

ARTICLE

<https://doi.org/10.1038/s42003-019-0734-6>

OPEN

# Whole-genome sequencing of 128 camels across Asia reveals origin and migration of domestic Bactrian camels

Liang Ming et al.<sup>#</sup>

The domestic Bactrian camels were treated as one of the principal means of locomotion between the eastern and western cultures in history. However, whether they originated from East Asia or Central Asia remains elusive. To address this question, we perform whole-genome sequencing of 128 camels across Asia. The extant wild and domestic Bactrian camels show remarkable genetic divergence, as they were split from dromedaries. The wild Bactrian camels also contribute little to the ancestry of domestic ones, although they share close habitat in East Asia. Interestingly, among the domestic Bactrian camels, those from Iran exhibit the largest genetic distance and the earliest split from all others in the phylogeny, despite evident admixture between domestic Bactrian camels and dromedaries living in Central Asia. Taken together, our study support the Central Asian origin of domestic Bactrian camels, which were then immigrated eastward to Mongolia where native wild Bactrian camels inhabit.

---

<sup>#</sup>A full list of authors and their affiliations appears at the end of the paper.

Camels (*Camelus*, *Camelini*) contain two extant domestic species, the one-humped dromedary (*Camelus dromedarius*) and the two-humped Bactrian camel (*Camelus bactrianus*)<sup>1,2</sup>. Although the former herds are mainly feed in North Africa and West Asia, the latter herds live in the cold desert areas of Northeast and Central Asia. The wild Bactrian camel (*Camelus ferus*), the only representative of the wild tribe *Camelini* as a result of the extinction of the wild dromedary<sup>3</sup>, is listed as critically endangered by the International Union for Conservation of Nature<sup>4</sup> with an estimation of a few hundreds to 2000 individuals<sup>5,6</sup>. Historically, the wild Bactrian camel was widely distributed throughout Asia, extending from the great bend of the Yellow River westward to central Kazakhstan (KAZA), but it can only be found in the Mongolian Gobi and the Chinese Taklimakan and Lop Noor deserts today<sup>7</sup>. Fossil and molecular evidence suggested that the ancestor of camels lived in North America and spread to Asia via the Bering land bridge around 11 or 16 million years ago<sup>8,9</sup>. Within the *Camelini*, the dromedaries and Bactrian camels were then split around 4 or 5 million years ago<sup>9,10</sup>. The domestication of camels, similar to many other innovations of domestic mammals such as horse-based transport<sup>11</sup>, greatly promoted human mobility and represented a great leap forward for human civilization. For example, the Bactrian camels were rightfully considered as the principal means of locomotion across the bridge between the eastern and western cultures in the time of the Silk Road<sup>12</sup>. Today, they still serve as valuable sources of meat, milk, and wool to people's livelihoods in arid and semi-arid areas<sup>1</sup>.

The origin of domestic dromedaries was recently revealed by world-wide sequencing of modern and ancient mitochondrial DNA (mtDNA), which suggested that they were at first domesticated in the southeast Arabian Peninsula<sup>13</sup>. However, the origin of domestic Bactrian camels remains controversial. One intuitive possibility was that the extant wild Bactrian camels were the progenitor of the domestic form, which were then dispersed from the Mongolian Plateau to the West gradually<sup>7,12</sup>. This hypothesis was supported by the presence of Camelid faunal remains at Neolithic sites near Mongolia (MG), although it was unclear these were the domestic as opposed to the wild ones<sup>12</sup>. Nevertheless, molecular studies based on mtDNAs<sup>9,14</sup> and Y chromosomes<sup>15</sup> discovered dramatic sequence variations between the wild and domestic Bactrian camels, suggesting that the extant wild Bactrian camel was a separate lineage<sup>14</sup>. Another possible place of origin was Iran (IRAN)<sup>1</sup>, where early skeletal remains of domestic Bactrian camels (around 2500–3000 BC) were discovered<sup>16</sup>. Although prehistoric mtDNAs of Bactrian camels supported the idea that the domestication took place in Central Asia rather than in MG or East Asia<sup>17</sup>, the incomplete archaeological findings and limited molecular markers provided little decisive information about the actual domestication history.

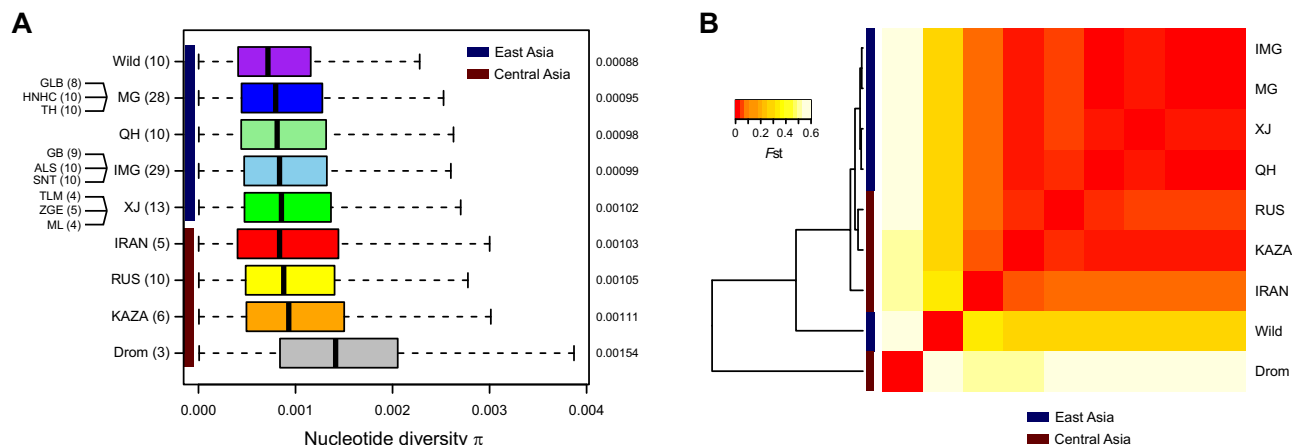
Whole-genome sequences contain much more molecular markers than mtDNAs, which were successfully used to portray the origin, migration, and admixture of humans<sup>18–20</sup> and domestic animals<sup>21</sup>, such as dogs<sup>22–25</sup> and pigs<sup>26–28</sup>. The published genome assembly of Bactrian camels<sup>10,29</sup> provides a new opportunity to examine the evolutionary relationship between the extant wild and domestic Bactrian camels, and trace their origin. In this study, we perform whole-genome sequencing of 128 camels including both domestic and extant wild Bactrian camels from their typical habitats. We include dromedaries as well, because they are not only the outgroup of Bactrian camels in phylogeny but also have a long history of hybridization with Bactrian camels in breeding practice<sup>2,30</sup>, especially in Central Asia. Our results support the Central Asian origin of domestic Bactrian camels and a roughly west-to-east route of migration back to the Mongolian Plateau.

## Results

**Sample collection and whole-genome sequencing.** A total of 105 domestic Bactrian camels across Asia, 19 wild Bactrian camels from Gobi-Altai region in MG, as well as 4 dromedaries from IRAN were gathered for this study (Supplementary Fig. 1 and Supplementary Table 1). The domestic Bactrian camels were chosen to cover as many major geographic regions as possible, including 55 from Inner MG (IMG), Xinjiang (XJ), and Qinghai of China, 28 from MG, 6 from KAZA, 10 from Russia (RUS) and 6 from IRAN. As a variety of local breeds were formed due to the wide utilization of domestic Bactrian camels in China and MG, eight different representative breeds were chosen from the regions. The other domestic Bactrian camels from Central Asia were living around the Caspian Sea.

