Palaeogenome reveals genetic contribution of extinct giant panda to extant populations

Gui-Lian Sheng[1,*], Nikolas Basler[2], Xue-Ping Ji[3], Johanna L. A. Paijmans[2,4], Michaela Preick[2], Stefanie Hartmann[2], Michael V. Westbury[2,5], Jun-Xia Yuan[1], Nina G. Jablonski[6], Federica Alberti[2], Georgios Xenikoudakis[2], Xin-Dong Hou[1], Bo Xiao[1], Jian-Hui Liu[3], Michael Hofreiter[2], Xu-Long Lai[1], Axel Barlow[2,*]

[1] State Key Laboratory of Biogeology and Environmental Geology, China University of Geosciences, Wuhan, Hubei 430074 China
[2] Institute for Biochemistry and Biology, University of Potsdam, Karl-Liebknecht-Strasse 24–25, 14476 Potsdam, Germany
[3] Yunnan Cultural Relics and Archaeology Institute, 15-1, Chunmingli, Chunyuanxiaoqu, Kunming, Yunnan 650118, China
[4] Present address: School of Archaeology and Ancient History, University of Leicester, Leicester, LE1 7RH, UK
[5] Present address: Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, DK-1350 Copenhagen K, Denmark
[6] Department of Anthropology, 409 Carpenter Building, The Pennsylvania State University, University Park, PA 16802, USA

*Corresponding authors: G.-L.S (email: glsheng@cug.edu.cn) and A.B. (email: axel.barlow.ab@gmail.com)

Lead contact: G.-L.S (email: glsheng@cug.edu.cn)

Summary: Historically, the giant panda was widely distributed from northern China to southwestern Asia [1]. As a result of range contraction and fragmentation, extant individuals are currently restricted to fragmented mountain ranges on the eastern margin of the Qinghai-Tibet plateau, where they are distributed among three major population clusters [2]. However, little is known about the genetic consequences of this dramatic range contraction. For example, were regions where giant pandas previously existed occupied by ancestors of present-day populations, or were these regions occupied by genetically distinct populations which are now extinct? If so, is there any contribution of these extinct populations to the genomes of giant pandas living today? To investigate these questions, we sequenced the nuclear genome of a ~5,000 year old giant panda from Jiangdongshan, Tengchong County in Yunnan Province, China. We find that this individual represents a genetically distinct population that diverged prior to the diversification of modern giant panda populations. We find evidence of differential admixture with this ancient population among modern individuals originating from different populations as well as within the same population. We also find evidence for directional gene flow, which transferred alleles from the ancient population into the modern giant panda lineages. A variable proportion of the genomes of extant individuals is therefore likely derived from the ancient population represented by our sequenced individual. Although extant giant panda populations retain reasonable genetic diversity, our results suggest that this represents only part of the genetic diversity this species harbored prior to its recent range contractions.

Results and Discussion

Ancient giant panda genome. We extracted DNA from 300 mg of bone powder sampled from a subfossil femur bone of a giant panda from Jiangdongshan, Tengchong County in Yunnan Province, south-western China, far south of the current distribution of giant pandas (Fig. 1). Remains of this individual have previously been radio-carbon dated at 5,025 ± 35 years before present [1], representing the last known record of the species from this region. Parts of the mitochondrial DNA sequence of this specimen have been obtained previously and its haplotype shown to nest within the phylogenetic diversity of modern giant panda mitochondrial DNA, as sister to a clade comprised of three haplotypes sampled from Mountains in Shaanxi Province [3, 4]. This may reflect either the ancient giant panda as a direct ancestor of this modern population, incomplete lineage sorting, or maternal gene flow among more diverged populations, none of which can be excluded based on mitochondrial evidence alone.

We converted twelve DNA extracts to Illumina sequencing libraries. A total of 1.75 billion reads were generated from these, of which 55 million could be mapped with high confidence to the giant panda nuclear reference genome assembly [5, 6], providing approximately 1.2x coverage of the genome of the ancient giant panda (Table S1). Analysis of this data indicated advanced DNA fragmentation and high levels of cytosine deamination at the terminal DNA fragment ends (Figure S1), consistent with the age of the sample. Contamination analysis provided no evidence of substantial contamination of the ancient sample by potentially contaminating mammalian DNA (Table S2).

Relationship to modern giant panda genomes. Previous studies have shown that modern giant pandas are distributed among three major geographic population clusters (Fig. 1; [2]). These comprise: a northeastern Qinling Mountains population (QIN); a western Minshan Mountains population (MIN); and a third southwestern population encompassing the Qionglai, Daxiangling, Xiaoxiangling, and Liangshan Mountains (QXL). Previous studies suggest that the modern MIN and QXL populations diverged around 2,800 years ago from a common ancestral population, which diverged from the population ancestral to QIN around 0.3 million years ago [2]. To gain insight into the relationship of the ancient giant panda to these populations, we carried out a principal components analysis (PCA). This analysis recovered the expected three modern population clusters. Within the QXL population, two individuals from Liangshan Mountains appear to be diverged considerably from other QXL individuals, which has been found by previous studies [3], and may reflect substructure within the QXL population. Regarding the ancient individual, the PCA showed that it does not cluster together with any specific extant population (Fig. 2a), in contrast to the relationships suggested by mitochondrial DNA (Figure S2). Phylogenetic analysis further suggests that the ancient individual represents a population which diverged prior to the diversification of modern populations (Fig. 2b), which is also supported by tree topology tests, which show that any pair of modern giant pandas always share a greater excess of derived alleles with each other than either of the two does

103 with the ancient giant panda (Fig. 2c). Overall, these results support the
104 ancient giant panda as representing a distinct and divergent extinct population
105 that diverged from all modern populations more than 0.3 million years ago. In
106 contrast, the estimated coalescence time of the ancient giant panda's
107 mitochondrial haplotype and its modern sister clade is ~14,150 years (95%
108 credibility interval 6,805–23,301 years, Figure S2). The close relationship of
109 the ancient giant panda's mitochondrial DNA with modern giant pandas
110 therefore most likely reflects maternal gene flow from an ancestor of the
111 modern populations into the extinct ancient population, since incomplete
112 lineage sorting would require a coalescence time older than the initial
113 divergence of all populations.
114
115 **Gene flow among giant panda populations.** We further investigated the
116 possibility of past gene flow among ancestors of the modern populations and
117 the extinct, ancient population using the D statistic [7, 8]. The D statistic is a
118 four taxon test of differential admixture among two closely related individuals
119 (P1 and P2) with a more distantly related candidate admixing lineage (P3),
120 which makes use of an outgroup (P4) for allele polarisation. Significant non-
121 zero D values suggest admixture with P3 subsequent to the divergence of P1
122 and P2, with negative and positive values indicating, respectively, P1 or P2 as
123 the lineage that is more admixed with P3. We tested combinations of
124 individuals consistent with their phylogeny (Fig. 2b) using the polar bear as
125 outgroup (P4). Specifically, we tested all combinations of: (((modern panda,
126 modern panda), ancient panda), polar bear). Since the ancient individual is in
127 P3, these tests will not be biased by increased rates of sequencing error in
128 the ancient dataset [9].
129
130 An additional source of bias for D statistics is the reference genome sequence
131 used for mapping [10]. Since the giant panda reference genome [6,7] likely
132 represents an ingroup to the investigated clade, mapping to this sequence will
133 be biased toward alleles found in the population from which the reference
134 genome descends, potentially leading to inflated estimates of admixture with
135 that population [11, 12]. Bias towards the reference allele is exacerbated for
136 ancient datasets since the expected baseline number of mismatches between
137 read and reference is higher due to increased rates of sequence errors in
138 paleogenomes compared to modern datasets. The giant panda reference
139 genome assembly is from a captive individual, "Jingjing", whose father
140 descends from the QXL population and whose mother descends from both
141 QXL and MIN populations. Correspondingly, D statistics calculated after
142 mapping reads to this reference suggest a general pattern of admixture
143 between the ancient population and the modern QXL and MIN populations,
144 relative to the modern QIN population (Figure S3). To further investigate
145 whether this result is driven by mapping reference bias, we remapped the
146 data to the reference genome assembly of the polar bear (Table S1), which
147 represents an outgroup to the investigated clade. These D statistics supported
148 very different patterns of admixture. To visualise these patterns we identified
149 one modern QXL individual (Qionglai Mountains, SRR504883) as showing the
150 least admixture with the ancient population, against which all other modern
151 individuals were compared (Fig. 3). Of these, all but one individual showed a
152 significant signal of admixture with the ancient population, relative to the least

admixed individual. Admixture levels appear highly variable within modern populations, with no obvious pattern of any single population being more or less admixed with the ancient population. Additional D statistic comparisons (Figure S3) further revealed differential levels of admixture with the ancient population within each modern population as well as within individual mountain ranges.

