

# Whole-genome sequencing of 128 camels across Asia provides insights into origin and migration of domestic Bactrian camels

Liang Ming<sup>1,2#</sup>, Liyun Yuan<sup>3,4#</sup>, Li Yi<sup>1#</sup>, Guohui Ding<sup>3</sup>, Surong Hasi<sup>5</sup>, Gangliang Chen<sup>6</sup>, Tuyatsetseg Jambl<sup>7</sup>, Nemat Hedayat-Evright<sup>8</sup>, Mijiddorj Batmunkh<sup>9,10</sup>, Garyaeva Khongr Badmaevna<sup>11</sup>, Tudeviin Gan-Erdene<sup>12</sup>, Batskh TS<sup>7</sup>, Wenbin Zhang<sup>13</sup>, Azhati Zulipikaer<sup>14</sup>, Hosblig<sup>15</sup>, Erdemt<sup>16</sup>, Arkady Natyrov<sup>17</sup>, Prmanshayev Mamay<sup>18</sup>, Narenbatu<sup>19</sup>, Gendalai Meng<sup>20</sup>, Choijilsuren Narangerel<sup>21</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>7</sup>, Yixue Li<sup>3,4,22,23\*</sup>, Zhen Wang<sup>3\*</sup>, Jirimutu<sup>1,2,9\*</sup>

1. Key Laboratory of Dairy Biotechnology and Engineering, Ministry of Education; College of Food Science and Engineering, Inner Mongolia Agricultural University, Huhhot, China
2. Inner Mongolia Institute of Camel Research, West Alax, Inner Mongolia, China
3. 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
4. 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
5. Key Laboratory of Clinical Diagnosis and Treatment Technology in Animal Disease, Ministry of Agriculture; College of Veterinary Medicine, Inner Mongolia Agricultural University, Huhhot, China
6. Bactrian Camel Academe of Altai, Xingjiang Wangyuan Camel Milk Limited Company, Fuhai County, Xijiang, China
7. College of Industrial Technology, Mongolian University of Science and Technology, Ulaanbaater, Mongolia

8. University of Mohaghegh Arabili, Ardabil, Iran
9. China-Mongolia Joint Laboratory for Biomacromolecule Research, Ulaanbaatar, Mongolia
10. Mongolian Wild Camel Protection Area, Ministry of Nature and Environment, Ulaanbaatar, Mongolia
11. Non-public joint-stock company, Kirovski plant, Republic of Kalmykia, Russia
12. Institute of Chemistry and Chemical Technology, Mongolian Academy of Sciences, Ulaanbaatar, Mongolia
13. Bactrian Camel Institute of Alxa, Inner Mongolia, China
14. Animal Science Institute, Xinjiang Academy of Animal Science, Urumqi, China
15. Animal Husbandry Bureau of North Urad, Bayannuur, Inner Mongolia, China
16. Animal Husbandry Workstation of West Sunid, Xiliingol, Inner Mongolia, China
17. Ministry of Education and Science of Russian Federation Kalmyk State University Agrarian Faculty, Republic of Kalmykia Russia
18. Kazakh National Agrarian University, Almaty, Kazakhstan
19. College of Animal Science, Inner Mongolia Agricultural University, Huhhot, China
20. Department of Pharmacy, Affiliated Hospital of Inner Mongolia Medical University, Huhhot, China
21. Institute of Technology, Ulaanbaatar, Mongolia
22. Shanghai Center for Bioinformation Technology, Shanghai Industrial Technology Institute, Shanghai, China
23. Collaborative Innovation Center for Genetics and Development, Fudan University, Shanghai, China

#These authors contributed equally to this work.

\*Correspondence should be addressed to Yixue Li (yxli@sibs.ac.cn), Zhen Wang (zwang01@sibs.ac.cn) or Jirimutu (yeluotuo1999@vip.163.com).