After DNA extraction, individual genomes were sequenced to an average of 13× coverage (Supplementary Fig. 2 and Supplementary Table 2). The sequence reads were aligned to our previous genome assembly of the Bactrian camel<sup>29</sup> for variant calling. After stringent filtering (Supplementary Fig. 3), we totally identified 13.83 million single-nucleotide polymorphisms (SNPs) and 1.41 million small indels. Notably, the transition to transversion ratio of raw SNPs (2.29) was lower than that reported in dromedaries (2.31–2.34)<sup>31</sup>, but it was increased to 2.44 after the filtering procedures, suggesting a quality improvement of identified SNPs. Functional annotation of the variants indicated that about 63.10% of them were intergenic, 33.62% were intronic, and 0.94% were exonic (Supplementary Table 3). There were 13.73 million, 6.39 million, and 10.55 million variants identified in the domestic Bactrian camels, wild Bactrian camels, and dromedaries, respectively. Although dromedaries were more divergent from both of the Bactrian camel species in phylogeny, the domestic Bactrian camels shared more variants with the dromedaries (66.73%) than with the wild Bactrian camel (39.31%) (Supplementary Fig. 4) due to the tremendous reduction of genetic variants observed in the extant wild Bactrian camel and to gene flow between dromedaries and domestic Bactrian camels. Among the domestic Bactrian camels, there were 12.68 million and 11.61 million variants identified in the East Asian and Central Asian populations, respectively (Supplementary Fig. 4). Although the domestic camels sampled from East Asia were more than those from Central Asia, the variant count private to each population showed no significant bias between the two areas ( $P$ -value = 0.77, two-tailed  $t$ -test; Supplementary Table 4).

**Genetic diversity and differentiation.** For a more detailed comparison of the genetic diversity among different populations, we first removed 14 individuals showing close genetic relationship with the remaining others (Supplementary Table 5). The pairwise nucleotide diversity  $\pi$  (Fig. 1a) of dromedaries ( $1.54 \times 10^{-3}$ ) was significantly higher than that of Bactrian camels from all geographic regions ( $0.88 \times 10^{-3}$ – $1.11 \times 10^{-3}$ ; Supplementary Table 6), which was in contrast to previous heterozygosity estimates based on individual genomes<sup>10</sup>. One important reason could be the hybridization practice between dromedaries and Bactrian camels in Central Asia<sup>30</sup>. Among the Bactrian camels, the wild population showed the lowest  $\pi$  ( $0.88 \times 10^{-3}$ ) compared with all of the domestic populations (Fig. 1a and Supplementary Table 6). Although this result violated many cases that wild animals usually have higher genetic diversity than their domestic counterparts such as dogs<sup>24</sup>, pigs<sup>27</sup>, and rabbits<sup>32</sup>, it would happen in the case of endangered wild animals with an extremely small population size such as horses<sup>33</sup>. In addition, the domestic populations living in Central Asia generally showed a higher diversity ( $1.03 \times 10^{-3}$ – $1.11 \times 10^{-3}$ ) than those living in East Asia ( $0.95 \times 10^{-3}$ – $1.02 \times 10^{-3}$ ; Fig. 1a). The tendency was also



**Fig. 1 Genetic diversity and differentiation of the camel populations.** **a** Nucleotide diversity  $\pi$ . The boxplot shows  $\pi$  for  $2.0 \times 10^5$  10 kb-sliding windows across the genome. The geographic origin and sample size of each population are shown on the left and the average of  $\pi$  are shown on the right. Multiple local breeds were sampled for MG, IMG, and XJ. Individuals with close genetic relationships were removed. The boxplot elements are defined as follows: center line, median; box limits, the third and first quartiles; whiskers,  $1.5 \times$  interquartile range. **b** Pairwise population differentiation  $F_{st}$ . The heatmap represents average  $F_{st}$  for  $2.0 \times 10^5$  10 kb-sliding windows. The dendrogram represents hierarchical clustering of the populations based on  $F_{st}$ .

confirmed by the Watterson's  $\theta$  (Supplementary Fig. 5). Again, the hybridization with dromedaries in Central Asia could account for the higher diversity in the region.

We then measured the pairwise genetic distance between the camel populations by Weir's  $F_{st}$  (Fig. 1b). The result was well in agreement with the known phylogeny, which indicated that the dromedaries had the highest  $F_{st}$  with the Bactrian camels (0.54–0.64) and the wild Bactrian camels had the second highest  $F_{st}$  with the domestic ones (0.27–0.31). The differentiation among the domestic Bactrian camels was much lower, in line with their recent single origin. Interestingly, among the domestic Bactrian camels, those from IRAN displayed the largest divergence with all others (0.05–0.06). To validate the population differentiation, we constructed a neighbor-joining (NJ) tree for all individuals based on their pairwise identity-by-state (IBS) matrix (Supplementary Fig. 6). The NJ tree also supported a monophyletic clade of all domestic Bactrian camels, within which IRAN formed the deepest branch.

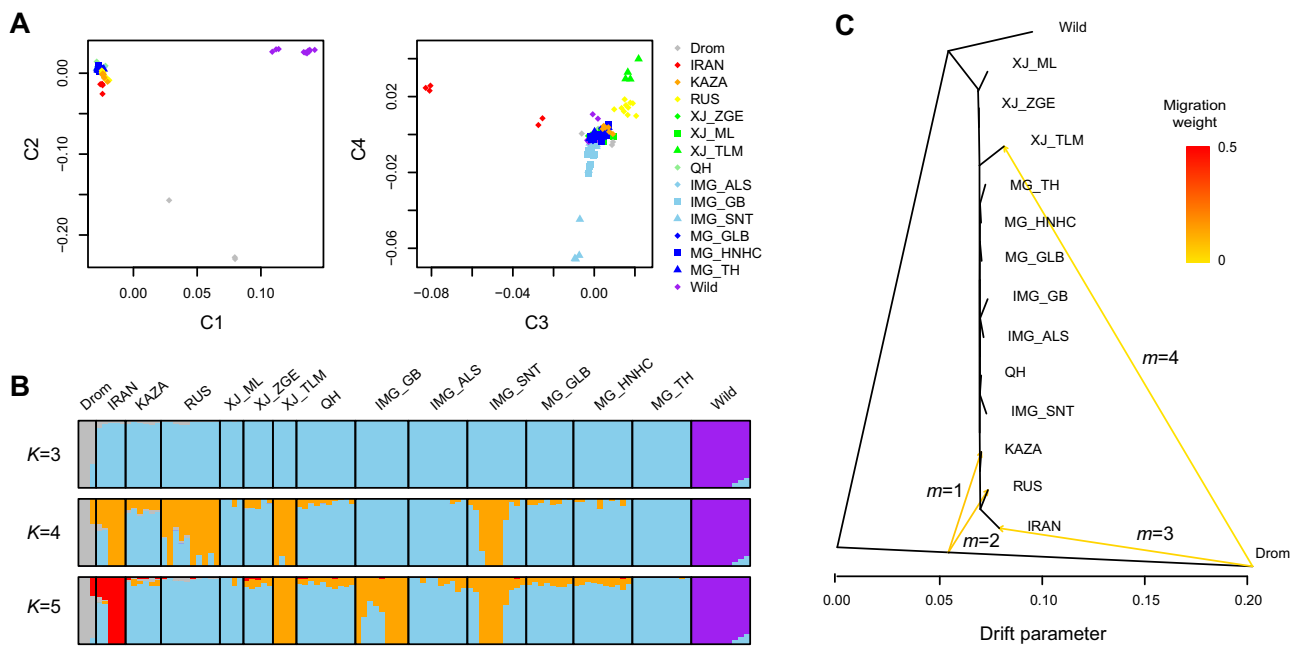
A potential issue with the population genomic estimates was the reference bias, where using a single reference genome would lead to low efficiency in variant calling for individuals that highly differed from it<sup>34</sup>. To investigate the bias, we compared the missing count of variants among the three species, taking the sequencing depth as a covariate (Supplementary Table 7). The analysis of variance (ANOVA) showed that although the domestic Bactrian camels had no significant difference with the wild ones ( $P$ -value = 0.50), they indeed had lower missing count than dromedaries ( $P$ -value =  $4.38 \times 10^{-3}$ ). To evaluate the impact of the bias on our estimates, we re-calculated the genetic diversity and  $F_{st}$  with just synonymous SNPs (Supplementary Fig. 7), as coding sequences were more likely to be invariant across species. As a result, the estimates based on the synonymous SNPs were in good agreement with the whole genome for all species, suggesting that the reference bias had only minor effects on our population genomic estimates.

