

Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation

Shancen Zhao^{1,2,10}, Pingping Zheng^{1,3,10}, Shanshan Dong^{2,10}, Xiangjiang Zhan^{1,10}, Qi Wu^{1,10}, Xiaosen Guo², Yibo Hu¹, Weiming He², Shanning Zhang⁴, Wei Fan², Lifeng Zhu¹, Dong Li², Xuemei Zhang², Quan Chen², Hemin Zhang⁵, Zhihe Zhang⁶, Xuelin Jin⁷, Jinguo Zhang⁸, Huanming Yang², Jian Wang², Jun Wang^{2,9} & Fuwen Wei¹

The panda lineage dates back to the late Miocene¹ and ultimately leads to only one extant species, the giant panda (*Ailuropoda melanoleuca*). Although global climate change and anthropogenic disturbances are recognized to shape animal population demography^{2,3} their contribution to panda population dynamics remains largely unknown. We sequenced the whole genomes of 34 pandas at an average 4.7-fold coverage and used this data set together with the previously deep-sequenced panda genome⁴ to reconstruct a continuous demographic history of pandas from their origin to the present. We identify two population expansions, two bottlenecks and two divergences. Evidence indicated that, whereas global changes in climate were the primary drivers of population fluctuation for millions of years, human activities likely underlie recent population divergence and serious decline. We identified three distinct panda populations that show genetic adaptation to their environments. However, in all three populations, anthropogenic activities have negatively affected pandas for 3,000 years.

We carried out whole-genome resequencing of 34 wild giant pandas (Fig. 1a and Supplementary Table 1). This sample constitutes ~2% of the current estimates of the entire wild panda population⁵, the highest percentage of individuals assessed for existing animal population genomics studies. Genome alignment indicated an average of 91.5% sequencing coverage and 4.7-fold depth for each individual relative to the panda's 2.25-Gb genome⁴. To improve SNP inference quality, we estimated the probabilities of individual genotypes and population allele frequencies for each site⁶ and identified a total of 13,020,055 SNPs with ≥99% probability of being variable over the panda population.

We inferred three distinct genetic clusters—Qinling (QIN), Minshan (MIN) and Qionglai-Daxiangling-Xiaoxiangling-Liangshan

(QXL)—among the current panda population using *frappe*⁷, Admixture⁸ and an allele-shared matrix (Online Methods) (Fig. 1 and Supplementary Fig. 1). Previous studies only showed a distinct QIN cluster⁹; our larger study revealed that the MIN and QXL populations were also genetically distinct. We found no population substructure present in the QIN or MIN population but detected two subpopulations within the QXL population ($K = 4$; Fig. 1b and Supplementary Fig. 1): one comprising Xiaoxiangling and some Qionglai individuals and the other comprising Daxiangling, Liangshan and the remaining Qionglai individuals. The fixation index (F_{ST})¹⁰ strongly supported this three-population stratification (Supplementary Table 2). Principal-components analysis (PCA)¹¹ provided additional corroborative evidence. The first eigenvector separated these three genetic populations ($P < 0.05$) (Fig. 1c and Supplementary Fig. 2). The second eigenvector indicated that the Liangshan population was separate from the other populations, but this assignment was ambiguous because of limited sampling in the Liangshan population ($n = 2$ individuals). Overall, the three populations showed similar genetic diversity ($1.04\text{--}1.30 \times 10^{-3}$ for Watterson's estimator (θ_w) and $1.13\text{--}1.37 \times 10^{-3}$ for the average pairwise diversity within populations (θ_π); Supplementary Table 3) as humans¹², confirming the results from a study using ten microsatellite loci that indicated that the panda has substantial genetic variability⁹.

To reconstruct the demographic history of the giant panda, we used the pairwise sequentially Markovian coalescent (PSMC) model¹³ to examine changes in the local density of heterozygotes across the panda genome⁴. PSMC analysis showed a well-defined demographic history from 8 million to 20,000 years ago (Fig. 2a), a period covering the chronological distribution of three fossil panda species or subspecies (primal panda *Ailurarctos lufengensis*, pygmy panda *Ailuropoda microta* and baconi panda *Ailuropoda melanoleuca baconi*)^{1,14}. Considering the time since the origin of the panda, demography showed population peaks at ~1 million years ago and

¹Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing, China. ²Shenzhen Key Laboratory of Transomics Biotechnologies, BGI-Shenzhen, Shenzhen, China. ³College of Life Sciences, University of the Chinese Academy of Sciences, Beijing, China. ⁴China Wildlife Conservation Society, Beijing, China. ⁵China Conservation and Research Center for the Giant Panda, Wolong, China. ⁶Chengdu Research Base of Giant Panda Breeding, Chengdu, China. ⁷Shaanxi Wild Animal Rescue and Research Center, Louguantai, Xi'an, China. ⁸Beijing Zoo, Beijing, China. ⁹Department of Biology, University of Copenhagen, Copenhagen, Denmark. ¹⁰These authors contributed equally to this work. Correspondence should be addressed to F.W. (weifw@ioz.ac.cn) or Jun Wang (wangj@genomics.org.cn).