## ABSTRACT

The domestic Bactrian camels were treated as the principal means of locomotion between the eastern and western cultures in history. To address the question of their origin, we performed whole-genome sequencing of 128 camels across Asia, including representative populations of domestic Bactrian camels from the Mongolian Plateau to the Caspian Sea, as well as the extant wild Bactrian camels and dromedaries. The domestic and wild Bactrian camels showed remarkable genetic divergence since they were split from dromedaries, confirming they were separated species. The wild Bactrian camels made also little contribution to the ancestry of domestic ones. Among the domestic Bactrian camels, those from Iran exhibited the largest genetic distance from others, and were the first population to separate in the phylogeny. Although evident admixture was observed between domestic Bactrian camels and dromedaries living around the Caspian Sea, the large genetic distance and basal position of Iranian Bactrian camels could not be explained by introgression alone. Taken together, our study favored the Iranian origin of domestic Bactrian camels, which were then immigrated eastward to Mongolia where the native wild Bactrian camels inhabited. This study illustrated the complex genomic landscape of migration underlying domestication in Bactrian camels.

## INTRODUCTION

Camels (*Camelus*, *Camelini*) contain two extant domestic species, the one-humped dromedary (*Camelus dromedarius*) and the two-humped Bactrian camel (*Camelus bactrianus*) [1, 2]. While the former herds are mainly feed in North Africa and West Asia, the latter herds wildly live in the cold desert areas of Northeast and Middle 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 [3], is listed as critically endangered by the International Union for Conservation of Nature [4] with an estimation of a few hundreds to 2,000 individuals [5, 6]. Historically, the wild Bactrian camel was widely distributed throughout Asia, extending from the great bend of the Yellow River westward to central Kazakhstan, but it can only be found in the Mongolian Gobi and the Chinese Taklimakan and Lop Noor deserts today [7]. 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 [8, 9]. Within the Camelini, the dromedaries and Bactrian camels were then split around 4 or 5 million years ago [9, 10]. The domestication of camels, like many other domestic mammals, has promoted unprecedented progress in cultural and economic development of human societies, representing 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.

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 [11]. However, the origin of domestic Bactrian camels is still a mystery. One intuitive possibility was the extant wild Bactrian camels were the progenitor of the domestic form, which were then dispersed from the Mongolian Plateau to west gradually [7, 12]. This hypothesis was supported by the presence of Camelid faunal remains at Neolithic sites near Mongolia, although it was unclear these were the domestic as opposed to the wild ones [12]. Despite the fact that the wild Bactrian camel has certain characters distinct from the domestic one, such as the lower, pyramid-shaped humps, the thinner, lithe legs, and the smaller and more slender

body [7], it is still difficult to tell them apart based on their morphological features alone. Nevertheless, molecular studies based on mtDNAs discovered dramatic sequence variations between the wild camels and domestic Bactrian camels, and proposed that the extant wild Bactrian camel was a separate lineage and not the direct progenitor of the domestic Bactrian camel [9, 13]. Another possible place of origin was Iran [1], where early skeletal remains of domestic Bactrian camels (around 2,500-3,000 BC) were discovered [14]. Although prehistoric mtDNAs of Bactrian camels supported the idea that the domestication took place in the southwestern Middle Asia rather than in Mongolia or East Asia [15], 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 [16-18] and domestic animals [19] such as dogs [20-23] and pigs [24-26]. The genome assembly of Bactrian camels were published recently [10, 27], which provided a new opportunity to examine the evolutionary relationship between the extant wild and domestic Bactrian camels and trace their origin. In this study, we performed whole-genome sequencing of 128 camels including both domestic and extant wild Bactrian camels from their typical habitats. We included dromedaries as well because they were not only the outgroup of Bactrian camels in phylogeny, but also had a long history of hybridization with Bactrian camels in breeding practice [2, 28], especially in Middle Asia. Our results supported the Iranian-origin hypothesis 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 Mongolia as well as 4 dromedaries from Iran were gathered for this study (Supplementary Figure S1 and Table S1). The domestic Bactrian camels were chosen to cover as many major

geographic regions as possible, including 55 from Inner Mongolia (IMG), Xinjiang (XJ) and Qinghai (QH) of China, 28 from Mongolia (MG), 6 from Kazakhstan (KAZA), 10 from Russia (RUS) and 6 from Iran (IRAN). As a variety of local breeds were formed due to the wide utilization of domestic Bactrian camels in China and Mongolia, 8 different representative breeds were chosen from the regions. The other domestic Bactrian camels from Middle Asia were living around the Caspian Sea.