**Population structure with admixture.** To reveal the overall population structure with potential admixture, we pruned the SNPs by removing those with high linkage disequilibrium and potential functional effects. The multidimensional scaling (MDS) analysis based on the pruned subset reproduced the similar result as the full set (Fig. 2a and Supplementary Fig. 8). As expected, the

dromedaries and wild Bactrian camels could be separated by the first and second coordinates, respectively. When the third coordinate was incorporated in the MDS, IRAN was separated from all other domestic Bactrian camels (Fig. 2a).

To estimate different ancestral proportions, we performed population structure analysis with Admixture<sup>35</sup> by assuming  $K$  ancestral populations (Fig. 2b). The cross-validation procedure supported that  $K=3$  was optimal (Supplementary Fig. 9), showing a clear division between the dromedaries, wild Bactrian camels, and domestic Bactrian camels. Evident introgression of domestic Bactrian camels into the Iranian dromedaries was observed, at least in one dromedary. Accordingly, the dromedary ancestry was prevalent in the Central Asian Bactrian camel populations including IRAN, KAZA, and RUS, with a proportion estimated as 1–10%. Moreover, we observed ancestry of domestic Bactrian camels in several wild individuals with a proportion of 7–15%. This could arise from ancestral polymorphisms, but it could also be caused by introgressive hybridization, which was observed with mtDNAs<sup>36</sup> and Y chromosomes<sup>15</sup>, and was proposed to threaten the genetic distinctiveness of the wild species. Surprisingly, the wild camels contributed nearly nothing to the ancestry of domestic populations, even to the Mongolian populations, which share close habitats with the wild camels. Although most domestic Bactrian camels lacked differentiation when  $K$  grew, IRAN was the first population to separate with a unique ancestry ( $K=5$ ; Fig. 2b).

As another method to examine the population structure with admixture, we inferred the population tree for the camels using TreeMix<sup>37</sup> (Fig. 2c). When no migration track was added, the tree topology again indicated that IRAN was the first population separated among all the domestic Bactrian camels (Supplementary Fig. 10). Increasing of migration tracks ( $m=1-3$ ) could greatly improve the fit of the model (Supplementary Fig. 11), which identified gene flows from dromedaries to domestic Bactrian camels in Central Asia, including KAZA, RUS, and IRAN with migration weights ranging from 4% to 9% (Supplementary Table 8). It was worth mentioning that although the migration track pointed at the end of the dromedary branch to IRAN, it pointed at the middle of the dromedary branch to KAZA and RUS (Fig. 2c). This could imply a ghost population related to the Iranian dromedary that contributed to the ancestry of KAZA and RUS. Additional migration track ( $m=4$ ) could



**Fig. 2 Population structure of the camels based on genome-wide SNPs. a** Multidimensional scaling (MDS) plot with coordinate 1–4 (C1–C4). **b** Admixture analysis assuming different number of ancestry  $K$ . The proportion of an individual’s genome assigned to each ancestry is represented by different colors. **c** TreeMix analysis with different assumption of migration events  $m$ . The migration weight is the proportion of ancestry received from the donor population.

continue to improve the fit of the model, which indicated migration of the dromedary to a XJ breed (Fig. 2c). Although TreeMix detected no strong signal of migration between the wild and domestic Bactrian camels, the residues showed moderate admixture between the wild and East Asian breeds (Supplementary Fig. 11). We then used the less-parameterized three- and four-population (F3/F4) test<sup>38</sup> to evaluate the statistical significance of these admixture events. Again, the F3 test strongly supported the admixture of dromedaries and Bactrian camels in KAZA, RUS, and IRAN (Supplementary Table 9). The more sensitive F4 test confirmed a significantly higher extent of admixture between dromedaries and Bactrian camels in Central Asia compared with those in East Asia (Supplementary Table 10). Among the latter, a higher extent of admixture with dromedaries was detected in XJ than in MG/IMG.

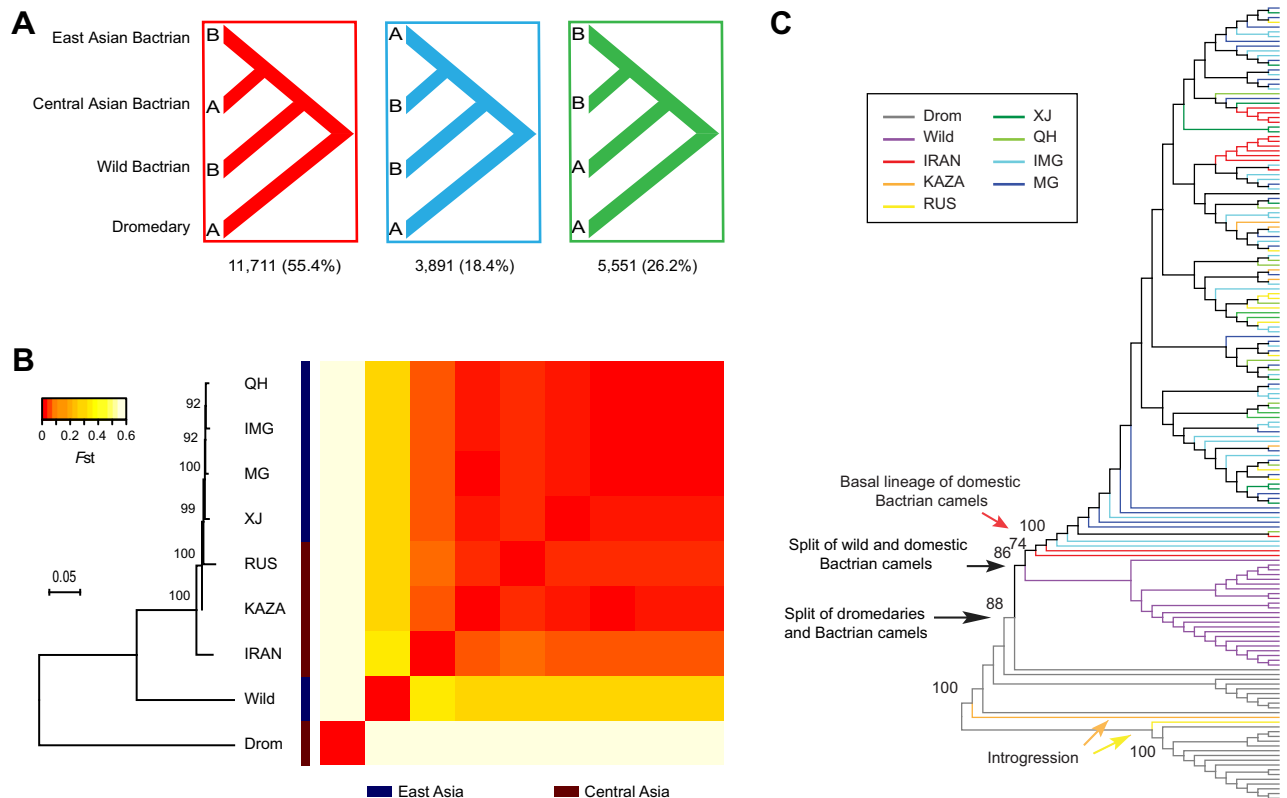
#### Evidence for Central Asian origin by removing introgression.