These complex patterns of admixture are unlikely to result solely from direct gene flow with the ancient population. Although the extinction date of the ancient population is unknown, the fact that the ancient individual sequenced in this study represents the last known occurrence of giant pandas in Yunnan Province suggests that extinction occurred shortly after ~5,000 years bp. By this reasoning, the ancient population had already gone extinct prior to the divergence of the MIN and QXL populations [2] by several thousand years. Variable diffusion of admixed alleles via gene flow among modern populations (i.e. including QIN), or with additional, so far unsampled extinct populations, is therefore required to explain the observed patterns of differential admixture both between and within the modern MIN and QXL populations.

The D statistic results suggest the possibility of survival of alleles from the ancient population in modern populations as a result of past admixture. However, D statistics cannot provide conclusive support for this hypothesis because the direction of gene flow is not explicitly tested. We therefore tested for directional gene flow from the ancient population into the modern populations using a previously described approach based on the distribution of phylogenetic tree topologies along non-overlapping 100 kb sliding genomic blocks [11]. For this test, we selected two MIN individuals, MIN+ and MIN-, which were found to be highly admixed and less admixed with the ancient population, respectively, relative to a QIN individual. The most commonly observed topology was the species tree: (((MIN+,MIN-),QIN),ancient)). We also observed a greater number of blocks where the MIN+ individual clustered with the ancient giant panda (((MIN-,QIN),(MIN+,ancient)), 1,190 blocks) than where the MIN- individual clustered with the ancient giant panda (((MIN+,QIN),(MIN-,ancient)), 1,126 blocks). This pattern indicates the transfer of alleles from the ancient population into the ancestors of the MIN+ individual, since unidirectional geneflow in the opposite direction would not be associated with such an imbalance in the observed frequency of these topologies. The observed imbalance equates to 6.4 Mb of the genome of MIN+ individual being derived from directional geneflow from the ancient population above that occurring in the MIN- individual. Since the test is relative, this estimate represents a conservative minimum since the complex overall patterns of admixture suggest a high likelihood that MIN- is itself admixed with the ancient population. Most importantly, the observed signal of admixture between the ancient giant panda and MIN+ at least in part reflects the presence of alleles derived from the ancient population in this modern giant panda. Ancestors of modern populations were therefore recipients of alleles from the ancient population which still persist in the genomes of living individuals.

Implications for our understanding of giant panda evolution. With an estimated census population size around 2,500 individuals [13, 14], the conservation status of the giant panda has recently been changed from "endangered" to "vulnerable" [15]. Giant pandas have also been shown to display moderate-to-high levels of genetic variation compared to other endangered carnivores, and even compared to humans [2, 6, 16-19]. Nonetheless, the current habitat range of giant panda is far more restricted than it was in the past [16, 20, 21], with genome data suggesting two major population bottlenecks at ~0.2 million years ago and ~20,000 years ago, respectively [2]. Elucidating the full impact of this range contraction in terms of both the loss of genetic diversity within surviving populations and the extinction of distinct and divergent populations is likely impossible based only on data from modern populations [22].

The palaeogenome of the ~5,000 year old giant panda from Yunnan Province presented here reveals an extinct, divergent population that is an outgroup to all extant giant panda populations, which must have diverged from them prior to 0.3 million years ago. This lost lineage survived through the Last Glacial Maximum and went extinct around the Middle Holocene. The sample used in this study represents the last known record of the giant panda in Yunnan Province before it disappeared from this region, and therefore probably approximates the extinction time of this lineage. However, genetically, the ancient population may not have gone fully extinct, since we found extensive evidence of differential admixture with it among all extant populations as well as evidence for the persistence of alleles from the ancient population in modern individuals.

Recently, independent studies have recovered mitochondrial haplotypes from Middle Holocene [3] and Late Pleistocene [23] giant pandas, which are sister to all modern giant panda haplotypes sampled thus far. Combined analysis of these ancient mitochondrial sequences suggests their monophyly indicating a divergent mitochondrial clade lost during the recent evolutionary history of giant pandas [4], mirroring the results of this study based on complete nuclear genomes. Given that the mitochondrial haplotype of the ancient giant panda investigated here almost certainly reflects a recent transfer of mitochondrial DNA from the ancestors of extant populations, and that this ancient individual was recovered from the same locality as the Middle Holocene lost clade individual, it could be tentatively suggested that the lost mitochondrial clade and the divergent ancient nuclear genome revealed by this study represent one and the same population. Genome sequencing of ancient giant pandas representing the lost mitochondrial clade therefore represents the next logical step in the study of the evolutionary history of the giant panda.

Acknowledgements

Author contributions

261 Conceptualization, G.-L.S., A.B., M.H., and X.-L.L.; Investigation, N.B., G.-
262 L.S., F.A., and M.P.; Resources, X.-P.J., N.G.J., J.-H.L., and M.H.;
263 Methodology, A.B., N.B., F.A., J.L.A.P., and G.X.; Formal Analysis, A.B., G.-
264 L.S., N.B., J.L.A.P., and S.H.; Software, A.B., N.B., and S.H.; Visualization,
265 G.-L.S., J.L.A.P., and A.B.; Writing – Original Draft, G.-L.S., A.B., M.H., and
266 M.V.W.; Writing – Review & Editing, G.-L.S., A.B., M.H., J.L.A.P., F.A., and
267 N.B.; Data Curation, N.B., X.-D.H., and B.X.; Funding Acquisition, G.-L.S., X.-
268 L.L., J.-X.Y., and M.H.; Project Administration, G.-L.S., M.H., and A.B.;
269 Supervision, G.-L.S., M.H., and A.B.
270
271

Declaration of interests

273 The authors declare no competing interests.
274

Main text figure legends

276

277 **Figure 1. Sampling locations of giant pandas investigated in this study**.
278 Photograph on the bottom left shows sampled the ancient giant panda femur
279 bone. The three modern giant panda populations (coloured areas) comprise:
280 Qinling population in Shaanxi Province (QIN); Minshan (MIN) population in
281 both Gansu and Sichuan Provinces; and the QXL population comprised of
282 Qionglai (QIO), Daxiangling (DXL), Xiaoxiangling (XXL), and Liangshan (LS)
283 in Sichuan Province. The red point indicates the approximate locality of the
284 ancient sample
285
286 **Figure 2. Relationship of the ancient giant panda to modern giant panda**
287 **genomes.** a. Ordination of individuals along the first and second components
288 of a PCA based on 409,165 variable transversion sites. Axis labels indicate
289 the percentage of variance explained by each component. Symbols for each
290 population are indicated in the key at the bottom left. Singleton positions were
291 excluded from this analysis, which conservatively reduces the overall
292 separation of the ancient from the modern individuals, but the ancient
293 population is still clearly distinct. b. Neighbour-joining phylogeny based on
294 403,235 transversion sites, rooted using the polar bear as outgroup (not
295 shown). Note that the QXL population is not recovered as monophyletic,
296 which likely reflects the substantial divergence of LS within this population
297 cluster (part a. and [2]). c. Topology tests for the position of the ancient panda
298 as basal to all modern pandas based on the excess of derived alleles that a
299 modern giant panda shares with the ancient giant panda and not with another
300 modern giant panda, compared to the excess of derived alleles that a modern