After DNA extraction, individual genomes were sequenced to an average of  $13\times$  coverage (Supplementary Figure S2 and Table S2). The sequence reads were aligned to our previous genome assembly of the Bactrian camel [27] for variant calling (Supplementary Figure S3). After stringent filtering, we totally identified 15.76 million SNPs and 1.97 million small indels (Supplementary Table S3). Functional annotation of the variants indicated that about 68.75% of them were intergenic, 30.18% were intronic and 0.77% were exonic (Supplementary Table S4 and Table S5). Among the variants, 15.49 million, 5.96 million and 10.31 million were identified in the domestic Bactrian camels, wild Bactrian camels and dromedaries, respectively (Supplementary Table S3). Although dromedaries were more divergent from both of the Bactrian camel species in phylogeny, the domestic Bactrian camels shared more variants with the dromedaries (52.69%) than with the wild Bactrian camel (31.14%) (Supplementary Figure S4) due to the tremendous reduction of genetic variants observed in the extant wild Bactrian camel.

### ***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 S6). The pairwise nucleotide diversity  $\pi$  (Figure 1A) of dromedaries ( $1.65\times 10^{-3}$ ) was significantly higher than that of Bactrian camels from all geographic regions ( $0.92\times 10^{-3}$ - $1.14\times 10^{-3}$ , Supplementary Table S7), which was in contrast to previous heterozygosity estimates based on individual genomes [10]. One important reason could be the hybridization practice between dromedaries and Bactrian camels in Middle Asia [28]. Among the Bactrian camels,

the wild population showed the lowest  $\pi$  ( $0.92 \times 10^{-3}$ ) compared with all of the domestic populations (Figure 1A and Supplementary Table S7). Although this result violated many cases that wild animals usually have higher genetic diversity than the domestic breeds derived from them such as dogs [22], pigs [25] and rabbits [29], it would happen due to the fact that the extant wild Bactrian camels are critically endangered with an extremely small population size [4]. In addition, the domestic populations living around the Caspian Sea generally showed a higher diversity than those living in the Mongolia Plateau (Figure 1A). The tendency was also confirmed by the Watterson's  $\theta$  (Supplementary Figure S5). Again, the hybridization with dromedaries in Middle Asia could account for the higher diversity in the region.

We then measured the pairwise genetic distance between the camel populations by Weir's  $F_{st}$  (Figure 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 (0.53-0.63), and the wild Bactrian camels had the second highest  $F_{st}$  with the domestic ones (0.27-0.30). The differentiation among the domestic Bactrian camels was much lower, implying a single-origin of them. 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 Figure S6). The NJ tree also supported IRAN as the first population separated from all the other domestic Bactrian camels, implying the possibility of Iranian origin of domestic Bactrian camels.

### ***Population structure with admixture***

To reveal the overall population structure with potential admixture, we pruned the SNPs by removing those with high linkage disequilibrium, extremely low minor allele frequency and potential functional effects. The multidimensional scaling (MDS) analysis based on the pruned subset reproduced the similar result as the full set (Figure 2A and Supplementary Figure S7). 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 (Figure 2A).

To estimate different ancestral proportions, we performed population structure analysis with Admixture [30] by assuming  $K$  ancestral populations (Figure 2B). The cross-validation procedure supported that  $K = 3$  was optimal (Supplementary Figure S8), 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 Caspian Bactrian camel populations including IRAN, KAZA and RUS, with a proportion estimated as 1%-10%. Moreover, we observed introgression of domestic Bactrian camels into several wild individuals with a proportion of 6%-15%. As the wild camels have become quite rare, it was proposed that the introgression could threaten the genetic distinctiveness of the wild species [31]. On the contrary, 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 ( $K = 5$ , Figure 2B).

As another method to examine the population structure with admixture, we inferred the population tree for the camels using TreeMix [32] (Figure 2C). When no migration track ( $m = 0$ ) was added, the tree topology again indicated that IRAN was the first lineage separated among all the domestic Bactrian camels. Increasing of migration tracks ( $m = 3$ ) could dramatically improve the fit of the model (Supplementary Figure S9), which identified gene flows from dromedaries to domestic Bactrian camels in Middle Asia, including KAZA, RUS and IRAN with migration weights ranging from 4% to 8% (Supplementary Table S8). It was worth mentioning that while 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 (Figure 2C). This could imply a ghost population related to the Iranian dromedary that contributed to the ancestry of KAZA and RUS. However, TreeMix detected no signal of migration between the wild and domestic Bactrian camels. We then used the less-parameterized three- and four-population ( $F_3/F_4$ ) test [33] to evaluate the statistical



significance of the admixture events. Again, the F3 test strongly supported the admixture of dromedaries and Bactrian camels in KAZA, RUS and IRAN (Supplementary Table S9), but it did not support the admixture between the wild and domestic Bactrian camels. The more sensitive F4 test confirmed a significantly higher extent of admixture between dromedaries and Bactrian camels in Middle Asia compared with those in Northeast Asia (Supplementary Table S10). Among the latter, a higher extent of admixture with dromedaries was detected in XJ than in MG/IMG.