East and Central Asia were the two alternative regions of domestication for Bactrian camels based on archaeological evidences<sup>1,12,17</sup>, but the most probable one remained unsolved. Although we observed the largest genetic differentiation between the Iranian population and all the other domestic ones, the existence of admixture between dromedaries and Bactrian camels in Central Asian would weaken the support for origin inference. To reduce this effect, we attempted to remove the introgressed segments of dromedaries from the Bactrian camel genomes by using the “BABA/ABBA” test<sup>39</sup>. We grouped the East and Central Asian populations, and compared allele sharing between the two groups with dromedaries (Fig. 3a). As the ancestry of Bactrian camels in one dromedary, as well as the ancestry of domestic camels in three wild ones (Fig. 2b), would be confounding factors, we removed the four individuals in the analysis. We used the statistic  $f_{ab}$ , a robust version of the Patterson’s  $D$ , to locate introgressed segments<sup>40</sup> and applied a strict significance level of  $Z\text{-score} = 2$  by using the Jackknife procedure (Supplementary Fig. 12). In a total of 21,153 non-overlapping 100 kb segments across the genome, there were far more segments with putative

signals of introgression in the Central Asian populations (11,711,  $Z\text{-score} > 2$ ) than in the East Asian populations (3891,  $Z\text{-score} < -2$ ) as expected. We then performed the Admixture analysis based on the remaining segments and confirmed that the introgression of dromedaries were effectively reduced (Supplementary Fig. 13). Re-calculation of the pairwise  $F_{st}$  after removing introgression still indicated that IRAN was the most differentiated one (0.04–0.06) among all the domestic populations (Fig. 3b). To gain more insights into the population phylogeny, we reconstructed the NJ tree based on the pairwise  $F_{st}$  and performed the bootstrap test (Fig. 3b). It confirmed that IRAN was the first one to separate among all the domestic Bactrian populations, followed by KAZA and RUS. The Bayesian binary Markov Chain Monte Carlo (MCMC) analysis based on the phylogeny strongly supported Central Asia as the ancestral area of domestic Bactrian camels (probability = 99.78%) and a subsequent dispersal route from Central to East Asia (Supplementary Fig. 14).

As an independent evidence, we also reconstructed the maximum likelihood tree of full-length mtDNAs based on the 128 samples we sequenced in this study, as well as 39 additional samples available from Genbank (Fig. 3c and Supplementary Table 11). Introgression of mtDNAs could easily be identified and excluded from the tree. For example, two newly sequenced mtDNAs from KAZA and RUS were clustered with dromedaries. Within the clade of domestic Bactrian camels, although most camels from different geographic regions were mixed, two mtDNAs from IRAN formed the most basal branches of the domestic populations (Fig. 3c). The Bayesian binary MCMC analysis again supported the Central Asian origin of domestic Bactrian camels (probability = 76.43%).

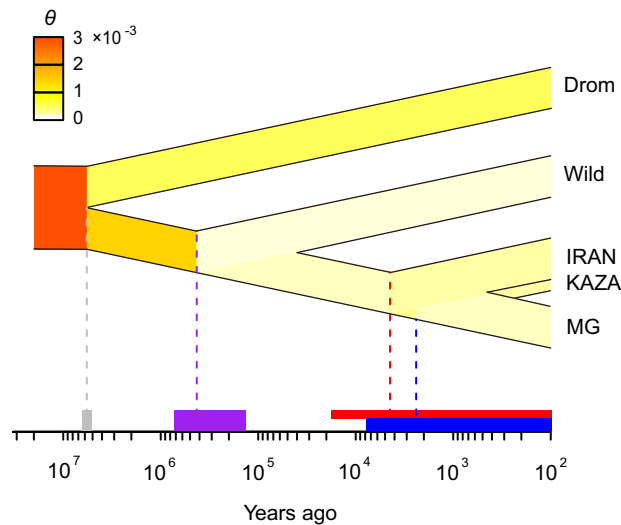
**Demographic history of Bactrian camels.** We performed several parametric modeling analyses to infer the demographic dynamics of the camels in history. Consistent with previous study<sup>10</sup>, the long-term trajectories of Bactrian camels based on the pairwise sequentially Markovian coalescent (PSMC) model<sup>41</sup> revealed a tremendous decrease in the effective population size of the



**Fig. 3 Identification of the origin of domestic Bactrian camels by removing introgression.** **a** BABA/ABBA analysis for introgression of dromedaries into domestic Bactrian camels. To focus on this introgression, one dromedary with the ancestry of Bactrian camels and three wild camels with the ancestry of domestic ones were removed. Number of 100 kb segments with significant  $f_0$  ( $|Z\text{-score}| > 2$ ) for each tree configuration is shown. **b** Neighbor-joining (NJ) tree of the populations after the introgressed segments were removed. The heatmap represents average pairwise  $F_{st}$  for  $5.1 \times 10^4$  10 kb-sliding windows. Bootstrap values of the NJ tree were calculated by randomly sampling five thousand 10 kb windows for 100 times. **c** Maximum likelihood tree of full-length mtDNAs. Populations are represented by different colors and sequences from Genbank are indicated by dots. Bootstrap values for main branches are labeled.

ancestral camels earlier than one million years ago (Supplementary Fig. 15). Although the long-term trajectories of the wild and domestic Bactrian camels were generally similar, they were obvious to diverge from each other as early as 0.4 million years ago, excluding the former as direct progenitors of the latter as previous mtDNA analyses<sup>9,14</sup>.

To explore the divergence time among the camel populations in more detail, we used the generalized phylogenetic coalescent sampler (G-PhoCS)<sup>42</sup>. Given the phylogeny of the camel populations, G-PhoCS could estimate the mutation-scaled population size and population divergence time based on unlinked neutral loci in individual genomes from each population. To reduce the model complexity, we only included dromedaries, wild Bactrian camels, and three representative populations (IRAN, KAZA, and MG) of domestic Bactrian camels (Supplementary Fig. 16 and Supplementary Table 12). According to Fig. 3b, IRAN and KAZA were the first two Central Asian populations to separate and the split of MG could indicate the dispersal from Central to East Asia. The age was calibrated by assuming the Bactrian camel and dromedary divergence of 5.73 million years according to the Timetree database<sup>43</sup>. When no migration band was incorporated, convergence of all parameter estimates could easily be achieved (Supplementary Fig. 17 and Supplementary Table 13). Similar to the PSMC results, the effective population size was generally decreased from ancestral to modern populations (Fig. 4). The divergence time between wild and domestic Bactrian camels was estimated as 0.43 million years



**Fig. 4 Parameter-based inference of demographic history with G-PhoCS.** The change in mutation-scaled effective population size  $\theta$  is represented by heat colors. The time in years were calibrated by the divergence time between dromedaries and Bactrian camels. 95% Confidence intervals are shown by bars on the time axis. The red and blue bar indicate IRAN-MG and KAZA-MG divergence, respectively. These estimates are based on the model without migration.



ago (95% confidence interval [CI]: 0.13–0.73 Mya; Fig. 4), which was slightly later than that based on mtDNAs (0.7<sup>14</sup> or 1.1 Mya<sup>9</sup>). Among the domestic populations, IRAN was separated from others about 4.45 thousand years ago (95% CI: 0.07–17.6 Kya) and then the Central and East Asian populations were separated about 2.40 thousand years ago (95% CI: 0.01–7.84 Kya; Fig. 4).