301 giant panda shares with another modern giant panda and not with the ancient
302 giant panda. These derived allele proportions (x axis) are expressed as D
303 statistics (black points) calculated for all combinations of individuals for the
304 topologies indicated on the y-axis. All Z-scores for ((ancient,modern),modern)
305 were > 3, whereas many Z-scores for ((modern,modern),ancient) were < 3.
306 Consistently lower D values for the topology ((modern,modern),ancient)
307 supports the ancient panda as basal to all modern pandas. The giant panda
308 was used as mapping reference for all these analyses. For mitochondrial
309 relationships see Figure S2.
310
311 **Figure 3. D statistic tests of differential admixture with the ancient giant**
312 **panda, relative to the least admixed modern giant panda**. These
313 correspond to all comparisons of the topology (((P1,P2),P3),P4), where P1 is
314 the least admixed modern panda, P2 is another modern panda, P3 is the
315 ancient panda, and P4 is the polar bear outgroup. Positive D values indicate
316 that the P2 individual is more admixed with P3 than P1 is with P3. D values
317 are indicated as points coloured according to the population origin of the P2
318 individual (consistent with Figures 1 and 2). The least admixed individual
319 originates from the QXL population. All comparisons are significant (Z > 3)
320 except the single smallest (leftmost) D value. The polar bear was used as
321 mapping reference for these analyses. D statistics tests of admixture with the
322 ancient giant panda for all pairs of modern giant pandas are shown in Figure
323 S3.
324
325
326 # STAR METHODS
327
328 ## CONTACT FOR REAGENT AND RESOURCE SHARING
329
330 Further information and requests for resources and reagents should be
331 directed to and will be fulfilled by the Lead Contact, Gui-Lian Sheng
332 (glsheng@cug.edu.cn)
333
334
335 ## METHOD DETAILS
336
337 ### *Laboratory procedures*
338 All DNA extraction and library preparation procedures were performed in
339 ancient-DNA-dedicated clean rooms following standard procedures to avoid
340 contamination. Negative (nuclease free water) controls were included in all
341 DNA extraction and library preparation steps.
342
343 We ground the ancient giant panda bone specimen to powder with a mortar
344 and pestle and separated the powder into ~25 mg aliquots. Initially, we
345 extracted DNA and prepared Illumina sequencing libraries from two of these
346 aliquots to assess the endogenous content of the ancient giant panda bone.
347 DNA extraction was carried out using a protocol optimised for the recovery of
348 short ancient DNA fragments [26], with slight modifications [24]. Bone powder
349 aliquots were each digested in 1 mL of extraction buffer (0.45 M EDTA, 0.25

350   mg/mL Proteinase K) overnight at 37 °C with rotation. Following
351   centrifugation, the supernatant was removed and combined with 13 mL of
352   binding buffer (5 M guanidine hydrochloride, 40% (vol/vol) isopropanol, 0.05%
353   Tween-20, and 90 mM sodium acetate) and passed through a commercial
354   silica spin column (Qiagen MinElute) with an extension reservoir (Zymo-spin
355   V) fitted. Two wash steps were then carried out (PE buffer, Qiagen) and the
356   DNA eluted in two steps each using 12.5 µL TET buffer (10mM Tris-HCl, 1
357   mM EDTA, 0.05% Tween-20).
358
359   We prepared Illumina sequencing libraries using a protocol based on single-
360   stranded DNA [27], with modifications [24], designed to efficiently recover
361   short ancient DNA fragments. The DNA extractions were quantified using a
362   Qubit 2.0 instrument (Fisher) with dsDNA HS Assay kit and the input volume
363   for library preparation was adjusted to 13 ng total input DNA to maintain the
364   efficiency of the single-stranded ligation reaction [27]. Input DNA was treated
365   with the enzymes uracil-DNA glycosylase and endonuclease VIII to excise
366   uracil residues resulting from cytosine deamination and to cleave DNA at
367   abasic sites, respectively. This involved 44 µL reactions with the following
368   reagent concentrations: 1.8x CircLigase buffer II, 4.5 mM $MnCl_2$, 0.11 U/µL of
369   uracil-DNA glycosylase, and 0.02 U/µL of endonuclease VIII. Residual
370   phosphate groups were then removed from the 5' and 3' DNA fragment ends
371   using 1 unit of FastAP. The double-stranded DNA was then heat denatured
372   and oligo CL78 ligated to the 3' end of the single strands by adding the
373   following reagents to a final volume of 80 µL and incubating overnight: 20%
374   (vol/vol) PEG-4000, 0.125 µM CL78, and 2.5 units/µL Circligase II. Ligation
375   products were then immobilised on streptavidin beads (MyOne C1) allowing
376   the removal of reagent mixtures for successive steps of the library
377   preparation. The CL9 extension primer was annealed to the complementary
378   CL78 oligo sequence and the strand complementary to the template single-
379   stranded molecules filled-in using Bst 2.0 polymerase in 50 µL reactions with
380   the following reagent concentrations: 1x isothermal amplification buffer, 250
381   µM of each dNTP, 2 µM CL9 extension primer, and 0.48 U/µL Bst 2.0
382   polymerase. 3' overhangs were then removed using T4 DNA polymerase in
383   100 µL reactions with the following reagent concentrations: 1x Buffer Tango,
384   0.025% (vol/vol) Tween 20, 100 µM of each dNTP, and 0.05 U/µL T4 DNA
385   polymerase. The double-stranded adapter (CL53/CL73) was then ligated to
386   the blunt-ended molecules using T4 DNA ligase in 100 µL reactions with the
387   following reagent concentrations: 1x T4 DNA ligase buffer. 5% (vol/vol) PEG-
388   4000, 0.025% (vol/vol) Tween 20, 100 µM double-stranded adapter, and 0.1
389   U/µL T4 DNA ligase. The library strand complementary to the original single-
390   stranded template molecule was then heat denatured and eluted in 25 µL TET
391   buffer.
392
393   The libraries were then PCR amplified, incorporating unique 8 bp index
394   sequences within both P5 and P7 adapters, using AccuPrime Pfx polymerase
395   in 80 µL reactions with the following reagent concentrations: 1x AccuPrime
396   Pfx reaction mix, 0.4 µM each of P5 and P7 indexing primers, and 0.025U/µL
397   AccuPrime Pfx polymerase. The appropriate number of cycles was
398   determined in advance by qPCR analysis of the unamplified library to identify
399   the cycle number corresponding to the point of inflection of the qPCR