### ***Evidence for Iranian origin by removing introgression***

Mongolia and Iran were the two alternative regions of domestication for Bactrian camels based on archaeological evidences [1, 12, 15], 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 Middle Asian would weaken the support for origin inference. To reduce this confounder effect, we attempted to remove the introgressed segments of dromedaries from the Bactrian camel genomes by using the ‘BABA/ABBA’ test [34]. We grouped the Mongolian and Caspian domestic populations, and compared allele sharing between the two groups with dromedaries (Figure 3A). We used the statistic  $f_d$ , a robust version of the Patterson’s D to locate introgressed segments [35], and applied a strict significance level of Z-score = 2 by using the Jackknife procedure (Supplementary Figure S10). In a total of 21,188 non-overlapping 100-kb segments across the genome, there were far more segments with significant signals of introgression in the Caspian populations (11,651,  $Z > 2$ ) than in the Mongolian populations (3,778,  $Z < -2$ ) as expected. We then performed the Admixture analysis based on the remaining segments and confirmed that the introgression of dromedaries were effectively removed (Supplementary Figure S11). 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 (Figure 3B). To gain more insights into the population phylogeny, we re-constructed the NJ tree based on the pairwise  $F_{st}$  and performed the bootstrap test (Figure 3B). It was shown that IRAN was the first one to separate among all the domestic Bactrian populations, followed by KAZA and RUS. These results implied the Iranian origin of domestic Bactrian camels

and a dispersal route from Middle Asia to Mongolia.

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 (Figure 3C and Supplementary Table S11). 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 lineages of the domestic populations (Figure 3C). Moreover, after removing the introgressed mtDNAs, IRAN displayed the highest nucleotide diversity (Supplementary Table S12). These mtDNA results were in agreement with the nuclear genome and again implicated the Iranian origin of domestic Bactrian camels.

### ***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 [10], the long-term trajectories of Bactrian camels based on the pairwise sequentially Markovian coalescent (PSMC) model [36] revealed a tremendous decrease in the effective population size of the ancestral camels earlier than one million years ago (Supplementary Figure S12). 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 0.5 million years ago. The whole-genome divergence was in agreement with previous mtDNA studies [9, 13], which proposed that the domestic Bactrian camels and the extant wild ones were separated so long that the former could not be directly derived from the latter.

In order to explore the divergence time among the camel populations in more detail, we used the generalized phylogenetic coalescent sampler (G-PhoCS) [37]. 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

(Supplementary Figure S13 and Supplementary Table S13). The time was calibrated by assuming the Bactrian camel and dromedary divergence of 5.73 million years in the Timetree database [38]. When no migration band was incorporated, convergence of all parameter estimates could easily be achieved (Supplementary Figure S14 and Supplementary Table S14). Similar to the PSMC results, the effective population size was generally decreased from ancestral to modern populations (Figure 4). The divergence time between wild and domestic Bactrian camels was estimated as 0.49 million years ago (95% confidence interval [CI]: 0.35-0.65 Mya, Figure 4), which was slightly later than that based on mtDNAs (0.7 Mya) [13]. Among the domestic populations, IRAN was separated from others about 10.2 thousand years ago (95% CI: 0.25-22.5 Kya), and then the Caspian and Mongolian populations were separated about 7.39 thousand years ago (95% CI: 0.01-18.5 Kya, Figure 4). To allow for gene flow, we also tried to introduce migration bands from dromedaries to Bactrian camel populations (Supplementary Figure S13 and Supplementary Table S13). The estimates could only converge when a migration band from Iranian dromedaries to IRAN, as well as a migration band from a ghost population to KAZA were introduced (Supplementary Figure S15 and Supplementary Table S14). The total migration rate was estimated as 1% for both migration bands, though the divergence time between Bactrian camel populations was indistinguishable with the migration model.