To allow for gene flow, we also tried to introduce migration bands from dromedaries to Bactrian camel populations (Supplementary Fig. 16 and Supplementary Table 12). The estimates could only converge when a migration band from Iranian dromedaries to IRAN and a migration band from a ghost population to KAZA were introduced (Supplementary Fig. 18 and Supplementary Table 13). Although the divergence time between wild and domestic Bactrian camels was not changed with the migration model (0.46 Mya, 95% CI: 0.24–0.71 Mya), the first divergence time of the domestic populations (0.19 Mya, 95% CI: 0.08–0.31 Mya) became unrealistic, because it was far beyond the known history of livestock domestication (11.5 Kya<sup>44</sup>). Besides, the total migration rate was only estimated as 0.27% and 0.16% for the migration band to IRAN and KAZA, respectively, much lower than that estimated with Admixture (1–10%). A possible reason for the poor estimation would be a more complex admixture history than the continuous migration model with constant rates assumed by G-PhoCS.

## Discussion

In this study, we characterized for the first time the whole-genome variations of camels across Asia, including domestic Bactrian camels representing a major subset of recognized breeds, extant wild Bactrian camels as well as dromedaries. As the extant wild Bactrian camels are going towards extinction, our research provided extremely valuable genetic resources of the living fossil. Also, considering the extensive utilization of domestic camels in transportation, milk and wool production, our data provided new options to implement genetic association study and marker-assistant selection for improving livestock productivity and future breeding effects. In addition, these data provided an unprecedented opportunity to trace the origin and migration of domestic Bactrian camels in history.

Previous studies found limited archaeological records and molecular markers for the first domestication of Bactrian camels in Central Asia rather than in East Asia<sup>17</sup>. Here we provided more solid evidences on the basis of the whole-genome sequences. The earliest branching among the domestic Bactrian camels occurred between IRAN and all the others, which was followed by the split between the Central and East Asian populations. Although evident introgression of dromedaries was observed in Central Asia, we demonstrated that it will not influence our results by removing the introgressed genomic segments. In contrast, although the extant wild and domestic Bactrian camels share close habitat in MG, our whole-genome analyses gave a coherent result to other mtDNA analyses that the two populations were separated by so long a time that the latter were not likely to originate from the former<sup>9,14</sup>. Furthermore, the extant wild camels contributed little to the gene pool of domestic populations, implying that the domestic populations in MG could possibly be immigrated there during more recent periods.

Based on these results, we proposed a comprehensive scenario for the origin and migration of the Bactrian camels (Supplementary Fig. 19). After the ancestor of camels moved from North America and split into dromedaries and Bactrian camels, the wild Bactrian camels spread from East to Central Asia about 0.43 million years ago (95% CI: 0.13–0.73 Mya). The Bactrian camels were first domesticated in Central Asia before 4.45 thousand years ago (95% CI: 0.07–17.6 Kya), which were then migrated

back to East Asia around 2.40 thousand years ago (95% CI: 0.01–7.84 Kya) with the increasing economic exchange and cooperation between the West and East. This scenario could resolve the mystery why the wild and domestic Bactrian camels from the Mongolian Plateau have such a large genetic distance. It should be noted that the timing of the events here were based on the model without admixture. Considering that the domestic Bactrian camels in Central Asia were further hybridized with dromedaries out of Arabia, the origin and migration age of the domestic Bactrian camels would be overestimated because of the increased genomic divergence.

Despite the insights gleaned from our data, it was important to note that the direct wild progenitor of domestic Bactrian camels were not found in Central Asia now and may no longer exist. However, there were records suggesting that the wild Bactrian camels were more widely distributed throughout Asia in history, extending from the great bend of the Yellow River westward to central Kazakhstan<sup>7</sup>. In future work, sequencing of ancient genomes from camel fossils will add to the picture of their early domestication. Another issue in our study was that although the occurrence of gene flow between dromedaries and Bactrian camels in Central Asia was convincingly detected, the actual admixture history remained largely unknown. First, the TreeMix analysis suggested that although the Iranian Bactrian camels and dromedaries were directly mixed, those from KAZA and RUS appeared to be mixed with a ghost population (Fig. 2c). Second, when the excessive shared alleles between Iranian dromedaries and Central Asian Bactrian camels were removed, KAZA continued to have reduced divergence from the dromedaries compared with the other populations (population tree in Fig. 3b). This branching pattern was also consistent with the ghost admixture model for KAZA. Third, a continuous migration model with constant rates implemented by G-PhoCS could only capture a small fraction of admixture, even though a ghost population was assumed. These results hinted at a more complex and possible multistage admixture history between Bactrian camels and dromedaries. As we only had a few dromedaries from IRAN, a future attempt to collect dromedaries from more diverse populations could help to decipher the admixture history.

## Methods

**Sample preparation.** Blood samples of 105 domestic Bactrian camels were collected from villages in China (55), MG (28), KAZA (6), RUS (10), and IRAN (6). Blood samples of four dromedaries were also collected from IRAN. The collections were made during routine veterinary treatments with the guidelines from the Camel Protection Association of Inner Mongolia. An endeavor was made to collect samples from unrelated individuals based on the information provided by the owners and local farmers. We collected 50 ml blood for each camel from the jugular vein after disinfection treatment, placed it in EDTA anticoagulant tubes, and then stored it at  $-80^{\circ}\text{C}$ . Ear skin samples (0.5 cm) of 19 wild Bactrian camels were collected from the Great Gobi-Strictly Protected Area A in MG. The wild Bactrian camels chosen were artificially reared and the research was reviewed and approved by the Great Gobi National Park. Proper surgical procedures were adopted in the collection. Local anesthesia (5% procaine hydrochloride) was applied to the ear and the wound was disinfected with iodophor and sulfonamide powder. The samples were eluted with phosphate-buffered saline solutions, placed in cryotubes and were stored at  $-80^{\circ}\text{C}$ .

**Genome sequencing.** The genomic DNA was extracted from 200  $\mu\text{l}$  blood samples with the QIAamp DNA Blood Mini kit (Qiagen) and from the skin samples with a standard phenol–chloroform method. The quality and integrity of DNA was controlled by OD260/280 ratio and agarose gel electrophoresis. For sequencing library preparation, the genomic DNA was sheared to fragments of 300–500 bp, which were then end repaired, 'A'-tailed, and ligated to Illumina sequencing adapters. The ligated products with sizes of 370–470 bp were selected on 2% agarose gels and then amplified by PCR. The libraries were sequenced on Illumina HiSeq platform with standard paired-end mode.

**Variant calling.** We used an in-house script to perform quality control on raw sequencing reads. Low-quality reads with ambiguous bases >10% were excluded.

The 3'-ends with base quality score <20 were trimmed and reads with length <35 bp were removed after trimming. Trimmed reads were mapped to the reference genome assembly of the Bactrian camel ([ftp://ftp.ncbi.nih.gov/genomes/Camelus\\_ferus/CHR\\_Un/cfe\\_ref\\_CB1\\_chrUn.fa.gz](ftp://ftp.ncbi.nih.gov/genomes/Camelus_ferus/CHR_Un/cfe_ref_CB1_chrUn.fa.gz)) using BWA-MEM (v0.7.12)<sup>45</sup> for each individual and then processed with SAMtools (v1.3.1)<sup>46</sup>. We followed the GATK pipeline (v3.2-2)<sup>47</sup> for variant calling. First, PCR duplicates were removed using Picard tools (v1.135) and local indel realignment were performed. Second, SNPs and small indels were called with UnifiedGenotyper across all 128 individuals simultaneously. Finally, the raw variants were filtered with the following criteria: variant quality score >40, sequencing depth summing all individuals >200 and <5000, minor allele frequency >1%, variants with <20% individuals with missing genotypes, root mean square of mapping quality >30, and biallelic variants. Total number of SNPs were reduced from 17.76 to 13.83 million after filtering. Functional annotation of variants were performed with ANNOVAR (v2013-06-21)<sup>48</sup> according to RefSeq ([ftp://ftp.ncbi.nih.gov/genomes/Camelus\\_ferus/GFF/ref\\_CB1\\_scaffolds.gff3.gz](ftp://ftp.ncbi.nih.gov/genomes/Camelus_ferus/GFF/ref_CB1_scaffolds.gff3.gz)).