400 amplification curve, correcting for differing reaction volume and template
401 amount in the subsequent library amplification PCR. The qPCR analysis
402 involved 10 µL reactions with the following reagent concentrations: 1x SYBR
403 green qPCR master mix, 0.2µM each of IS7 and IS8 amplification primers,
404 and 0.2% of the unamplified library. After amplification, the indexed libraries
405 were quantified using a TapeStation 2200 instrument (Agilent) with D1000
406 screen tape and reagents, and a Qubit with dsDNA HS Assay kit.
407
408 We then performed test sequencing of the single stranded libraries on an
409 Illumina NextSeq 500 sequencing platform [28] using the custom CL72 R1
410 sequencing primer [27] generating approximately one million 75 bp single-end
411 reads for each library. Mapping of this test data (for methodological details
412 refer to the "Data processing" section below) to the reference genome
413 assembly of the giant panda [5, 6] indicated that the two libraries initially
414 prepared from the ancient giant panda bone did not contain a sufficient
415 content of endogenous DNA for deeper sequencing (Table S1). We then took
416 a further ten bone powder aliquots and pretreated them with 0.5% (vol/vol)
417 bleach solution (sodium hypochlorite) at room temperature for 15 min before
418 DNA extraction, in an attempt to reduce the proportion of contaminant DNA
419 [24, 25]. Qubit quantification indicated low concentrations for the resulting
420 DNA extracts (maximally 0.766 ng/µL), and so 20 µL of each extract was
421 used to prepare a further ten single-stranded libraries, respectively. Library
422 preparation followed the procedure described above except that the Klenow
423 Fragment of DNA polymerase I was used for the fill-in step [25]. This was
424 carried out in 50 µL reactions with the following reagent concentrations: 1x
425 Klenow buffer, 200 µM of each dNTP, 2 µM CL9_Phos extension primer, and
426 10 U/µL Klenow Fragment of DNA polymerase I. Since the resulting
427 molecules are blunt-ended, removal of overhangs using Bst 2.0 was not
428 required. Test sequencing of these pretreated libraries indicated that they
429 provided sufficient endogenous data yield for deeper sequencing (Table S1).
430 Therefore, these ten libraries were pooled in equal molarity and sequenced on
431 a single lane of the Illumina HiSeq 4000 platform using the custom CL72 R1
432 sequencing primer and the Gesaffelstein index 2 sequencing primer [28]
433 producing 100 bp paired-end reads.
434
435 To assess library complexity, we used lc_extrap in the Preseq package (http://
436 smithlabresearch.org/software/preseq/) to predict the endogenous data yield
437 from further sequencing of the pretreated libraries. The ten libraries varied in
438 their predicted complexity (Table S3) and were re-pooled based on the
439 Preseq result in order to minimise the overall level of sequence duplication
440 during further sequencing on five HiSeq lanes (Table S1).
441
442 ***Data processing***
443 All data processing was carried out within the BEARCAVE v.ce78f40 data
444 analysis and storage environment (available at:
445 https://github.com/nikolasbasler/BEARCAVE), which provides a resource for
446 data processing and the establishment of a common sequencing data
447 repository. The BEARCAVE v.ce78f40 distribution is freely available and can
448 be used to replicate the described analyses.
449

450   For data processing using BEARCAVE, raw reads from each library were
451   treated independently. Cutadapt v1.12 [29] was used to trim Illumina adapter
452   sequences from all reads and discard sequences shorter than 30 bp. Flash
453   v1.2.11 [30] was used to merge overlapping read pairs. Unmerged paired
454   reads were discarded as long DNA fragments are likely to represent modern
455   contamination. The merged reads were mapped to the both the nuclear
456   genome assembly of the giant panda [6] and the polar bear [31] using the
457   "aln" and "samse" algorithms in bwa v0.7.15 [32]. For mapping to polar bear,
458   the number of allowed mismatches was relaxed by setting the -n flag in bwa
459   aln to 0.01 rather than the default 0.04, replicating the approach used by a
460   previous study which mapped reads in the opposite direction, from polar bear
461   to the giant panda reference [12]. Sequences with a map quality score less
462   than 30 were removed and the alignment sorted by 5' mapping position using
463   the "view" and "sort" algorithms in samtools v1.3.1 [33]. Potential PCR
464   duplicates generated during library amplification were eliminated by using
465   "rmdup" in samtools. The resulting bam files for each library were then
466   merged into a single bam file using samtools "merge". Full details of mapping
467   statistics are provided in (Table S1).

468

469   Short read data of 49 modern giant pandas (average 4.7x fold coverage) [2]
470   and a polar bear ([31], accession SRS463472) were downloaded from the
471   European Nucleotide Archive and processed using identical methods, except
472   that both merged and unmerged read pairs were used for mapping and the
473   relaxed mismatch setting was used when mapping polar bear reads to the
474   giant panda reference genome assembly.

475
476
477   **QUANTIFICATION AND STATISTICAL ANALYSIS**
478
479   ***Ancient DNA authenticity and contamination***
480   We checked for the presence on DNA fragmentation and cytosine
481   deamination typical of ancient DNA using the program mapDamage v2.0.8
482   [34], with merged reference scaffolds and Bayesian estimation disabled. The
483   alignment length distribution of reads mapped to the giant panda reference
484   nuclear genome revealed high levels of DNA fragmentation (Figure S1). 94%
485   of alignments are < 100 bp, and the modal alignment length is 31 bp, which
486   represents an overestimate since reads < 30 bp were discarded prior to
487   mapping. The mapped reads also show elevated levels of $C \rightarrow T$ substitutions
488   relative to the reference genome at their terminal ends (Figure S1). These
489   patterns are consistent with the postmortem degradation of endogenous
490   molecules expected for ancient samples, supporting the authenticity of our
491   data.

492

493   We investigated potential modern mammalian contamination of the ancient
494   giant panda sample by comparing the percentage of reads mapping uniquely
495   to the giant panda reference genome with the percentage mapping uniquely
496   to the genomes of potential contaminating mammals using the program
497   FastQscreen v0.10.0 [35], using bwa v0.7.8 aligner and default parameters.
498   The potential contaminating genomes used in the comparison were human,
499   cow, pig, cat, dog and mouse. The proportion of reads mapping uniquely to

500 panda in this analysis was 7.75%, which is ~100-fold greater than values for
501 any other genome tested (maximum 0.07% for human) (Table S2). Although
502 these values do not represent absolute measures of contamination, their ratio
503 is indicative of a very low ratio of endogenous to contaminant DNA in the
504 ancient giant panda sample, and is further likely to be an overestimate due to
505 greater quality and contiguity of the test genomes in comparison to the giant
506 panda reference.
507
508 ***Mitochondrial phylogeny***
509 Modern [2] and ancient giant panda reads were mapped to a giant panda
510 mitochondrial genome ([36]; GenBank accession FM177761.1) using the
511 procedures described above (see Table S1). All read alignments were
512 checked by eye for the presence of polymorphic positions, which would
513 suggest that an appreciable proportion of reads mapping derive from nuclear
514 mitochondrial DNA segments (NUMTs), but such sites were not observed in
515 any read alignment. A consensus fasta sequence based on maximum
516 effective base depth [37], which takes into account both base and mapping
517 quality scores, was generated using "doFasta 3" in ANGSD v0.916 [38]. The
518 mitochondrial genomes of the 49 modern giant pandas [2] and that of the
519 ancient individual analysed here were aligned along with the published
520 mitochondrial genome sequence of an ancient giant panda from Guangxi,
521 China [23], using the MUSCLE algorithm [39] implemented in the software
522 MEGA X v10.0.5 [40], with default parameters. The resulting sequence
523 alignment was checked by eye and a 367 bp section of the d-loop containing
524 a microsatellite repeat motif was removed as this cannot be reliably
525 reconstructed from short read data. The final alignment comprised 16,445
526 aligned positions of which 245 were variable and 119 were parsimony
527 informative.
528
529 Phylogenetic relationships and coalescence times of the mitochondrial
530 sequences were then estimated using BEAST v1.8.2 [41]. This analysis was
531 based on a previous analysis of mitochondrial sequences of modern and
532 ancient giant pandas [4]. Phylogeny and coalescence times were estimated
533 under a piecewise-constant Bayesian Skyline tree model with 10 groups,
534 assuming a strict molecular clock and a GTR+G substitution model. Time
535 calibration was achieved by fixing the tip dates of the ancient samples to their
536 median calibrated radio-carbon ages and by applying a normal prior on the
537 coalescence time of all modern panda haplotypes with a mean age of 72,000
538 years and a standard deviation of 10,000 years, based on a previous study
539 [22]. The per-lineage substitution rate was estimated within an open, uniform
540 prior of 0–20% per million years. Default settings were retained for all other
541 priors. The MCMC chain ran for sufficient time to achieve convergence and
542 adequate posterior sampling of all parameters (effective sample sizes > 200),
543 determined using the program Tracer v1.6 [42]. The maximum clade
544 credibility tree was selected from the posterior sample with node heights
545 centred on the median from the posterior sample using TreeAnnotator v1.8.2
546 [43], and visualised in FigTree v1.4.2 [44] (Figure S2).
547
548 ***Relationships among giant panda genomes***