Received 13 March; accepted 14 November; published online 16 December 2012; doi:10.1038/ng.2494

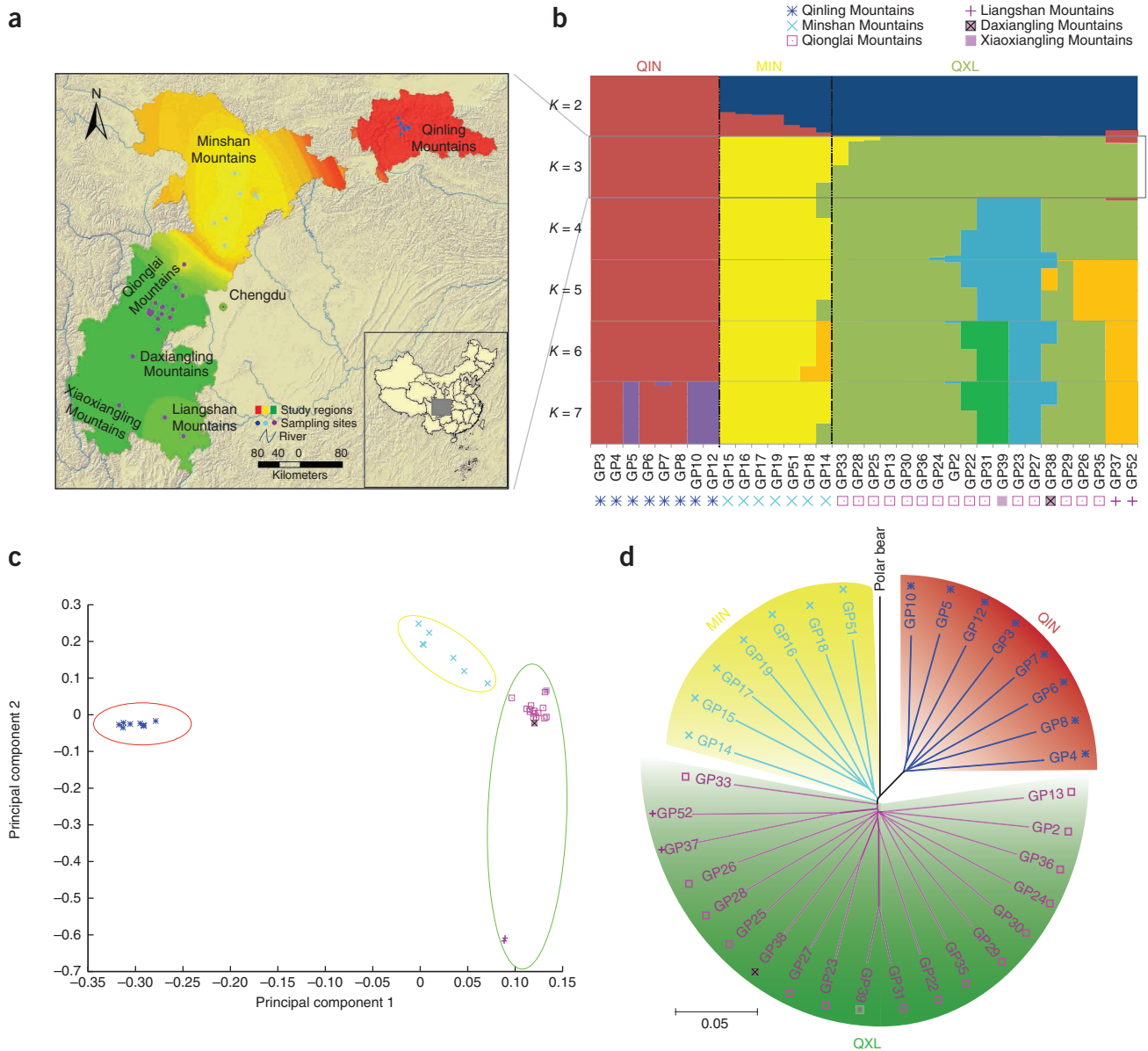


Figure 1 Current geographic populations of the giant panda and inferred genetic populations. **(a)** Sampling sites and genetic structure detected by *frappe* analysis ($K = 3$ populations) were mapped using ArcGIS v9.2 on the basis of the proportion of an individual's ancestry attributed to a given population. The genetic QIN population is shown in red, the MIN population is shown in yellow, and the QXL population is shown in green. Inset, the shaded area represents current panda habitats. **(b)** Genetic populations of the studied pandas inferred by *frappe* analysis. The number of populations (K) was predefined from 2 to 7. Symbols following each panda ID indicate where sampling occurred. **(c)** Results obtained from PCA using autosomal SNPs. Principal components 1 and 2 are shown. **(d)** A rooted neighbor-joining tree constructed from the allele-shared matrix of SNPs among the wild pandas, with the polar bear as an outgroup. The scale bar represents the p distance.