**Population statistics and structure.** Summary population statistics, including pairwise nucleotide diversity  $\pi$ , Watterson's  $\theta$ , and Weir's  $F_{st}$  across 10 kb-sliding windows were calculated by VCFtools (v0.1.12b)<sup>49</sup>. Pairwise kinships between the samples were inferred by KING (v2.1.3)<sup>50</sup> and one of the paired individuals with close relationship was removed. For population structure analyses, SNPs in approximate linkage disequilibrium with each other were pruned by PLINK (v1.07)<sup>51</sup> (-indep-pairwise 50 5 0.5). SNPs located within exons and flanking 1 kb regions were also excluded. As a result, 2.08 million SNPs were preserved. MDS and pairwise distance matrix based on IBS were calculated using the -mds-plot 4 and -distance-matrix option in PLINK, respectively. The distance matrix was used to construct the NJ tree by Phylip (v3.69)<sup>52</sup>. One hundred random datasets were generated with -thin 0.1 option in PLINK and bootstrap values were retrieved from the consensus tree reconstructed by Phylip. The population ancestry was inferred by ADMIXTURE (v1.3.0)<sup>53</sup> with a fast maximum likelihood method. The optimum number of ancestral clusters  $K$  was estimated with the fivefold cross-validation procedure.

**TreeMix analysis and admixture tests.** Migration events among camel populations were inferred using TreeMix (v1.12)<sup>37</sup> with migration number  $m = 0-5$ . The threepop/fourpop module from the TreeMix package was used to perform the F3/F4 test<sup>38,53</sup> with -k 500. In the F3 test (Z; X, Y), one focal population (Z) was tested as a mixture of population X and Y. A large negative value of F3 score (standardized to Z-score with the Jackknife procedure) would indicate a very strong signal of Z as a mixture of X and Y. In our analysis, we ran F3 tests with all configurations of the populations. In the more sensitive F4 test (Y, Z; W, X), where W is an outgroup of Y and Z, the admixture bias between Y and Z with X was tested. If Y (or Z) have more admixture with X, it will show significant negative (or positive) F4 score (standardized to Z-score with the Jackknife procedure). To focus on the admixture between the domestic Bactrian camels and dromedaries, we set the population configuration as (Y, Z; wild, drom), where Y and Z were two domestic populations.

**Local introgression test.** To select the local genomic regions with significant introgression between dromedaries and Bactrian camels after their divergence, we used an in-house script to perform the BABA/ABBA test<sup>39</sup> across 100 kb-sliding windows. For the tree configuration (Y, Z; W, X), the original Patterson's  $D$  statistic can be calculated as a normalized F4 score<sup>53</sup>:

$$D = \frac{E((p_Y - p_Z)(p_W - p_X))}{E((p_Y + p_Z - 2p_Y p_Z)(p_W + p_X - 2p_W p_X))} \quad (1)$$

where  $p_X$  is the frequency of a given allele in population X and the expectations  $E()$  are estimated by averaging all SNPs in a window. The more robust  $f_d$  statistic for local genomic regions, which is a special form of F4 ratio and directly measures the proportion of introgression<sup>40</sup>, can be formulated as:

$$f_d = \begin{cases} \frac{E((p_Y - p_Z)(p_W - p_X))}{E(\max((p_X - p_Y)(p_X - p_W), (p_Z - p_Y)(p_Z - p_W)))} & (D > 0) \\ \frac{E((p_Y - p_Z)(p_W - p_X))}{E(\max((p_X - p_Z)(p_X - p_W), (p_Y - p_Z)(p_Y - p_W)))} & (D < 0) \end{cases} \quad (2)$$

We used the population configuration (East Asian, Central Asian; wild, drom) to perform the test. The  $f_d$  statistic in each window was evaluated by the Z-score with the Jackknife procedure:

$$Z(f_d) = \frac{E(\hat{f}_d)}{\sqrt{\text{var}(\hat{f}_d) \times n}} \quad (3)$$

where  $\hat{f}_d$  is estimated with a 10 kb block removed each time and  $n$  is the repetition times.

**Population phylogeny and mtDNA analysis.** Population distance was measured with average  $F_{st}$  across 10 kb windows. To minimize linkage disequilibrium and

perform the bootstrap test, five thousand 10 kb windows located at least 100 kb apart were randomly sampled for 100 times. The NJ tree and consensus tree were reconstructed by Phylip<sup>52</sup>. We compiled the full-length mtDNA sequences of camels from those we sequenced in the study and those collected from GenBank. The sequences were aligned using ClustalW<sup>54</sup>. The control regions were deleted, because they were missing in many sequences and not well aligned. The maximum likelihood tree was constructed using MEGA (v6.06)<sup>55</sup> with 1000 random bootstrap runs. The Tamura-Nei model and uniform substitution rates among sites were adopted. The ancestral area inference was performed with the Bayesian binary MCMC method implemented in RASP (v4.0)<sup>56</sup>. The MCMC was run for 50,000 iterations with 100 iterations between two samples and the first 100 samples were discarded.

**G-PhoCS analysis.** To prepare for the G-PhoCS (v1.3)<sup>42</sup> input, we implemented the following filters to the genome to reduce the effects of selection and sequencing errors: exons and 1 kb flanking regions; gap regions in the genome assembly; and regions with repeat sequence annotations. Altogether, 47% of the genome were excluded. We then randomly collected ten thousand 1 kb loci located at least 30 kb apart, to ensure sufficient inter-locus recombination. Multiple sequence alignments for the loci from individual genomes per population were retrieved by vcf-consensus in VCFtools<sup>49</sup>, with heterozygous genotypes represented by the International Union of Pure and Applied Chemistry code and missing genotypes masked by "N." Recommended gamma priors were used in the G-PhoCS analysis for the mutation-scaled population size  $\theta$ , population divergence time  $\tau$ , as well as migration rate  $m$ . The MCMC was run for 100,000 burn-in iterations and 500,000 sampling iterations with 10 iterations between two traced samples. The automatic fine-tuning procedure was done during the first 10,000 iterations. The convergence and mixing of the MCMC trace were monitored by Tracer (v1.6, available from <http://tree.bio.ed.ac.uk/software/tracer/>). Because of the stochastic nature of the MCMC algorithm, we tested the models on independent datasets and accepted the results only if two independent runs achieved similar estimates. We explored four models as follows: no migration (model 1); a single migration band from the dromedary to IRAN (model 2); two migration bands from the dromedary to IRAN and KAZA, respectively (model 3); and a migration band from the dromedary to IRAN and another from a ghost population to KAZA (model 4). All loci of the ghost population were set as "N." Only the models 1 and 4 showed convergence within ten independent runs. The time scale in years was calibrated according to a consensus divergence time of Bactrian camels and dromedaries (5.73 Mya in TimeTree<sup>43</sup>). The total migration rate per band  $M$  was calculated with  $M = m\tau_{mm}$ , where  $m$  was the mutation-scaled migration rate per generation and  $\tau_{mm}$  was the mutation-scaled time span of the migration band.