549 Individuals of unknown provenance and captive individuals of mixed
550 population ancestry (indicated in Figure S2) were excluded from these
551 analyses. The giant panda genome assembly was used as mapping
552 reference. A covariance matrix was calculated using single base identity by
553 state (IBS) in ANGSD v0.916, with the following filters applied. Transition sites
554 were identified using genotype likelihoods and excluded. Singleton sites were
555 excluded (1/N < -minFreq < 2/N, where N = number of individuals). We
556 furthermore only considered sites without missing data (-minInd N), a
557 minimum base quality score of 30 (-minQ 30), minimum mapping quality score
558 of 30 (-minMapQ 30), and minimum scaffold length of 1 Mb. PCA of the
559 covariance matrix was then carried out using the "eigen" function in R [45],
560 which was also used to visualise the results. The removal of singletons in this
561 analysis provides an effective means of removing sequencing errors, which
562 are known to occur in high abundance in ancient datasets, but this approach
563 is sensitive to unbalanced sampling of populations [9]. Since the ancient
564 population investigated here is represented by only one individual, many
565 alleles unique to that population will have also been removed. This effect
566 makes this analysis highly conservative since the divergence of the ancient
567 population from modern ones will be underestimated. The observation of the
568 ancient genome as distinct under this conservative approach thus provides
569 robust support for the ancient population being distinct.
570
571 Relationships assuming a phylogenetic model of evolution were estimated
572 using neighbour-joining phylogenetic analysis. A distance matrix was
573 calculated using ANGSD including the polar bear as outgroup and applying
574 the same filters as used for PCA. The neighbour-joining tree was then
575 calculated using the "nj" function and rooted using the "root" function in the R
576 package ape [46]. Support for the ancient giant panda as basal to the modern
577 giant panda clade was further assessed using an approach based on D
578 statistics [7, 8]. Specifically, for each pair of modern giant pandas, we
579 calculated the excess of derived alleles that a modern giant panda shares
580 with the ancient giant panda and not with the other modern giant panda, and
581 compared these to the excess of derived alleles that a modern giant panda
582 shares with the other modern giant panda and not with the ancient giant
583 panda. These values correspond, respectively, to D statistics calculated for
584 tree topologies (((ancient,modern),modern),outgroup) and
585 (((modern,modern),ancient),outgroup). For the latter, D values were converted
586 to their absolute value, which effectively places the modern panda that shares
587 more derived alleles with the ancient panda in P2. Consistently lower D
588 values for a particular topology across all comparisons supports that topology
589 as correct since D values in the alternative topology are inflated because they
590 reflect derived alleles shared through both admixture and direct ancestry,
591 whereas those for the correct topology reflect only derived alleles shared
592 through admixture. D statistics for this topology test were calculated in
593 ANGSD using single read sampling (doAbbababa 1), requiring minimum base
594 and map qualities of 30, excluding transitions, and only considering scaffolds
595 > 1 Mb, with the polar bear as outgroup. Statistical support was assessed
596 using a weighted-block jackknife test using 5 Mb non-overlapping blocks, with
597 absolute Z-scores > 3 considered as supported.
598

Admixture tests

The polar bear genome assembly was used as mapping reference for all admixture tests, since preliminary analyses using the giant panda as mapping reference suggested an effect of mapping reference bias. The polar bear reference sequence was also used as outgroup. We computed D-statistics for the topology (((modern,modern),ancient),polar bear). Although increased rates of error in ancient datasets have been show to confound D statistics [9], these tests should not be substantially affected since the ancient individual is in P3. Specifically, assuming an equal occurrence of singleton sites in P1 and P2, errors in the P3 individual should not cause an imbalance in the frequency of either ABBA or BABA sites. D-statistics were calculated as described above. For significance testing, we applied the weighted block jackknife test using 5 Mb non-overlapping blocks, with D values more than three standard errors different from zero (Z > 3) considered as statistically significant. The phylogenetic test of directional admixture was based on that described in a previous study [11]. A majority-rule consensus sequence for scaffolds > 1 Mb was generated for the test individuals using ANGSD (-doFasta 2). A custom perl script was then used to divide the aligned sequences into non-overlapping 100 kb blocks. Blocks where any single individual had > 50% missing data were excluded, and the remainder converted into binary characters to exclude transitions (R: 0, Y: 1). 14,933 blocks remained after filtering. The phylogeny of each block was then computed under the BINGAMMA model with RAxML v8.2.10 [47] using the polar bear as outgroup to root the trees. The occurrence of each of the 15 possible rooted tree topologies was then counted.

DATA AND SOFTWARE AVAILABILITY

The raw fastq DNA sequence data files generated from the ancient giant panda bone sample have been deposited in the European Nucleotide Archive under ID codes ERX3266492 to ERX3266503, and ERX3266568 to ERX3266597.

References

1.   Jablonski, N.G., Ji, X., Liu, H., Li, Z., Flynn, L.J., and Li, Z. (2012). Remains of Holocene giant pandas from Jiangdong Mountain (Yunnan, China) and their relevance to the evolution of quaternary environments in south-western China. Historical Biology *24*, 527-536.

2.   Zhao, S., Zheng, P., Dong, S., Zhan, X., Wu, Q., Guo, X., Hu, Y., He, W., Zhang, S., Fan, W., et al. (2013). Whole-genome sequencing of giant pandas provides insights into demographic history and local adaptation. Nature Genetics *45*, 67-71.

3.   Sheng, G.-L., Barlow, A., Cooper, A., Hou, X.-D., Ji, X.-P., Jablonski, N., Zhong, B.-J., Liu, H., Flynn, L., Yuan, J.-X., et al. (2018). Ancient DNA from Giant Panda (*Ailuropoda melanoleuca*) of South-Western China Reveals Genetic Diversity Loss during the Holocene. Genes *9*, 198. https://doi.org/10.3390/genes9040198