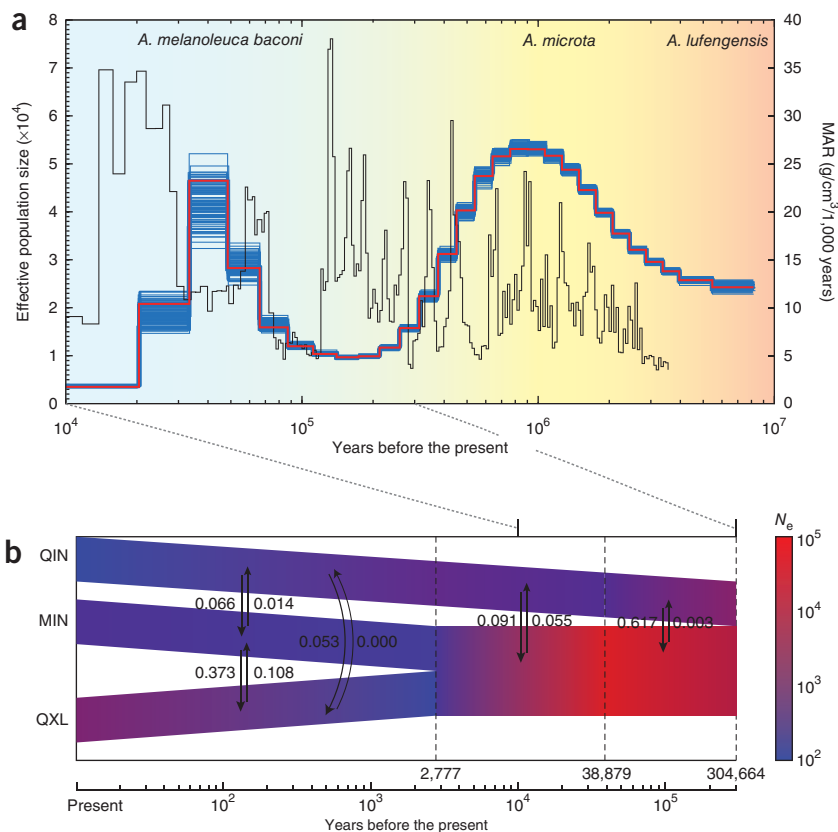
~40,000 years ago and population bottlenecks at ~0.2 million years ago and ~20,000 years ago (**Fig. 2a**). Notably, we found that these fluctuations in effective population size (N_e) were significantly negatively correlated with changes in the amount of atmospheric dust, as inferred by the mass accumulation rate (MAR) of Chinese loess¹⁵ (Pearson's correlation $R = -0.30$, $P < 0.05$), an index indicating cold and dry or warm and wet climatic periods in China.

The first population expansion coincided with a dietary switch to bamboo ~3 million years ago when pygmy pandas emerged¹⁶. Fossil evidence indicates that the earliest (primal) pandas were omnivores or carnivores, living in swamp habitats lacking bamboo¹, whereas pygmy pandas mainly ate bamboo, as indicated by their specialized cranial and dental adaptations^{16,17}. This hypothesis is supported at

the molecular level by the concurrent pseudogenization of the umami taste gene *Tas1r1* associated with the pandas' decreased reliance on meat¹⁸. The low levels of MAR during that time (**Fig. 2a**) indicate warm and wet weather conditions, which were ideal for the spread of bamboo forests.

The panda population declined around 0.7 million years ago, and the first bottleneck occurred about 0.2 million years ago (**Fig. 2a**), around the same time as the two largest Pleistocene glaciations in China, the Naynayxungla Glaciation (0.78–0.50 million years ago) and the Penultimate Glaciation (0.30–0.13 million years ago)¹⁹. Additionally, fossil evidence indicated that, from ~0.75 million years ago, the pygmy panda had been replaced by the subspecies *A. melanoleuca baconi*, which has the largest body size of all the panda species¹⁴.

Figure 2 Demographic history of the giant panda reconstructed from the reference and population resequencing genomes. **(a)** PSMC result showing demographic history from the panda's origin to 10,000 years ago. The red line represents the estimated effective population size (N_e), and the 100 thin blue curves represent the PSMC estimates for 100 sequences randomly resampled from the original sequence. The brown line shows the MAR of Chinese loess¹⁵. Generation time (g) = 12 years, and neutral mutation rate per generation (μ) = 1.29×10^{-8} . The approximate chronological ranges of three fossil panda species or subspecies (pygmy, baconi and giant panda) are shaded in pink, orange and blue, respectively. Note that PSMC simulation cannot detect population changes more recent than 20,000 years ago. **(b)** ∂adi result showing the demographic history of the panda from ~300,000 years ago to the present. The density of the heatmap shows fluctuations in effective population size. Two population divergences (304,664 and 2,777 years ago) and a population peak (38,879 years ago) are indicated by dashed lines. The average number of migrants per year between any two populations in each time interval is shown beside each arrow.



A cold climate, as evidenced by high MAR (Fig. 2a), might have contributed to the extinction of pygmy pandas and facilitated the origin of baconi pandas, or, possibly, the larger baconi pandas evolved from pygmy pandas as they adapted to the extreme weather.

The second population expansion occurred after the retreat of the Penultimate Glaciation¹⁹ (Fig. 2a, MAR decline), and the panda population reached its pinnacle between 30,000–50,000 years ago. The warm weather during the Greatest Lake Period (30,000–40,000 years ago) could have contributed to the population expansion, as would the alpine conifer forests, the primary habitat for pandas²⁰, having reached their greatest extent at this time (Supplementary Fig. 3)²¹. The second population bottleneck occurred during the last glacial maximum (~20,000 years ago), when substantial alpine glaciations (for example, Gongga glacial II; ref. 19) would likely have resulted in extensive loss of panda habitats.

Reconstruction of more recent panda demographic history could not be carried out using the PSMC approach because the power of this approach is greatly reduced for events occurring more recently than 20,000 years ago (Fig. 2a), owing to the limited number of recombination events in a single genome in this relatively short time interval¹³. We therefore used diffusion approximations for demographic inference (∂adi)²² to simulate recent demographic fluctuations on the basis of the SNPs we identified in our panda populations. The results of ∂adi analysis overlapped with and supported the PSMC findings of the second population expansion and its subsequent decline (Supplementary Fig. 4) and provided information on panda population history up to the present.