## 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 Middle Asia rather than in East Asia [15]. Here, we provided more evidences on the basis of the whole-genome sequences to support that hypothesis, particularly for the domestication in the area of Iran. We found that IRAN exhibited the largest genetic distance from all other domestic populations, and it harbored a unique ancestry not present in other populations. Further phylogenetic analysis revealed that IRAN was also the first one to separate among the domestic populations, which was followed by the other Caspian and Mongolian populations. Although evident introgression of dromedaries was observed in the Caspian populations, 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 Mongolia, 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 [9, 13]. Furthermore, the extant wild camels contributed little to the gene pool of domestic populations, implying that the domestic populations in Mongolia 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 Figure S16). After the ancestor of camels moved from North America and split into dromedaries and Bactrian camels, the wild Bactrian camels spread from East to Middle Asia about 0.5 million years ago. The Bactrian camels were first domesticated in the area of Iran around 10 thousand years ago, which were then migrated back to East Asia with the increasing economic exchange and cooperation between West and East. In more recent periods, the domestic Bactrian camels around the Caspian Sea were further hybridized with domestic dromedaries out of Arabia. This scenario could well resolve the mystery why the wild and domestic Bactrian camels from the Mongolian Plateau have so large genetic distance. It was also similar to what happened in human evolution. Modern genomic evidence suggested that although many groups of ancient hominids occupied the continent in history, they were globally extinct and

substituted by only a small group of people out of Africa [18].

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 Iran now, which 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 [7]. In future work, sequencing of ancient genomes from camel fossils will add to the picture of their early domestication.

## **METHODS**

### ***Sample preparation***

Blood samples of 105 domestic Bactrian camels were collected from villages in China (55), Mongolia (28), Kazakhstan (6), Russia (10) and Iran (6). Blood samples of 4 dromedaries were also collected from Iran. The collection were made during routine veterinary treatments with approval 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 °C. Skin samples from the ear (0.5 cm) of 19 wild Bactrian camels were collected from the Great Gobi-Strictly Protected Area “A” in Mongolia. 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 PBS buffered solutions, placed in cryotubes and stored at -80 °C.

### ***Genome sequencing***

The genomic DNA was extracted from 200 µl 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 marked by Illumina sequencers in the FASTQ files were removed. 3'-ends with base quality score < 20 were trimmed, and reads with length < 35-bp were removed after trimming. Trimmed reads were

then 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) [39] for each individual with default parameters. We followed the GATK pipeline (v3.2-2) [40] for variant calling. Firstly, PCR duplicates were removed using Picard tools (v1.135), and local indel realignment were performed. Secondly, SNPs and small indels were called across all 128 individuals simultaneously. Finally, the raw variants were filtered with the following criteria: 1) variant quality score > 40; 2) average sequencing depth of individuals 1-40; 3) individuals with missing genotypes < 20. Total number of SNPs were reduced from 17.76 to 15.76 million after filtering. Functional annotation of variants were performed with ANNOVAR (v2013-06-21) [41] 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) [42]. Pairwise kinships between the samples were inferred by KING (v2.1.3) [43], and one of the paired individuals with close relationship was removed. The following SNPs were further removed by PLINK (v1.07) [44] for population structure analyses: 1) multiallelic sites; 2) SNPs in approximate linkage disequilibrium with each other (--indep 50 5 0.2); 3) SNPs with extremely low minor allele frequency (--maf 0.005); 4) SNPs located within exons and flanking 1-kb regions. A total of 2.65 million SNPs were preserved. Multidimensional scaling 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 neighbour-joining tree by Phylip (v3.69) [45]. 100 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) [30] with a fast maximum likelihood method. The optimum number of ancestral clusters  $K$  was estimated with the 5-fold cross-validation procedure.

### ***TreeMix analysis and admixture tests***

Migration events among camel populations were inferred using TreeMix (v1.12) [32] with migration number  $m = 0-5$ . The threepop/fourpop module from the TreeMix package was used to perform the F3/F4 test [33, 46] 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, drome), 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 [34] 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 [46]:

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

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 [35], 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 (Mongolian, Caspian; wild, drome) to performed 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.