4.   Barlow, A., Sheng, G.-L., Lai, X.-L., Hofreiter, M., and Paijmans, J.L.

649 (2018). Once lost, twice found: Combined analysis of ancient giant
650 panda sequences characterises extinct clade. Journal of Biogeography
651 *45*, 1-3.
652 5. Hu, Y., Wu, Q., Ma, S., Ma, T., Shan, L., Wang, X., Nie, Y., Ning, Z.,
653 Yan, L., Xiu, Y., et al. (2017). Comparative genomics reveals
654 convergent evolution between the bamboo-eating giant and red
655 pandas. Proc. Natl. Acad. Sci. USA *114*, 1081-1086.
656 6. Li, R., Fan, W., Tian, G., Zhu, H., He, L., Cai, J., Huang, Q., Cai, Q., Li,
657 B., Bai, Y., et al. (2010). The sequence and de novo assembly of the
658 giant panda genome. Nature *463*, 311-317.
659 7. Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher,
660 M., Patterson, N., Li, H., Zhai, W., Fritz, M.H., et al. (2010). A draft
661 sequence of the Neandertal genome. Science *328*, 710-722.
662 8. Durand, E.Y., Patterson, N., Reich, D., and Slatkin, M. (2011). Testing
663 for ancient admixture between closely related populations. Molecular
664 Biology and Evolution *28*, 2239-2252.
665 9. Barlow, A., Hartmann, S., Gonzalez, J., Hofreiter, M., and Paijmans,
666 J.L.A. (2018). Consensify: a method for generating pseudohaploid
667 genome sequences from palaeogenomic datasets with reduced error
668 rates. bioRxiv DOI: 10.1101/498915.
669 10. Günther, T., and Nettelblad, C. (2018). The presence and impact of
670 reference bias on population genomic studies of prehistoric human
671 populations. bioRxiv DOI: 10.1101/487983.
672 11. Barlow, A., Cahill, J.A., Hartmann, S., Theunert, C., Xenikoudakis, G.,
673 Fortes, G.G., Paijmans, J.L.A., Rabeder, G., Frischauf, C., Grandal-
674 d'Anglade, A., et al. (2018). Partial genomic survival of cave bears in
675 living brown bears. Nature Ecology & Evolution *2*, 1563-1570.
676 12. Cahill, J.A., Heintzman, P.D., Harris, K., Teasdale, M.D., Kapp, J.,
677 Soares, A.E.R., Stirling, I., Bradley, D., Edwards, C.J., Graim, K., *et al.*
678 (2018). Genomic Evidence of Widespread Admixture from Polar Bears
679 into Brown Bears during the Last Ice Age. Mol. Biol. Evol. *35(5)*, 1120-
680 1129. DOI: 10.1093/molbev/msy018
681 13. Wei, F., Hu, Y., Zhu, L., Bruford, M.W., Zhan, X., and Zhang, L. (2012).
682 Black and white and read all over: the past, present and future of giant
683 panda genetics. Molecular Ecology *21*, 5660-5674.
684 14. Zhan, X., Li, M., Zhang, Z., Goossens, B., Chen, Y., Wang, H., Bruford,
685 M.W., and Wei, F. (2006). Molecular censusing doubles giant panda
686 population estimate in a key nature reserve. Current Biology *16*, R451-
687 452.
688 15. Swaisgood, R., Wang, D., and Wei, F. (2016). *Ailuropoda melanoleuca*,
689 giant panda. IUCN Redlist of Threatened Species 2016,
690 e.T712A102080907.
691 16. Hu, Y., Qi, D., Wang, H., and Wei, F. (2010). Genetic evidence of
692 recent population contraction in the southernmost population of giant
693 pandas. Genetica *138*, 1297-1306.
694 17. Wei, F., Hu, Y., Yan, L., Nie, Y., Wu, Q., and Zhang, Z. (2015). Giant
695 pandas are not an evolutionary cul-de-sac: evidence from
696 multidisciplinary research. Molecular Biology and Evolution *32*, 4-12.
697 18. Zhang, B., Li, M., Zhang, Z., Goossens, B., Zhu, L., Zhang, S., Hu, J.,
698 Bruford, M.W., and Wei, F. (2007). Genetic viability and population

699   history of the giant panda, putting an end to the "evolutionary dead
700   end"? Molecular Biology and Evolution *24*, 1801-1810.

701 19. Hu, Y., Zhan, X., Qi, D., and Wei, F. (2010). Spatial genetic structure
702   and dispersal of giant pandas on a mountain-range scale.
703   Conservation Genetics *11*, 2145-2155.

704 20. Wei, F., Costanza, R., Dai, Q., Stoeckl, N., Gu, X., Farber, S., Nie, Y.,
705   Kubiszewski, I., Hu, Y., Swaisgood, R., et al. (2018). The Value of
706   Ecosystem Services from Giant Panda Reserves. Current Biology *28*,
707   2174-2180.

708 21. Zhu, J., and Long, Z. (1983). The vicissitudes of the giant panda. Acta
709   Zoologica Sinica *29*, 93-104.

710 22. Hofreiter, M., Paijmans, J.L.A., Goodchild, H., Speller, C.F., Barlow, A.,
711   Fortes, G.G., Thomas, J.A., Ludwig, A., and Collins, M.J. (2015). The
712   future of ancient DNA: Technical advances and conceptual shifts.
713   BioEssays *37*, 284-293.

714 23. Min-Shan Ko, A., Zhang, Y., Yang, M.A., Hu, Y., Cao, P., Feng, X.,
715   Zhang, L., Wei, F., and Fu, Q. (2018). Mitochondrial genome of a
716   22,000-year-old giant panda from southern China reveals a new panda
717   lineage. Current Biology *28*, R693-R694.

718 24. Basler, N., Xenikoudakis, G., Westbury, M.V., Song, L., Sheng, G., and
719   Barlow, A. (2017). Reduction of the contaminant fraction of DNA
720   obtained from an ancient giant panda bone. BMC Res Notes *10*, 754.

721 25. Korlevic, P., Gerber, T., Gansauge, M.T., Hajdinjak, M., Nagel, S.,
722   Aximu-Petri, A., and Meyer, M. (2015). Reducing microbial and human
723   contamination in DNA extractions from ancient bones and teeth.
724   BioTechniques *59*, 87-93.

725 26. Dabney, J., Knapp, M., Glocke, I., Gansauge, M.T., Weihmann, A.,
726   Nickel, B., Valdiosera, C., Garcia, N., Pääbo, S., Arsuaga, J.L., et al.
727   (2013). Complete mitochondrial genome sequence of a Middle
728   Pleistocene cave bear reconstructed from ultrashort DNA fragments.
729   Proceedings of the National Academy of Sciences of the United States
730   of America *110*, 15758-15763.

731 27. Gansauge, M.T., and Meyer, M. (2013). Single-stranded DNA library
732   preparation for the sequencing of ancient or damaged DNA. Nature
733   Protocols *8*, 737-748.

734 28. Paijmans, J.L., Baleka, S., Henneberger, K., Taron, U.H., Trinks, A.,
735   Westbury, M., and Barlow, A. (2017). Sequencing single-stranded
736   libraries on the Illumina NextSeq500 platform. arXiv:1711.11004v1, 1-
737   5.

738 29. Martin, M. (2011). Cutadapt removes adapter sequences from high-
739   throughput sequencing reads. EMBnet.journal *17*, 10.

740 30. Magoč, T., and Salzberg, S.L. (2011). FLASH: fast length adjustment
741   of short reads to improve genome assemblies. Bioinformatics *27*, 2957-
742   2963.

743 31. Liu, S., Lorenzen, E.D., Fumagalli, M., Li, B., Harris, K., Xiong, Z.,
744   Zhou, L., Korneliussen, T.S., Somel, M., Babbitt, C., et al. (2014).
745   Population genomics reveal recent speciation and rapid evolutionary
746   adaptation in polar bears. Cell *157*, 785-794.

747 32. Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment
748   with Burrows-Wheeler transform. Bioinformatics *26*, 589-595.