This simulation showed that the QIN and non-QIN populations diverged ~0.3 million years ago (95% confidence interval (CI) of 0.1–0.7 million years ago; Fig. 2b), corresponding with the onset of the Penultimate Glaciation¹⁹. About 40,000 years ago (CI = 4,900–58,900 years ago), the non-QIN population expanded by 300%, while the QIN population lost ~80% of its initial effective size; this occurred at a time when there was marked concurrent habitat expansion in the

regions inhabited by non-QIN pandas (Supplementary Fig. 3). After this event, the non-QIN population began to decline, while the QIN population remained stable. Our data showed that, about ~2,800 years ago (CI = 400–4,100 years ago), the non-QIN cluster diverged into the MIN and QXL populations, which gave rise to today's pattern of three genetically distinct panda populations. These three populations further fluctuated but in different ways: the QIN population decreased, the MIN population increased slightly, and the QXL population increased more substantially (Fig. 2b and Supplementary Table 4). The QIN population's decline correlated with the most extensive linkage disequilibrium (LD; Supplementary Fig. 5).

Probable causes of the QIN population decline include habitat loss and human activities. Arboreal pollen studies have indicated that there was a continuing and extensive decline in forest habitats in northern China, including the area populated by QIN pandas, around 4,000 years ago²³; however, paleobiological studies have indicated that this was unlikely to have been caused by concurrent changes in climate, as there was no differential impact on the populations of wet habitat-adapted species (for example, *Pinus* species) and dry habitat-adapted species (for example, *Quercus* species)²⁴. Instead, there is evidence indicating that deforestation in the QIN-populated area was associated with anthropogenic disturbances. At the beginning of the Spring and Autumn Period in China (770–486 BC), revolutionary improvements in farming technology greatly advanced local agricultural development and increased the capacity to reclaim woodland (Supplementary Note). For the next 2,500 years (~500 BC–present), the northern Qinling region remained one of the most prosperous areas in central China, giving rise to centralized geopolitical power and extensive human settlements²⁴, but resulting in depletion of the surrounding forest (Supplementary Note). Additionally, humans hunted and raised giant pandas for entertainment, sacrifice (during the Han Dynasty), and rewards (during the Tang Dynasty) (Supplementary Note).

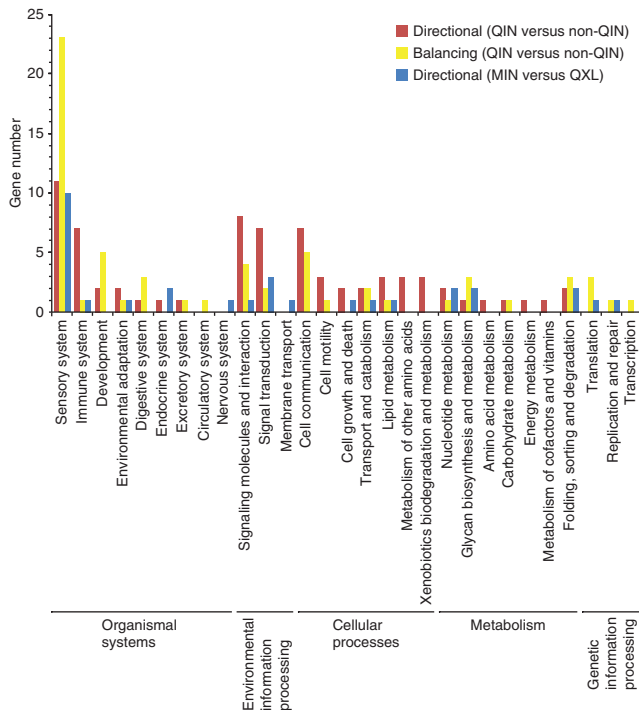


Figure 3 Annotation of genes containing selected SNPs on the basis of the KEGG database. Red bars, the number of genes under directional selection between the QIN and non-QIN panda populations; yellow bars, the number of genes under balancing selection between the QIN and non-QIN populations; blue bars, the number of genes under directional selection between the MIN and QXL populations.

Together, along with lack of evidence for a climatic effect, these human activities seem to have had an important role in the decline of the QIN population. It does remain possible that stochastic ecological events (for example, large-scale bamboo flowering and die-off or infectious disease) or panda-relevant physiological problems (for example, degenerative reproductive ability) contributed to this decline. However, there is no solid evidence to indicate that panda reproductive ability was compromised⁹ nor any support for stochastic events, leaving anthropogenic disturbance as the most plausible explanation for the decline.

The non-QIN (MIN and QXL) clusters were geographically divided by the Min River valley, along which the ancient Shu people established a kingdom (6,700–2,300 years ago) and built the most important road connecting their kingdom with the outside world (**Supplementary Fig. 6**)²⁵. Such a geographic barrier, accompanied by regional deforestation and human activity, might have established the initial separation of the two populations ~2,800 years ago (**Fig. 2b**). The increase in the MIN and QXL populations (**Fig. 2b**) coincided with the retreat of the Shu people from panda habitats to the lowlands and abandonment of the road²⁵. Regional reduction in human activities should have allowed habitat recovery in these regions. Alternatively, colonization of new habitats would have enabled population expansion. About 2,400 years ago, a new north-south road was established in the Qinling Mountains (**Supplementary Fig. 6**), which would have placed more anthropogenic pressure on the QIN pandas, resulting in the decline of this population.