**Statistics and reproducibility.** Standard statistical tests were performed with R (v3.4.2). Specifically, the count of population-specific variants was compared between East Asia ( $n = 4$ ) and Central Asia ( $n = 3$ ) with the two-tailed  $t$ -test. The nucleotide diversity  $\pi$  was compared between the populations with the two-tailed  $t$ -test based on twenty thousand 10 kb windows separated by 100 kb with each other. The ANOVA of missing count of variants was performed for domestic Bactrian camels ( $n = 105$ ), wild Bactrian camels ( $n = 19$ ), and dromedaries ( $n = 4$ ), with sequencing depth as a covariant.

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

## Data availability

The raw data generated from this study have been submitted to the NCBI Sequence Read Archive (<http://www.ncbi.nlm.nih.gov/sra/>) under accession number SRP107089. The raw data are also available from NODE (<http://www.biosino.org/node/>) under accession number OEP000024. The datasets to reproduce the main figures have been submitted to Dryad (<https://doi.org/10.5061/dryad.tx95x69sz>).

Received: 16 July 2019; Accepted: 10 December 2019;

Published online: 07 January 2020

## References

1. Yam, B. A. Z. & Khomeiri, M. Introduction to camel origin, history, raising, characteristics, and wool, hair and skin: a review. *Int. J. Res. Innov. Earth Sci.* **2**, 177–187 (2015).
2. Burger, P. A. The history of Old World camelids in the light of molecular genetics. *Tropical Anim. Health Prod.* **48**, 905–913 (2016).
3. Köhler-Rollefson, I. U. *Camelus dromedarius*. *Mamm. Species* **375**, 1–8 (1991).
4. Hare, J. *Camelus ferus*. *The IUCN Red List of Threatened Species 2008* <https://doi.org/10.2305/IUCN.UK.2008.RLTS.T63543A12689285.en> (2008).
5. Hare, J. The wild Bactrian camel *Camelus bactrianus ferus* in China: the need for urgent action. *Oryx* **31**, 45–48 (1997).

6. Reading, R. P., Mix, H., Lhagvasuren, B. & Blumer, E. S. Status of wild Bactrian camels and other large ungulates in south-western Mongolia. *Oryx* **33**, 247–255 (1999).
7. Schaller, G. B. in *Wildlife of the Tibetan Steppes* 151–162 (Univ. Chicago Press, 2000).
8. Webb, S. D. in *Pleistocene Mammals of Florida* (ed. David W. S.) 170–214 (Univ. Press of Florida, 1974).
9. Mohandesan, E. et al. Mitogenome sequencing in the genus *Camelus* reveals evidence for purifying selection and long-term divergence between wild and domestic Bactrian camels. *Sci. Rep.* **7**, 9970 (2017).
10. Wu, H. et al. Camelid genomes reveal evolution and adaptation to desert environments. *Nat. Commun.* **5**, 5188 (2014).
11. Loog, L. et al. Estimating mobility using sparse data: application to human genetic variation. *Proc. Natl Acad. Sci. USA* **114**, 12213–12218 (2017).
12. Potts, D. Bactrian camels and bactrian-dromedary hybrids. *Silk Road* **3**, 50–58 (2005).
13. Almathen, F. et al. Ancient and modern DNA reveal dynamics of domestication and cross-continental dispersal of the dromedary. *Proc. Natl Acad. Sci. USA* **113**, 6707–6712 (2016).
14. Ji, R. et al. Monophyletic origin of domestic bactrian camel (*Camelus bactrianus*) and its evolutionary relationship with the extant wild camel (*Camelus bactrianus ferus*). *Anim. Genet.* **40**, 377–382 (2009).
15. Felkel, S. et al. A first Y-chromosomal haplotype network to investigate male-driven population dynamics in domestic and wild Bactrian camels. *Front. Genet.* **10**, 423 (2019).
16. Compagnoni, B. & Tosi, M. in *Approaches to Faunal Analysis in the Middle East* (eds Richard H. Meadow & Melinda A. Zeder) 119–128 (Peabody Museum Press, 1978).
17. Trinks, A., Burger, P., Benecke, N. & Burger, J. B. in *Camels in Asia and North Africa: Interdisciplinary Perspectives on their Past and Present Significance* (eds Eva-Maria Knoll & Pamela Burger) 79–86 (Austrian Academy of Sciences Press, 2012).
18. Veeramah, K. R. & Hammer, M. F. The impact of whole-genome sequencing on the reconstruction of human population history. *Nat. Rev. Genet.* **15**, 149–162 (2014).
19. Schraiber, J. G. & Akey, J. M. Methods and models for unravelling human evolutionary history. *Nat. Rev. Genet.* **16**, 727–740 (2015).
20. Nielsen, R. et al. Tracing the peopling of the world through genomics. *Nature* **541**, 302–310 (2017).
21. Wang, G. D., Xie, H. B., Peng, M. S., Irwin, D. & Zhang, Y. P. Domestication genomics: evidence from animals. *Annu. Rev. Anim. Biosci.* **2**, 65–84 (2014).
22. Fan, Z. et al. Worldwide patterns of genomic variation and admixture in gray wolves. *Genome Res.* **26**, 163–173 (2016).
23. Frantz, L. A. et al. Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science* **352**, 1228–1231 (2016).
24. Wang, G.-D. et al. Out of southern East Asia: the natural history of domestic dogs across the world. *Cell Res.* **26**, 21–33 (2016).
25. Botigue, L. R. et al. Ancient European dog genomes reveal continuity since the Early Neolithic. *Nat. Commun.* **8**, 16082 (2017).
26. Li, M. et al. Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nat. Genet.* **45**, 1431–1438 (2013).
27. Ai, H. et al. Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nat. Genet.* **47**, 217–225 (2015).
28. Frantz, L. A. et al. Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. *Nat. Genet.* **47**, 1141–1148 (2015).
29. Jirimutu et al. Genome sequences of wild and domestic bactrian camels. *Nat. Commun.* **3**, 1202 (2012).
30. Faye, B. & Konuspayeva, G. in *Camels in Asia and North Africa: Interdisciplinary Perspectives on their Past and Present Significance* (eds Eva-Maria Knoll & Pamela Burger) 27–33 (Austrian Academy of Sciences Press, 2012).
31. Khalkhali-Evrigh, R., Hafezian, S. H., Hedayat-Evrigh, N., Farhadi, A. & Bakhtiarzadeh, M. R. Genetic variants analysis of three dromedary camels using whole genome sequencing data. *PLoS ONE* **13**, e0204028 <https://doi.org/10.1371/journal.pone.0204028> (2018).
32. Carneiro, M. et al. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science* **345**, 1074–1079 (2014).
33. Sarkissian, C. D. et al. Evolutionary genomics and conservation of the endangered Przewalski's horse. *Curr. Biol.* **25**, 2577–2583 (2015).
34. Gunther, T. & Nettelblad, C. The presence and impact of reference bias on population genomic studies of prehistoric human populations. *PLoS Genet.* **15**, e1008302 (2019).
35. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
36. Silbermayr, K. & Burger, P. in *Camels in Asia and North Africa: Interdisciplinary Perspectives on their Past and Present Significance* (eds Eva-Maria Knoll & Pamela Burger) 69–76 (Austrian Academy of Sciences Press, 2012).
37. Pickrell, J. K. & Pritchard, J. K. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**, e1002967 (2012).
38. Peter, B. M. Admixture, population structure, and F-statistics. *Genetics* **202**, 1485–1501 (2016).
39. Durand, E. Y., Patterson, N., Reich, D. & Slatkin, M. Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* **28**, 2239–2252 (2011).
40. Martin, S. H., Davey, J. W. & Jiggins, C. D. Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Mol. Biol. Evol.* **32**, 244–257 (2015).
41. Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).
42. Gronau, I., Hubisz, M. J., Gulko, B., Danko, C. G. & Siepel, A. Bayesian inference of ancient human demography from individual genome sequences. *Nat. Genet.* **43**, 1031–1034 (2011).
43. Kumar, S., Stecher, G., Suleski, M. & Hedges, S. B. TimeTree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).
44. Larson, G. & Fuller, D. Q. The evolution of animal domestication. *Annu. Rev. Ecol. Evol. Syst.* **45**, 115–136 (2014).
45. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
46. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
47. DePristo, M. A. et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
48. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164–e164 (2010).
49. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
50. Manichaikul, A. et al. Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
51. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
52. Felsenstein, J. Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods Enzymol.* **266**, 418–427 (1996).
53. Patterson, N. et al. Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
54. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
55. Tamara, K., Stecher, G., Peterson, D., Filipowski, A. & Kumar, S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).
56. Ali, S. S., Yu, Y., Pfosser, M. & Wetschnig, W. Inferences of biogeographical histories within subfamily Hyacinthoideae using S-DIVA and Bayesian binary MCMC analysis implemented in RASP (Reconstruct Ancestral State in Phylogenies). *Ann. Bot. Lond.* **109**, 95–107 (2012).