749 33. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N.,
750   Marth, G., Abecasis, G., Durbin, R., and Subgroup, G.P.D.P. (2009).
751   The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*,
752   2078-2079.
753 34. Jonsson, H., Ginolhac, A., Schubert, M., Johnson, P.L., and Orlando, L.
754   (2013). mapDamage2.0: fast approximate Bayesian estimates of
755   ancient DNA damage parameters. Bioinformatics *29,* 1682-1684.
756 35. Wingett, S.W., and Andrews, S. (2018). FastQ Screen: A tool for multi-
757   genome mapping and quality control. F1000Res *7*, 1338.
758 36. Krause, J., Unger, T., Nocon, A., Malaspinas, A.S., Kolokotronis, S.O.,
759   Stiller, M., Soibelzon, L., Spriggs, H., Dear, P.H., Briggs, A.W., et al.
760   (2008). Mitochondrial genomes reveal an explosive radiation of extinct
761   and extant bears near the Miocene-Pliocene boundary. BMC
762   Evolutionary Biology *8*, 220.
763 37. Wang, Y., Lu, J., Yu, J., Gibbs, R.A., and Yu, F. (2013). An integrative
764   variant analysis pipeline for accurate genotype/haplotype inference in
765   population NGS data. Genome Research *23*, 833-842.
766 38. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD:
767   Analysis of Next Generation Sequencing Data. BMC Bioinformatics *15*,
768   356.
769 39.Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high
770   accuracy and high throughput. Nucleic Acids Research *32*, 1792-1797.
771 40.Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA
772   X: Molecular evolutionary genetics analysis across computing
773   platforms. Molecular Biology and Evolution *35*, 1547-1549.
774 41. Drummond, A.J., Suchard, M.A., Xie, D., and Rambaut, A. (2012).
775   Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular
776   Biology and Evolution *29*, 1969-1973.
777 42.Rambaut, A., Suchard, M.A., Xie, D., and Drummond, A.J. (2014). Tracer
778   v1.6. Available from: http://tree.bio.ed.ac.uk/software/tracer/.
779 43. Helfrich, P., Rieb, E., Abrami, G., Lucking, A., and Mehler, A. (2018).
780   TreeAnnotator: Versatile visual annotation of hierarchical text relations.
781   LREC 2018: 11th edition of the Language Resources and Evaluation
782   ConferenceAt: Miyazaki, Japan.
783 44. Rambaut, A. (2009). FigTree, version 1.4.3. Available from:
784   http://tree.bio.ed.ac.uk/software/figtree.
785 45. Team, R.C. (2017). R: A language and environment for statistical
786   computing. Retrieved from: http://www.r-project.org/.
787 46. Paradis, E., Claude, J., and Strimmer, K. (2004). APE: Analyses of
788   Phylogenetics and Evolution in R language. Bioinformatics *20*, 289-
789   290.
790 47. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic
791   analysis and post-analysis of large phylogenies. Bioinformatics *30*,
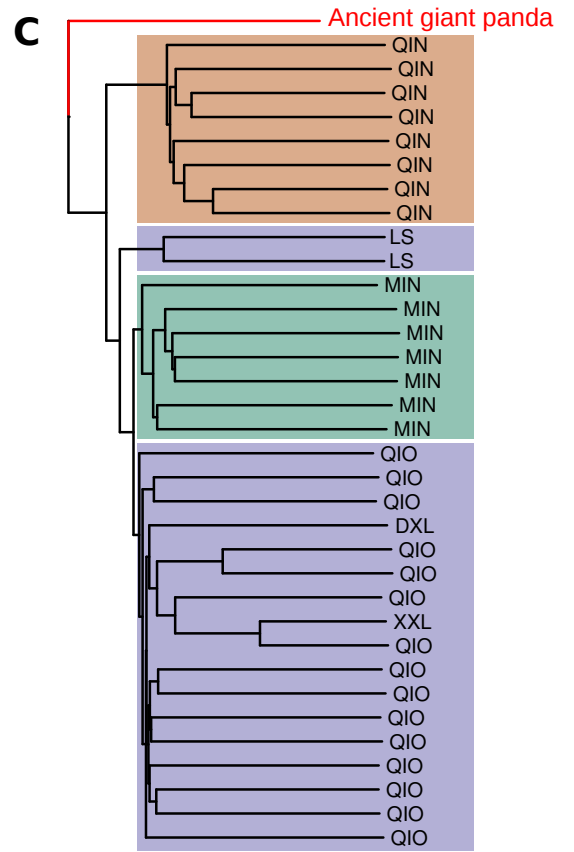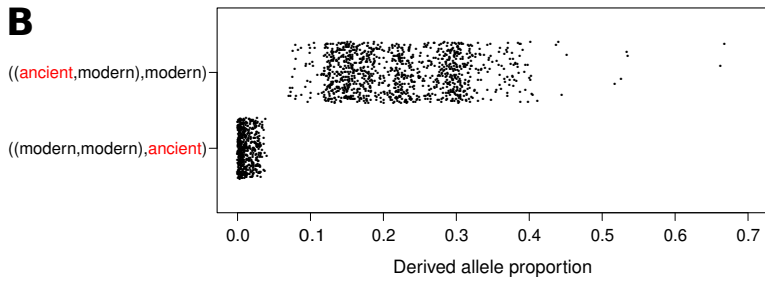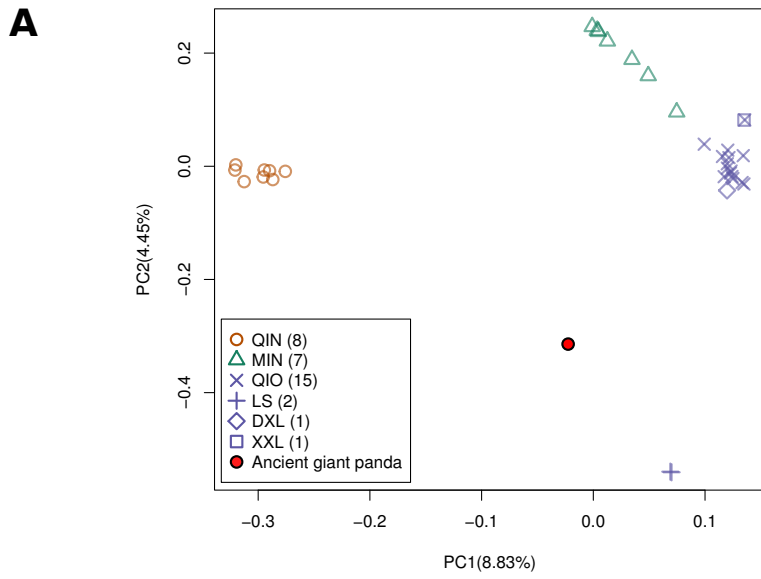792   1312-1313.
793

MIN

QIN

QIO

DXL

XXL

LS

Jiangdongshan

A

PC2(4.45%)

- ○ QIN (8)
- △ MIN (7)
- × QIO (15)
- + LS (2)
- ◇ DXL (1)
- □ XXL (1)
- ● Ancient giant panda

PC1(8.83%)

B

((ancient,modern),modern)

((modern,modern),ancient)

Derived allele proportion

C

Ancient giant panda

QIN
QIN
QIN
QIN
QIN
QIN
QIN
QIN
LS
LS
MIN
MIN
MIN
MIN
MIN
MIN
MIN
QIO
QIO
QIO
DXL
QIO
QIO
XXL
QIO
QIO
QIO
QIO
QIO
QIO
QIO
QIO