To examine local population adaptation, we used coalescence-based simulation methods^{26,27}, a Bayesian test²⁸ and a ‘model-free’ global F_{ST} test²⁹ to detect selection signals in coding DNA sequences (CDS) across the whole genome (Online Methods). Between QIN and non-QIN populations, a total of 111 (134 SNPs) and 152 (212 SNPs) genes were

found to be under directional and balancing selection, respectively (**Supplementary Table 5**). KEGG annotation showed that the largest groups of selected genes were involved in the sensory system (**Fig. 3**). Of these, two genes, *Tas2r49* and *Tas2r3*, were directionally selected across the two panda populations (**Supplementary Table 6**). Studies in humans have indicated that these genes are functionally relevant to bitter taste³⁰, and they may have a similar role in pandas. A derived allele frequency (DAF) test³¹ indicated that *Tas2r49* was positively selected in the QIN population. Consistent with this finding, field observations showed that QIN pandas, compared to non-QIN (for example, QXL³²) pandas, consume more bamboo leaves³³, which are higher in alkaloids (a major bitter component) than other parts of the plant.

We also found 8 olfactory receptor genes under directional selection and 24 under balancing selection (**Supplementary Tables 6 and 7**). Odor perception as a form of olfactory communication is crucial for panda reproduction and survival, given their solitary existence in dense forest³⁴. However, only ligands of the gene *OR52R1* have been identified in giant panda scent marks³⁵ (**Supplementary Table 7**). More detailed analysis of the function of the other olfactory receptor genes might therefore be worthwhile.

The MIN and QXL populations, compared to the QIN and non-QIN populations, have fewer directionally selected genes ($n = 44$; **Supplementary Table 8**), indicating less variation in the selection processes between them, which is consistent with their lower interpopulation habitat heterogeneity⁵ and genetic differentiation (**Supplementary Table 2**). The largest group of selected genes in these populations is related to the sensory system (**Fig. 3**), but we saw no directional selection signals for the two receptor genes involved in bitter taste, *Tas2r3* and *Tas2r49*. We also identified eight olfactory receptor genes under directional selection in the MIN and QXL populations, but only *OR51L1* overlapped with those identified in the QIN and non-QIN populations (**Supplementary Tables 6 and 8**).

Giant pandas once inhabited most of China and neighboring countries in southeast Asia, but today they are confined to six relatively isolated mountain habitats in western China^{36,37}. In this study, integration of genomic and population genomics approaches provided a continuous outline of the history of the panda population and demonstrated that recent anthropogenic disturbances are likely a major reason for the panda's current endangered status. Although the presence of substantial genetic diversity in panda populations improves our chances of saving this iconic species, human activities have already fragmented some populations (for example, the QXL population) into small geographically isolated populations (for example, Xiaoxiangling), putting them at greater risk of extinction in the long term³⁷. For such small populations, translocation of wild-caught individuals or release of captive-bred individuals might be a useful means for genetic rescue by reestablishing gene flow. However, our data indicate that it will be important to monitor the evidence for selection and local adaptation in these fragmented panda populations, as reintroduction candidates ill-suited to a particular environment will be unlikely to promote the development of a robust population. This study may also serve as a model for other endangered species in assessing and establishing the most effective long-term conservation solutions.

URLs. Giant panda genome, <http://gigadb.org/giant-panda/>; bear *Ursus maritimus* genome, <http://gigadb.org/polar-bear/>; LASTZ (at the Miller Lab website), http://www.bx.psu.edu/miller_lab/; TreeBeST, <http://treesoft.sourceforge.net/treebest.shtml>; *frappe*, <http://med.stanford.edu/tanglab/software/frappe.html>; Ensembl, <http://asia.ensembl.org/index.html>; KEGG, <http://www.genome.jp/kegg/>; GenBank, <http://www.ncbi.nlm.nih.gov/genbank/>.

METHODS

Methods and any associated references are available in the [online version of the paper](#).

Accession codes. Panda resequencing reads have been deposited in the NCBI Sequence Read Archive (SRA) under accession [SRA053353](#).

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

This study was supported by grants from the National Natural Science Foundation of China (31230011), the Knowledge Innovation Program of the Chinese Academy of Sciences (KSCX2-EW-Z-4) and the State Forestry Administration of China. We thank the Chongqing Zoo, the Fuzhou Research Center of the Giant Panda, the Shanghai Zoo, the Shanghai Wildlife Park and the Zhengzhou Zoo for assistance during sample collection. We acknowledge T. Meng for generation of the panda distribution map, R.N. Gutenkunst for suggestions on analysis with *daði*, H. Li for suggestions on PSMC simulations and L. Goodman, J. Elser, M. Holyoak, S. Kumar and R.R. Swaisgood for comments and revisions of this manuscript. We also thank G. Tian, M. Jian, H. Jiang, M. Zhao, Q. Zhang, B. Wang, Y. Huang, G. Wang, C. Lin and F. Xi for laboratory assistance and B. Li for assistance on polar bear data analysis.