### ***Mitochondrial DNA analysis***

We compiled the full-length mtDNA sequences of camels from both we sequenced in the study and those collected from GenBank. The sequences were aligned using ClustalW [47]. 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) [48] with 1000 random bootstrap runs. The Tamura-Nei model and uniform substitution rates among sites were adopted. The nucleotide diversity of mtDNAs were calculated by DNASP (v5.1.0.1) [49].

### ***G-PhoCS analysis***

To prepare for the G-PhoCS (v1.3) [37] input, we implemented the following filters to the genome to reduce the effects of selection and sequencing errors: 1) exons and 1-kb flanking regions; 2) gap regions in the genome assembly; 3) regions with repeat sequence annotations. Altogether, 47% of the genome were excluded. We then randomly collected 10,000 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 [42], with heterozygous genotypes represented by the IUPAC 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 the stochastic nature of the MCMC algorithm, we tested the models on independent datasets, and accepted the results only if two independent runs achieved the similar estimates. Simple model without migration had a high chance to converge, but

none of the runs with migration bands directly from dromedaries to IRAN and/or KAZA showed convergence. This could be improved by incorporating a ghost population closely related to the dromedary and with migration band to KAZA. All loci of the ghost population were set as ‘N’. The time scale in years was calibrated according to a consensus divergence time of Bactrian camels and dromedaries, which was 5.73 million years ago in TimeTree [38]. The total migration rate across a given migration band was calculated with  $M = m\tau_m$ , where  $\tau_m$  was the time span of the migration band.

### ***Data availability***

The 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 data are also available from NODE (<http://www.biosino.org/node/>) under accession number OEP000024.

### **ACKNOWLEDGMENTS**

This work was supported by grants from the International S&T Cooperation Program of China (2015DFR30680, ky201401002), the National Natural Science Foundation of China (31360397, 31560710), the National Key R&D Program of China (2017YFA0505500, 2016YFC0901704), the special project of the Inner Mongolia Autonomous Region, 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.

### **AUTHOR CONTRIBUTIONS**

Y.L., Z.W. and J. designed the study. 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., S., S. and A. collected samples. L.M., L.Y., G.D., J.H. and L.H. performed sequencing experiments. L.M., L.Y.Y., L.Y., W.L. and Z.W. analyzed the data. L.M., L.Y.Y., Z.W. and J. wrote the manuscript.

### **COMPETING FINANCIAL INTERESTS**

The authors declare no competing financial interests.

## FIGURE LEGENED

**Figure 1. Genetic diversity and differentiation of the camel populations.** (A) Nucleotide diversity  $\pi$ . The boxplot shows  $\pi$  for 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. (B) Pairwise population differentiation  $F_{st}$ . The heatmap represents average  $F_{st}$  for 10-kb sliding windows. The dendrogram represents hierarchical clustering of the populations based on  $F_{st}$ .

**Figure 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.

**Figure 3. Identification of the origin of domestic Bactrian camels by removing introgression.** (A) BABA/ABBA analysis for introgression of dromedaries into domestic Bactrian camels. Number of 100-kb segments with significant  $f_d$  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 10-kb sliding windows. Bootstrap values of the NJ tree were calculated by re-sampling the 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.

