Testing for ancient admixture between closely related populations. *Mol. Biol. evolution* **28**, 2239–2252 (2011).
88. Skoglund, P. et al. Genetic evidence for two founding populations of the Americas. *Nature* **525**, 104–108 (2015).
89. Browning, B. L. & Browning, S. R. Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics* **194**, 459–471 (2013).
90. Pickrell, J. & Pritchard, J. Inference of population splits and mixtures from genome-wide allele frequency data. *Nature Precedings* **8**, 1–1 (2012).
91. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).
92. Terhorst, J., Kamm, J. A. & Song, Y. S. Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nat. Genet.* **49**, 303–309 (2017).
93. Zhang, F. et al. A note on captive breeding and reproductive parameters of the Chinese pangolin, *Manis pentadactyla Linnaeus*, 1758. *ZooKeys* **129**, 129–144 (2016).
94. Nei, M. & Li, W.-H. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl Acad. Sci.* **76**, 5269–5273 (1979).
95. Dobrynin, P. et al. Genomic legacy of the African cheetah, *Acinonyx jubatus*. *Genome Biol.* **16**, 1–20 (2015).
96. Kyriazis, C. C., Wayne, R. K. & Lohmueller, K. E. Strongly deleterious mutations are a primary determinant of extinction risk due to inbreeding depression. *Evolution Lett.* **5**, 33–47 (2021).
97. Li, W.-H., Wu, C.-I. & Luo, C.-C. Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. *J. Mol. Evolution* **21**, 58–71 (1984).
98. Feng, S. et al. The genomic footprints of the fall and recovery of the crested ibis. *Curr. Biol.* **29**, 340–349. e347 (2019).
99. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).
100. Zhou, Y. et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1–10 (2019).
101. Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A map of recent positive selection in the human genome. *PLoS Biol.* **4**, e72 (2006).
102. Team, R. D. C. *The R Reference manual: base package*. (Network Theory, 2004).
103. Guo, X. et al. CNSA: a data repository for archiving omics data. *Database* **2020**, baaa055 (2020).
104. Chen, F. Z. et al. CNGbDb: china national genebank database. *Yi Chuan= Hereditas* **42**, 799–809 (2020).

## Acknowledgements

This project was financially supported by funding from the Guangdong Provincial Key Laboratory of Genome Read and Write (grant No. 2017B030301011). This work was also supported by China National GeneBank, Guangdong Academy of Forestry and Technology Innovation Project of Guangdong (No. 2022KJCX008). Finally, we are thankful to the China National GeneBank for producing the sequencing data and Guangdong Provincial Academician Workstation of BGI Synthetic Genomics (No. 2017B090904014).

## Author contributions

T.L., H.L. and Y.H. conceived and designed the project. T.L. and Y.Z. were responsible for collecting the data. Y.H. and H.M.L. performed DNA extraction, library construction, and sequencing. Q.W., H.M.L., Q.L., L.H., S.Y., and L.Z. coordinated the data analysis. Q.W., M.S., and H.M.L. contributed to data analysis. Q.W. participated in data interpretation and visualization. T.L., S.K.S., D.Z., and Q.W. wrote the manuscript. T.L., H.L., and Y.H. provided supervision. All authors have read and approved the final manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s42003-022-03757-3>.

**Correspondence** and requests for materials should be addressed to Tianming Lan, Huan Liu or Yan Hua.

**Peer review information** *Communications Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Caitlin Karniski and Zhijuan Qiu.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022