AUTHOR CONTRIBUTIONS

F.W. designed the research and interpreted data. Jun Wang led the genome sequencing and supervised the analysis. P.Z., S. Zhang, L.Z., H.Z., Z.Z., X.J. and J.Z. prepared the samples. S. Zhao, P.Z., X. Zhan, Y.H., Jian Wang and H.Y. performed research. S. Zhao, Q.W., S.D., X. Zhan, P.Z., X.G., W.H., W.F., D.L., X. Zhang and Q.C. analyzed the data. X. Zhan, F.W., P.Z., S. Zhao, Q.W. and S.D. wrote and revised the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at <http://www.nature.com/doi/10.1038/ng.2494>.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Qiu, Z. & Qi, G. *Ailuropoda* found from the late Miocene deposits in Lufeng, Yunnan. *Vertebrata Palasiatica* **27**, 153–169 (1989).
- Root, T.L. & Schneider, S.H. Ecology and climate: research strategies and implications. *Science* **269**, 334–341 (1995).
- Hewitt, G. The genetic legacy of the Quaternary ice ages. *Nature* **405**, 907–913 (2000).
- Li, R. *et al.* The complete genome sequence of the giant panda. *Nature* **463**, 311–317 (2010).
- State Forestry Administration. *The 3rd National Survey Report on Giant Panda in China* (Science Press, Beijing, 2006).
- Yi, X. *et al.* Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* **329**, 75–78 (2010).
- Tang, H., Peng, J., Wang, P. & Risch, N.J. Estimation of individual admixture: analytical and study design considerations. *Genet. Epidemiol.* **28**, 289–301 (2005).
- Alexander, D.H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
- Zhang, B. *et al.* Genetic viability and population history of the giant panda, putting an end to the “Evolutionary Dead End”? *Mol. Biol. Evol.* **24**, 1801–1810 (2007).
- Weir, B.S. & Cockerham, C. Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**, 1358 (1984).
- Patterson, N., Price, A.L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
- The Bovine HapMap Consortium. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* **324**, 528–532 (2009).
- Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).
- Wang, J. On the taxonomic status of species, geological distribution and evolutionary history of *Ailuropoda*. *Acta Zool. Sinica* **20**, 191–201 (1974).
- Sun, Y.B. & An, Z.S. Late Pliocene-Pleistocene changes in mass accumulation rates of eolian deposits on the central Chinese Loess Plateau. *J. Geophys. Res.* **110**, D23101 (2005).
- Pei, W. Evolutionary history of giant pandas. *Acta Zool. Sinica* **20**, 188–190 (1974).
- Jin, C. *et al.* The first skull of the earliest giant panda. *Proc. Natl. Acad. Sci. USA* **104**, 10932–10937 (2007).
- Zhao, H., Yang, J., Xu, H. & Zhang, J. Pseudogenization of the umami taste receptor gene *Tas1r1* in the giant panda coincided with its dietary switch to bamboo. *Mol. Biol. Evol.* **27**, 2669–2673 (2010).
- Zheng, B., Xu, Q. & Shen, Y. The relationship between climate change and Quaternary glacial cycles on the Qinghai-Tibetan Plateau: review and speculation. *Quat. Int.* **97–98**, 93–101 (2002).
- Hu, J. & Wei, F. Comparative ecology of giant pandas in the five mountain ranges of their distribution in China. in *Giant Pandas: Biology and Conservation* (eds. Lindburg, D. & Baragona, K.) 137–148 (University of California Press, London, 2004).
- Zhan, X., Zheng, Y., Wei, F., Bruford, M.W. & Jia, C. Molecular evidence for Pleistocene refugia at the eastern edge of the Tibetan Plateau. *Mol. Ecol.* **20**, 3014–3026 (2011).
- Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H. & Bustamante, C.D. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* **5**, e1000695 (2009).
- Ren, G. & Beug, H.J. Mapping Holocene pollen data and vegetation of China. *Quat. Sci. Rev.* **21**, 1395–1422 (2002).
- Ren, G. Decline of the mid-to-late Holocene forests in China: climatic change or human impact? *J. Quaternary Sci.* **15**, 273–281 (2000).
- Ren, N. *Illustrations and Annotations of Huayang Guo Zhi* (Shanghai Ancient Books Publishing House, Shanghai, 1987).
- Excoffier, L. & Lischer, H.E.L. Arlequin suite ver3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**, 564–567 (2010).
- Beaumont, M.A. & Nichols, R.A. Evaluating loci for use in the genetic analysis of population structure. *P. Roy. Soc. Lond. B. Bio.* **263**, 1619–1626 (1996).
- Foll, M. & Gaggiotti, O. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics* **180**, 977–993 (2008).
- Cockerham, C.C. & Weir, B. Estimation of gene flow from *F*-statistics. *Evolution* **47**, 855–863 (1993).
- Meyerhof, W. *et al.* The molecular receptive ranges of human TAS2R bitter taste receptors. *Chem. Senses* **35**, 157–170 (2010).
- Sabeti, P.C. *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918 (2007).
- Schaller, G.B., Hu, J., Pan, W. & Zhu, J. *The Giant Panda of Wolong* (University of Chicago Press, Chicago, 1985).
- Pan, W. *et al.* *A Chance for Lasting Survival* (Beijing University Press, Beijing, 2001).
- Swaisgood, R.R. *et al.* Chemical communication in giant pandas. in *Giant Pandas: Biology and Conservation* (eds. Lindburg, D.G. & Baragona, K.) 106–120 (University of California Press, Berkeley, CA, 2004).
- Hagey, L. & MacDonald, E. Chemical cues identify gender and individuality in giant pandas (*Ailuropoda melanoleuca*). *J. Chem. Ecol.* **29**, 1479–1488 (2003).
- Zhan, X. *et al.* Molecular censusing doubles giant panda population estimate in a key nature reserve. *Curr. Biol.* **16**, R451–R452 (2006).
- Zhu, L. *et al.* Drastic reduction of the smallest and most isolated giant panda population: implications for conservation. *Conserv. Biol.* **24**, 1299–1306 (2010).