## Acknowledgements

This work was supported by grants from the International S&T Cooperation Program of China (2015DFR30680 and ky201401002), the National Natural Science Foundation of China (31360397 and 31560710), the National Key R&D Program of China (2017YFA0505500, 2016YFC0901704, 2017YFC0907505, and 2017YFC0908405), the special project of the Inner Mongolia Autonomous Region, the Chinese Academy of Sciences (KFJ-STZ-QYZD-126 and ZDBS-SSW-DQC-02), and the Youth Innovation Promotion Association CAS (2017325). We thank Dr Feng Qiu and Mr Ze Xu from BasePair Biotechnology Co., Ltd, for technical assistance on NGS. We thank Anhui Engineering Laboratory for Big Data of Precision Medicine, Suzhou, for providing computational resources.

## Author contributions

Y.L., Z.W., and J. designed research. S.H., G.C., T.J., N.H.-E., M.B., G.K.B., T.G.-E., B.T., W.Z., A.Z., H., E., A.N., P.M., N., G.M., C.N., O.K., Sirendalai, Sarentuya, and A. collected samples. L.M., L. Yi, G.D., J.H., and L.H. performed sequencing. L.M., L. Yuan, L. Yi, W.L., and Z.W. analyzed data. L.M., L. Yuan, Z.W., and J. wrote the manuscript.

## Competing interests

The authors declare no competing interests.



**Additional information**

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s42003-019-0734-6>.

**Correspondence** and requests for materials should be addressed to Y.L., Z.W. or Jirimutu.

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

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



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

© The Author(s) 2020

Liang Ming<sup>1,2,25</sup>, Liyun Yuan<sup>3,4,25</sup>, Li Yi<sup>1,25</sup>, Guohui Ding<sup>3,5</sup>, Surong Hasi<sup>6</sup>, Gangliang Chen<sup>7</sup>, Tuyatsetseg Jambal<sup>8</sup>, Nemat Hedayat-Evrigh<sup>9</sup>, Mijiddorj Batmunkh<sup>10,11</sup>, Garyaeva Khongr Badmaevna<sup>12</sup>, Tudeviin Gan-Erdene<sup>13</sup>, Batsukh Ts<sup>8</sup>, Wenbin Zhang<sup>14</sup>, Azhati Zulipikaer<sup>15</sup>, Hosblig<sup>16</sup>, Erdemt<sup>17</sup>, Arkady Natyrov<sup>18</sup>, Prmanshayev Mamay<sup>19</sup>, Narenbatu<sup>20</sup>, Gendalai Meng<sup>21</sup>, Choijilsuren Narangerel<sup>22</sup>, Orgodol Khongorzul<sup>1</sup>, Jing He<sup>1</sup>, Le Hai<sup>1</sup>, Weili Lin<sup>3</sup>, Sirendalai<sup>2</sup>, Sarentuya<sup>2</sup>, Aiyisi<sup>8</sup>, Yixue Li<sup>3,4,23,24\*</sup>, Zhen Wang<sup>10</sup><sup>3\*</sup> & Jirimutu<sup>1,2,10\*</sup>

<sup>1</sup>Key Laboratory of Dairy Biotechnology and Engineering, Ministry of Education, College of Food Science and Engineering, Inner Mongolia Agricultural University, Huhhot, China. <sup>2</sup>Inner Mongolia Institute of Camel Research, West Alax, Inner Mongolia, China. <sup>3</sup>Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institute of Nutrition and Health, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China. <sup>4</sup>Bio-Med Big Data Center, CAS-MPG Partner Institute for Computational Biology, Shanghai Institute of Nutrition and Health, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China. <sup>5</sup>Gui'an Bio-Med Big Data Center, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Guiyang, China. <sup>6</sup>Key Laboratory of Clinical Diagnosis and Treatment Technology in Animal Disease, Ministry of Agriculture, College of Veterinary Medicine, Inner Mongolia Agricultural University, Huhhot, China. <sup>7</sup>Bactrian Camel Academe of Altai, Xingjiang Wangyuan Camel Milk Limited Company, Fuhai County, Xijiang, China. <sup>8</sup>College of Industrial Technology, Mongolian University of Science and Technology, Ulaanbaatar, Mongolia. <sup>9</sup>University of Mohaghegh Arabili, Ardabil, Iran. <sup>10</sup>China-Mongolia Joint Laboratory for Biomacromolecule Research, Ulaanbaatar, Mongolia. <sup>11</sup>Mongolian Wild Camel Protection Area, Ministry of Nature and Environment, Ulaanbaatar, Mongolia. <sup>12</sup>Kirovski Plant, Non-Public Joint-Stock Company, Republic of Kalmykia, Russia. <sup>13</sup>Institute of Chemistry and Chemical Technology, Mongolian Academy of Sciences, Ulaanbaatar, Mongolia. <sup>14</sup>Bactrian Camel Institute of Alxa, Inner Mongolia, China. <sup>15</sup>Animal Science Institute, Xinjiang Academy of Animal Science, Urumqi, China. <sup>16</sup>Animal Husbandry Bureau of North Urad, Bayannuur, Inner Mongolia, China. <sup>17</sup>Animal Husbandry Workstation of West Sunid, Xiliingol, Inner Mongolia, China. <sup>18</sup>Agrarian Faculty, Kalmyk State University, Republic of Kalmykia, Russia. <sup>19</sup>Kazakh National Agrarian University, Almaty, Kazakhstan. <sup>20</sup>College of Animal Science, Inner Mongolia Agricultural University, Huhhot, China. <sup>21</sup>Department of Pharmacy, Affiliated Hospital of Inner Mongolia Medical University, Huhhot, China. <sup>22</sup>Institute of Technology, Ulaanbaatar, Mongolia. <sup>23</sup>Shanghai Center for Bioinformation Technology, Shanghai Industrial Technology Institute, Shanghai, China. <sup>24</sup>Collaborative Innovation Center for Genetics and Development, Fudan University, Shanghai, China. <sup>25</sup>These authors contributed equally: Liang Ming, Liyun Yuan, and Li Yi. \*email: [yxli@sibs.ac.cn](mailto:yxli@sibs.ac.cn); [zwang01@sibs.ac.cn](mailto:zwang01@sibs.ac.cn); [yeluotuo1999@vip.163.com](mailto:yeluotuo1999@vip.163.com)