**Figure 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 indicates IRAN-MG and KAZA-MG divergence, respectively.

## Reference

1. Yam BAZ, Khomeiri M: Introduction to Camel origin, history, raising, characteristics, and wool, hair and skin: A Review. *International Journal of Research and Innovations in Earth Science* 2015, 2(6):177-187.
2. Burger PA: The history of Old World camelids in the light of molecular genetics. *Tropical animal health and production* 2016, 48(5):905-913.
3. Köhler-Rollefson IU: Camelus dromedarius. *Mammalian Species* 1991, 375:1-8.
4. *Camelus ferus*. The IUCN Red List of Threatened Species 2008 [<http://dx.doi.org/10.2305/IUCN.UK.2008.RLTS.T63543A12689285.en>]
5. Hare J: The wild Bactrian camel Camelus bactrianus ferus in China: The need for urgent action. *Oryx* 1997, 31(1):45-48.
6. Reading RP, Mix H, Lhagvasuren B, Blumer ES: Status of wild Bactrian camels and other large ungulates in south-western Mongolia. *Oryx* 1999, 33(3):247-255.
7. Schaller GB: Wild Bactrian Camel. In: *Wildlife of the Tibetan Steppes*. Chicago and London: University of Chicago Press; 2000: 151-162.
8. Webb SD: Pleistocene Llamas of Florida, with a brief review of the Lamini. In: *Pleistocene mammals of Florida*. Edited by Webb SD. Gainesville, FL: University Press of Florida; 1974: 170-214.
9. Mohandesan E, Fitak RR, Corander J, Yadamsuren A, Chuluunbat B, Abdelhadi O, Raziq A, Nagy P, Stalder G, Walzer C *et al*: Mitogenome Sequencing in the Genus Camelus Reveals Evidence for Purifying Selection and Long-term Divergence between Wild and Domestic Bactrian Camels. *Scientific reports* 2017, 7(1):9970.
10. Wu H, Guang X, Al-Fageeh MB, Cao J, Pan S, Zhou H, Zhang L, Abutarboush MH, Xing Y, Xie Z *et al*: Camelid genomes reveal evolution and adaptation to desert environments. *Nature communications* 2014, 5:5188.
11. Almathen F, Charruau P, Mohandesan E, Mwacharo JM, Orozco-terWengel P, Pitt D, Abdussamad AM, Uerpmann M, Uerpmann H-P, De Cupere B: Ancient and modern DNA reveal dynamics of domestication and cross-continental dispersal of the dromedary. *Proceedings of the National Academy of Sciences* 2016, 113(24):6707-6712.
12. Potts D: Bactrian camels and bactrian-dromedary hybrids. *The Silk Road* 2005, 3(1):50-58.
13. Ji R, Cui P, Ding F, Geng J, Gao H, Zhang H, Yu J, Hu S, Meng H: Monophyletic origin of domestic bactrian camel (*Camelus bactrianus*) and its evolutionary relationship with the extant wild camel (*Camelus bactrianus ferus*). *Animal genetics* 2009, 40(4):377-382.
14. Compagnoni B, Tosi M: The camel: Its distribution and state of domestication in the Middle East during the third millennium B.C. in light of the finds from Shahr-i Sokhta. In: *Approaches to Faunal Analysis in the Middle East*. Edited by Meadow RH, Zeder MA. New Haven: Peabody Museum Press; 1978: 119-128.
15. Trinks A, Burger P, Benecke N, Burger JB: Ancient DNA Reveals Domestication Process: The Case of the Two-Humped Camel. In: *Camels in Asia and North Africa: Interdisciplinary Perspectives on their Past and Present Significance*. Edited by Knoll E-M, Burger P. Wien (Austria): Austrian Academy of Sciences Press; 2012: 79-86.
16. Veeramah KR, Hammer MF: The impact of whole-genome sequencing on the reconstruction of human population history. *Nature reviews Genetics* 2014, 15(3):149-162.
17. Schraiber JG, Akey JM: Methods and models for unravelling human evolutionary history. *Nature*