ONLINE METHODS

Sampling information. Blood and tissue samples were obtained from 34 wild giant pandas. Sampling covered the 6 main geographic distributions, with 8 individuals from the Qinling Mountains, 7 from the Minshan Mountains, 15 from the Qionglai Mountains, 2 from the Liangshan Mountains, 1 from the Daxiangling Mountains and 1 from the Xiaoxiangling Mountains (Supplementary Table 1).

Library construction and sequencing. Genomic DNA was extracted from blood or muscle samples. For each individual, 1–3 µg of DNA was sheared into fragments of 200–800 bp with the Covaris system. DNA fragments were then treated according to the Illumina DNA sample preparation protocol: fragments were end repaired, A-tailed, ligated to paired-end adaptors and PCR amplified with 500-bp inserts for library construction. Sequencing was performed on the Illumina HiSeq 2000 platform, and 100-bp paired-end reads were generated.

Read alignment. We used the Burrows-Wheeler Aligner³⁸ to map paired-end reads to the reference genome⁴. First, the reference was indexed. Second, the command 'aln -t 3 -e 10' was used to find the suffix array coordinates of good matches for each read. Third, the command 'sampe -a 500 -o 1000' converted suffix array coordinates to pseudochromosomal coordinates and paired reads. Other parameters were set to the default.

Population SNP detection. We adopted an algorithm using a Bayesian approach to detect population SNPs⁶. Details of SNP calling are provided in the Supplementary Note.

SNP validation. A total of 366 SNPs from 4 pandas were validated by PCR and Sanger sequencing. In the test, 355 SNPs were confirmed to be polymorphic, 8 of which were homozygous; 3 SNPs were erroneously inferred. The error rate for population SNP calling was estimated to be 0.82–3.01%.

Principal-components analysis. We conducted PCA on autosomal biallelic SNPs using EIGENSOFT3.0 software¹¹. Eigenvectors from the covariance matrix were generated with the R function reigen, and significance levels were determined using the Tracey-Widom test (Supplementary Table 9).

Phylogenetic tree inference. We identified homologous regions between the panda and polar bear genomes using LASTZ (see URLs) and extracted SNPs within syntenic regions. Genotypes at the polar bear SNPs were considered to be the outgroup at corresponding positions. A neighbor-joining rooted tree was generated by TreeBeST (see URLs).

Population structure analyses. Genetic structure was inferred using the programs *frappe*⁷ and *Admixture*⁸, which implement an expectation-maximization algorithm and a block-relaxation algorithm, respectively. To explore the convergence of individuals, we predefined the number of genetic clusters *K* from 2–7 and ran both programs 5 times. The maximum iteration of the expectation-maximization algorithm was set to 10,000 in the *frappe* analysis. Default methods and settings were used in *Admixture* analysis.

θ_p , θ_w and F_{ST} calculations. The average pairwise diversity within a population (θ_p)³⁹ and Watterson's estimator (θ_w)⁴⁰ were calculated with sliding windows of different sizes (10, 100 and 500 kb) that had 90% overlap between adjacent windows. Population differentiation was measured by pairwise F_{ST} among three panda populations¹⁰.

Linkage disequilibrium. To evaluate LD decay, the correlation coefficient (r^2) between any two loci was calculated using Haploview⁴¹. Parameters were set as follows: -maxdistance 100, -dprime, -minMAF 0.01, -hwcutoff 0.0001 and -minGeno 0.6. Average r^2 was calculated for pairwise markers with the same distance, and LD decay was drawn using an R script.

Demographic history reconstruction using the PSMC approach. For the autosomal sequences⁴, scaffolds shorter than 50 kb (~2.6% of all scaffolds) were excluded to improve the accuracy of inferring historical recombination

events, and a total of 1,680,757 heterozygous loci were used to reconstruct demographic history with the PSMC model¹⁵. Parameters were set as follows: -N30, -t15, -r5 and -p '4+25*2+4+6'.

The estimated time to the most recent common ancestor (TMRCA) is given in units of $2N_0$ time, and the relative population size (N_e) at state t was scaled to N_0 (the present effective population size). The neutral mutation rate μ was used to infer N_0 and scale the TMRCA and N_e values into chronological time. The sequence divergence between the panda and polar bear was estimated to be 3.53%. Divergence between the two species was estimated to have occurred 16.4 million years ago (Supplementary Note) and a mean generation time (g) for pandas was set at 12 years⁴². Therefore, we calculated $\mu = (0.0353 \times 12) / (2 \times 16.4 \times 10^6) = 1.29 \times 10^{-8}$ mutations per generation for the giant panda.

Following Li's procedure¹³, we applied a bootstrapping approach, repeating sampling 100 times to estimate the variance of simulated results.

Correlation statistics for N_e inferred by PSMC results and MAR. We extracted N_e values for every time interval from the PSMC simulation results. We then averaged the corresponding MAR for the same intervals. We used Pearson's correlation in SPSS16.0 software to estimate the correlation of the two factors ($n = 46$).