D value
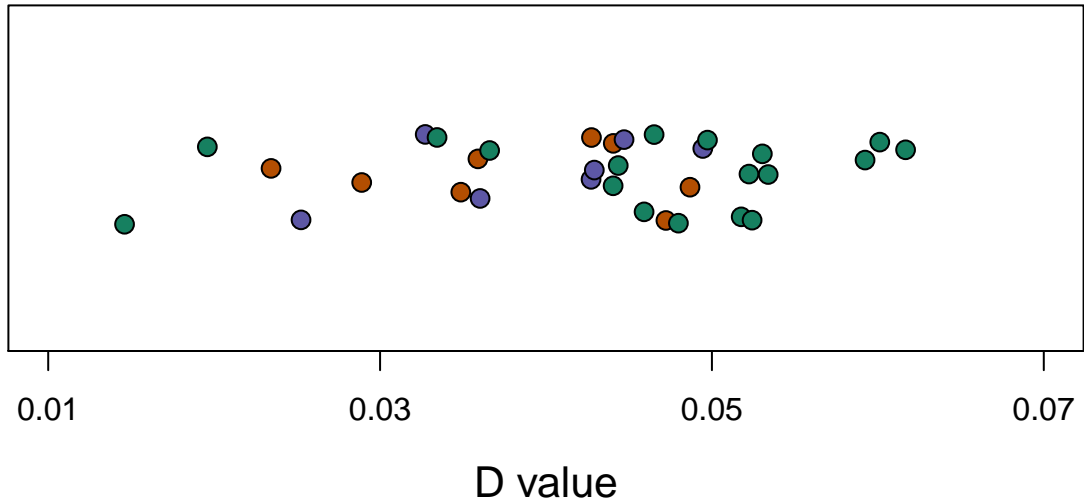
KEY RESOURCES TABLE

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOUR	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
Guanidine hydrochloride	Roth	Cat#0037.1
QIAGEN MinElute kit	Qiagen	Cat#28004
Critical Commercial Assays		
D1000 Screen Tape (Tapestation2200)	Agilent	Cat#5067-5582
dsDNA HS Assay Kit (Qubit 2.0)	Therm ofisher	Cat#Q32851
Deposited Data		
aGP2-01_test	This paper	ENA: ERX3266492
aGP2-02_test	This paper	ENA: ERX3266493
aGP2-03_test	This paper	ENA: ERX3266494
aGP2-03_1st	This paper	ENA: ERX3266568
aGP2-03_2nd	This paper	ENA: ERX3266569
aGP2-03_3rd	This paper	ENA: ERX3266570
aGP2-04_test	This paper	ENA: ERX3266495
aGP2-04_1st	This paper	ENA: ERX3266571
aGP2-04_2nd	This paper	ENA: ERX3266572
aGP2-04_3rd	This paper	ENA: ERX3266573
aGP2-05_test	This paper	ENA: ERX3266496
aGP2-05_1st	This paper	ENA: ERX3266574
aGP2-05_2nd	This paper	ENA: ERX3266575
aGP2-05_3rd	This paper	ENA: ERX3266576
aGP2-06_test	This paper	ENA: ERX3266497
aGP2-06_1st	This paper	ENA: ERX3266577
aGP2-06_2nd	This paper	ENA: ERX3266578
aGP2-06_3rd	This paper	ENA: ERX3266579

aGP2-07_test	This paper	ENA: ERX3266498
aGP2-07_1st	This paper	ENA: ERX3266580
aGP2-07_2nd	This paper	ENA: ERX3266581
aGP2-07_3rd	This paper	ENA: ERX3266582
aGP2-08_test	This paper	ENA: ERX3266499
aGP2-08_1st	This paper	ENA: ERX3266583
aGP2-08_2nd	This paper	ENA: ERX3266584
aGP2-08_3rd	This paper	ENA: ERX3266585
aGP2-09_test	This paper	ENA: ERX3266500
aGP2-09_1st	This paper	ENA: ERX3266586
aGP2-09_2nd	This paper	ENA: ERX3266587
aGP2-09_3rd	This paper	ENA: ERX3266588
aGP2-10_test	This paper	ENA: ERX3266501
aGP2-10_1st	This paper	ENA: ERX3266589
aGP2-10_2nd	This paper	ENA: ERX3266590
aGP2-10_3rd	This paper	ENA: ERX3266591
aGP2-11_test	This paper	ENA: ERX3266502
aGP2-11_1st	This paper	ENA: ERX3266592
aGP2-11_2nd	This paper	ENA: ERX3266593
aGP2-11_3rd	This paper	ENA: ERX3266594
aGP2-12_test	This paper	ENA: ERX3266503
aGP2-12_1st	This paper	ENA: ERX3266595
aGP2-12_2nd	This paper	ENA: ERX3266596
aGP2-12_3rd	This paper	ENA: ERX3266597
Oligonucleotides		

CL9_Phos extension primer: GTGACTGGAGTTCAGACGTGTGCTCTTCC*GA*TC*T (* = phosphothioate linkage)	[25]	Sigma Aldrich
CL9 extension primer: GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT	[27]	Sigma Aldrich
Double-stranded adapter Strand 1 (CL53): CGACGCTCTTC-ddC (ddC = dideoxycytidine) Strand 2 (CL73): [Phosphate]GGAAGAGCGTCGTGTAGGGAAAGAG*T*G*T*A (* = phosphothioate linkage)	[27]	Sigma Aldrich
CL78: AGATCGGAAG[C3Spacer] $_{10}$ [TEG-biotin] (TEG =triethylene glycol spacer)	[27]	Sigma Aldrich
P5 indexing primer: AATGATACGGCGACCACCGAGATCTACACnnnnnnnnACACTCTTTCCCTACACGACGCTCTT	[27]	Sigma Aldrich
P7 indexing primer: CAAGCAGAAGACGGCATACGAGATnnnnnnnnnGTGACTGGAGTTCAGACGTGT	[27]	Sigma Aldrich
IS7 amplification primer: ACACTCTTTCCCTACACGAC	[27]	Sigma Aldrich
IS8 amplification primer: GTGACTGGAGTTCAGACGTGT	[27]	Sigma Aldrich
CL72 R1 sequencing primer : ACACTCTTTCCCTACACGACGCTCTTCC	[27]	Sigma Aldrich
Gesaffelstein index 2 sequencing primer: GGAAGAGCGTCGTGTAGGGAAAGAGTGT	[28]	Sigma Aldrich
Software and Algorithms		
BEARCAVE ce78f40	-	https://github.com/nikolasbasler/BEARCAVE
Cutadapt v1.12	[29]	https://cutadapt.readthedocs.io/en/stable/index.html
Flash v1.2.11	[30]	https://ccb.jhu.edu/software/FLASH/
BWA v0.7.15 and v0.7.8	[32]	http://bio-bwa.sourceforge.net/
Samtools v1.3.1	[33]	https://sourceforge.net/projects/samtools/files/samtools/
PreSeq	-	http://smithlabresearch.org/software/preseq/
MapDamage v2.0.8	[34]	https://ginolhac.github.io/mapDamage/
FastQscreen v0.10.0	[35]	https://www.bioinformatics.babraham.ac.uk/projects/fastq_screen/
ANGSD v0.916	[38]	http://www.popgen.dk/angsd
MEGA X v10.0.5	[40]	https://www.megasoftware.net/dload_mac_beta
BEAST v1.8.2	[41]	http://beast.community/index.html
Tracer v1.6	[42]	https://github.com/beast-dev/tracer/

TreeAnnotator v1.8.2	[43]	http://beast.community/treeannotator
FigTree v1.4.2	[44]	http://tree.bio.ed.ac.uk/software/figtree/
RaxML v8.2.10	[47]	https://github.com/stamatak/standard-RAxML
Other		
Proteinase K	Promega	Cat#V3021
Zymo-spin V column extension reservoir	Zymo	Cat#C1016-50
Circligase II	Biozym	Cat#131402(CL9021K)
Endonuclease VIII	NEB	Cat#A0299S
Uracil-DNA glycosylase (Afu UDG)	NEB	Cat#M0279S
FastAP	Thermo Fisher	Cat#EF0651
MyOne C1 streptavidin beads	Thermo Fisher	Cat#65001
Bst 2.0 polymerase	NEB	Cat#M0537S
T4 DNA Polymerase	Thermo Fisher	Cat#EP0061
Buffer Tango (10x)	Thermo Fisher	Cat#BY5
T4 DNA ligase	Thermo Fisher	Cat#EL0011
Accuprime Pfx	Thermo Fisher	Cat#12344024
PEG-4000	Thermo Fisher	Cat#EP0061
Klenow fragment of DNA polymerase I	Thermo Fisher	Cat#EP0051
SYBR green PCR MasterMix	Thermo Fisher	Cat#4309155