- reviews Genetics* 2015, 16(12):727-740.
18. Nielsen R, Akey JM, Jakobsson M, Pritchard JK, Tishkoff S, Willerslev E: Tracing the peopling of the world through genomics. *Nature* 2017, 541(7637):302-310.
  19. Wang GD, Xie HB, Peng MS, Irwin D, Zhang YP: Domestication genomics: evidence from animals. *Annual review of animal biosciences* 2014, 2:65-84.
  20. Fan Z, Silva P, Gronau I, Wang S, Armero AS, Schweizer RM, Ramirez O, Pollinger J, Galaverni M, Ortega Del-Vecchyo D *et al*: Worldwide patterns of genomic variation and admixture in gray wolves. *Genome research* 2016, 26(2):163-173.
  21. Frantz LA, Mullin VE, Pionnier-Capitan M, Lebrasseur O, Ollivier M, Perri A, Linderholm A, Mattiangeli V, Teasdale MD, Dimopoulos EA *et al*: Genomic and archaeological evidence suggest a dual origin of domestic dogs. *Science* 2016, 352(6290):1228-1231.
  22. Wang G-D, Zhai W, Yang H-C, Wang L, Zhong L, Liu Y-H, Fan R-X, Yin T-T, Zhu C-L, Poyarkov AD: Out of southern East Asia: the natural history of domestic dogs across the world. *Cell research* 2016, 26(1):21-33.
  23. Botigue LR, Song S, Scheu A, Gopalan S, Pendleton AL, Oetjens M, Taravella AM, Seregely T, Zeeb-Lanz A, Arbogast RM *et al*: Ancient European dog genomes reveal continuity since the Early Neolithic. *Nature communications* 2017, 8:16082.
  24. Li M, Tian S, Jin L, Zhou G, Li Y, Zhang Y, Wang T, Yeung CK, Chen L, Ma J *et al*: Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nature genetics* 2013, 45(12):1431-1438.
  25. Ai H, Fang X, Yang B, Huang Z, Chen H, Mao L, Zhang F, Zhang L, Cui L, He W *et al*: Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. *Nature genetics* 2015, 47(3):217-225.
  26. Frantz LA, Schraiber JG, Madsen O, Megens HJ, Cagan A, Bosse M, Paudel Y, Crooijmans RP, Larson G, Groenen MA: Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. *Nature genetics* 2015, 47(10):1141-1148.
  27. Jirimutu, Wang Z, Ding G, Chen G, Sun Y, Sun Z, Zhang H, Wang L, Hasi S, Zhang Y *et al*: Genome sequences of wild and domestic bactrian camels. *Nature communications* 2012, 3:1202.
  28. Faye B, Konuspayeva G: The Encounter between Bactrian and Dromedary Camels in Central Asia. In: *Camels in Asia and North Africa: Interdisciplinary Perspectives on their Past and Present Significance*. Edited by Knoll E-M, Burger P. Wien (Austria): Austrian Academy of Sciences Press; 2012: 27-33.
  29. Carneiro M, Rubin CJ, Di Palma F, Albert FW, Alfoldi J, Martinez Barrio A, Pielberg G, Rafati N, Sayyab S, Turner-Maier J *et al*: Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science* 2014, 345(6200):1074-1079.
  30. Alexander DH, Novembre J, Lange K: Fast model-based estimation of ancestry in unrelated individuals. *Genome research* 2009, 19(9):1655-1664.
  31. Silbermayr K, Burger P: Hybridization: A Threat to the Genetic Distinctiveness of the Last Wild Old World Camel Species. In: *Camels in Asia and North Africa: Interdisciplinary Perspectives on their Past and Present Significance*. Edited by Knoll E-M, Burger P. Wien (Austria): Austrian Academy of Sciences Press; 2012: 69-76.
  32. Pickrell JK, Pritchard JK: Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS genetics* 2012, 8(11):e1002967.

33. Peter BM: Admixture, Population Structure, and F-Statistics. *Genetics* 2016, 202(4):1485-1501.
34. Durand EY, Patterson N, Reich D, Slatkin M: Testing for ancient admixture between closely related populations. *Molecular biology and evolution* 2011, 28(8):2239-2252.
35. Martin SH, Davey JW, Jiggins CD: Evaluating the use of ABBA-BABA statistics to locate introgressed loci. *Molecular biology and evolution* 2015, 32(1):244-257.
36. Li H, Durbin R: Inference of human population history from individual whole-genome sequences. *Nature* 2011, 475(7357):493-496.
37. Gronau I, Hubisz MJ, Gulko B, Danko CG, Siepel A: Bayesian inference of ancient human demography from individual genome sequences. *Nature genetics* 2011, 43(10):1031-1034.
38. Kumar S, Stecher G, Suleski M, Hedges SB: TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Molecular biology and evolution* 2017, 34(7):1812-1819.
39. Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009, 25(14):1754-1760.
40. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, Del Angel G, Rivas MA, Hanna M: A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature genetics* 2011, 43(5):491-498.
41. Wang K, Li M, Hakonarson H: ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research* 2010, 38(16):e164-e164.
42. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST *et al*: The variant call format and VCFtools. *Bioinformatics* 2011, 27(15):2156-2158.
43. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM: Robust relationship inference in genome-wide association studies. *Bioinformatics* 2010, 26(22):2867-2873.
44. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ: PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* 2007, 81(3):559-575.
45. Felsenstein J: Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. *Methods in enzymology* 1996, 266:418-427.
46. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, Reich D: Ancient admixture in human history. *Genetics* 2012, 192(3):1065-1093.
47. Thompson JD, Higgins DG, Gibson TJ: CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic acids research* 1994, 22(22):4673-4680.
48. Tamura K, Stecher G, Peterson D, Filipinski A, Kumar S: MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular biology and evolution* 2013, 30(12):2725-2729.
49. Librado P, Rozas J: DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 2009, 25(11):1451-1452.