Recent demographic history inference using $\partial a \partial i$. Of the SNPs identified in the 34 resequenced pandas, we only considered those from intergenic regions in autosomal sequences to ensure their neutrality. To minimize the effect of low-coverage sequencing, SNPs with more than 40-fold sequencing coverage at the population level were retained for the $\partial a \partial i$ ²² simulations. The polar bear genome sequence was used to infer ancestral alleles, and a statistical procedure was performed to correct ascertainment bias of the ancestral state, in which the trinucleotide substitution matrix specific for carnivores was kindly provided by the authors⁴³ (D.G. Hwang, University of Washington).

Four divergence models among three genetic populations of pandas were considered (Supplementary Note). The model with the maximum log-likelihood value was chosen as the optimal one (Supplementary Table 10). The ancestral population size (N_a) was estimated on the basis of the calculated θ value and the mutation rate. Population size and chronological split time were derived from parameters scaled by N_a . Nonparametric bootstrapping was performed 50 times to determine the variance of each parameter using the 50 new data sets with equal numbers of loci (111,161) sampled with replacement from the original data set.

Detecting SNPs under selection for pairwise populations. We performed F_{ST} -based approaches to investigate the selection signals across the whole genome. On the basis of the population structure detected, we defined two pairs of populations to detect the selection signals: (i) the QIN and non-QIN (MIN and QXL) pair and (ii) the MIN and QXL pair. We chose SNPs from CDS regions and excluded those with minor allele frequency of <0.05. A total of 37,999 and 37,405 loci were used for the analyses of the two comparisons, respectively.

First, the two pairs of populations were tested using the finite island model (with the FDIST approach²⁷) in Arlequin²⁶ to detect outliers. Considering the hierarchical genetic structure within the QIN and non-QIN populations (Fig. 2b), this pair was also analyzed using the hierarchical island model implemented in the same software²⁶ to estimate loci that were outliers with respect to F_{ST} between the two populations as well as outliers with respect to pairwise F_{CT} values (the proportion of total genetic variance due to differences among groups of populations). Parameters were set as default values, with the exception of our setting 200,000 simulations and allowing 10% missing data in each test. The *P* value for each locus was estimated using a kernel density approach. After completing the analysis, we performed false discovery rate (FDR) correction of *P* values, and each locus received a *q* value⁴⁴. We considered the loci with *q* value < 0.05 as possible outliers (Supplementary Figs. 7a–c and 8a). Then, a Bayesian test was performed using the program BayeScan²⁸. We ran 20 pilot runs of 50,000 iterations with an additional burn-in of 500,000 iterations and a thinning interval of 20. Other parameters were set at the default values. Because recent studies suggested that BayeScan is a conservative estimate^{45,46}, loci with FDR *q* < 0.1 were considered to be outliers in this analysis (Supplementary Figs. 7d and 8b). For each pair of populations, we also measured the pairwise global F_{ST} values²⁹ for every locus across the whole genome to detect the highly differential

SNPs between populations. The top 1% of SNPs ranked with raw F_{ST} values were considered to be potential outliers (**Supplementary Figs. 7e and 8c**).

To minimize the detection of false positives, we considered those loci identified by two or more methods to be outliers (**Supplementary Tables 11 and 12**) as true selected loci. In addition, for the QIN and non-QIN populations, we generated a derived allele frequency distribution for (i) the alleles under balancing selection, (ii) the alleles under directional selection and (iii) the unselected alleles. As expected, compared with directionally or unselected alleles, there was an enrichment in intermediate frequency for the derived allele frequency of alleles subjected to balancing selection (**Supplementary Fig. 9**), indicating that the false positives in the SNP set of balancing selection should be limited.

Derived allele frequency test. We used the DAF test³¹ to localize the signal of selection to populations. First, we inferred ancestral alleles using the polar bear genome. We then calculated and compared the derived allele frequency of each locus under directional selection between two populations to detect which population harbored a higher frequency of derived alleles. Populations with higher frequencies of derived alleles were assumed to be under positive selection.

Annotation of loci under selection. We annotated genes with selected SNPs using Ensembl (see URLs) and KEGG (see URLs) and then classified each

gene according to the KEGG pathways and KEGG Brite function databases (**Supplementary Tables 13 and 14**). Olfactory and taste receptor genes were examined using Ensembl, KEGG and GenBank annotations.

38. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
39. Tajima, F. Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**, 437–460 (1983).
40. Watterson, G.A. On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**, 256–276 (1975).
41. Barrett, J.C. *et al.* Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
42. Wei, F. *et al.* A study on the life table of wild giant pandas. *Acta Theriol. Sinica* **9**, 81–86 (1989).
43. Hwang, D.G. & Green, P. Bayesian Markov chain Monte Carlo sequence analysis reveals varying neutral substitution patterns in mammalian evolution. *Proc. Natl. Acad. Sci. USA* **101**, 13994–14001 (2004).
44. Storey, J.D. A direct approach to false discovery rates. *J. R. Stat. Soc., B* **64**, 479–498 (2002).
45. Huang, K., Whitlock, R., Press, M.C. & Scholes, J.D. Variation for host range within and among populations of the parasitic plant *Striga hermonthica*. *Heredity* **108**, 96–104 (2012).
46. Buckley, J., Butlin, R.K. & Bridle, J.R. Evidence for evolutionary change associated with the recent range expansion of the British butterfly, *Aricia agestis*, in response to climate change. *Mol. Ecol.* **21**, 267–280 (2012).