

ORIGINAL RESEARCH

Population genetic analysis of the domestic Bactrian camel in China by RAD-seq

Chenmiao Liu | Huiling Chen | Zhanjun Ren  | Chengdong Zhang | Xuejiao Yang

College of Animal Science and Technology, Northwest A&F University, Yangling, China

Correspondence

Zhanjun Ren, College of Animal Science and Technology, Northwest A&F University, Yangling, Shaanxi 712100, China.
Email: Renzhanjun@nwsuaf.edu.cn

Funding information

National Natural Science Foundation of China, Grant/Award Number: 31172178

Abstract

Restriction site-associated DNA sequencing (RAD-seq) is one of the most effective high-throughput sequencing technologies for SNP development and utilization and has been applied to studying the origin and evolution of various species. The domestic Bactrian camels play an important role in economic trade and cultural construction. They are precious species resources and indispensable animals in China's agricultural production. Recently, the rapid development of modern transportation and agriculture, and the deterioration of the environment have led to a sharp decline in the number of camels. Although there have been some reports on the evolution history of the domestic Bactrian camel in China, the origin, evolutionary relationship, and genetic diversity of the camels are unclear due to the limitations of sample size and sequencing technology. Therefore, 47 samples of seven domestic Bactrian camel species from four regions (Inner Mongolia, Gansu, Qinghai, and Xinjiang) were prepared for RAD-seq analysis to study the evolutionary relationship and genetic diversity. In addition, seven domestic Bactrian camel species are located in different ecological zones, forming different characteristics and having potential development value. A total of 6,487,849 SNPs were genotyped. On the one hand, the filtered SNP information was used to conduct polymorphism mapping construction, LD attenuation analysis, and nucleotide diversity analysis. The results showed that the number of SNPs in Dongjiang camel was the highest, the LD coefficient decayed the fastest, and the nucleotide diversity was the highest. It indicates that Dongjiang camel has the highest genetic diversity. On the other hand, the filtered SNPs information was used to construct the phylogenetic tree, and F_{ST} analysis, inbreeding coefficient analysis, principal component analysis, and population structure analysis were carried out. The results showed that Nanjiang camel and Beijing camels grouped together, and the other five Bactrian camel populations gathered into another branch. It may be because the mountains in the northern part of Xinjiang and the desert in the middle isolate the two groups from the other five groups.

KEYWORDS

Bactrian camel, genetic relationship, RAD sequencing, single nucleotide polymorphisms

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *Ecology and Evolution* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

The domestic Bactrian camel has been historically used as a transportation animal in the desert area of northwestern China and played an important role in ancient trade between China and Europe. Thus, the camel is very important for studying the livestock production in ancient northern China, and economic and cultural exchange between China and the West. With the rapid development of modern agriculture and transportation tools, the number of camels in China has dropped from 373,000 in 1994 to 250,000 today. Therefore, it is necessary to study the genetic diversity of Bactrian camels to clarify their phylogenetic relationship and lay the foundation for the protection of their biodiversity (Chen, Ren, Zhao, Zhang, & Yang, 2018). The genome of camel covers genetic information, which could be used to analyze the genetic relationship among different camels (He, Han, & Ma, 2009). The genetic diversity of six Bactrian camel populations in China was analyzed by microsatellite. It was found that Nanjiang camel and Beijiing camel gathered together, and the four groups of Alashan camel, Qinghai camel, Subei camel, and Sunite camel gathered together. Studies had shown that the genetic diversity of the domestic Bactrian camel in China was abundant and there was a certain gene flow among populations (Bai, Zhou, Li, & Feng, 2017). He (2011) studied the origin and domestication of nine domesticated Bactrian camel populations in China and one Mongolian domesticated Bactrian camel population by partial sequencing of mitochondrial D-loop. Phylogenetic median-joining network analysis showed that the domestic Bactrian camel populations came from a branch. However, the use of RAD-seq for the development and utilization of genomic SNPs has not been reported in the genetic resources of Chinese Bactrian camel populations.

Owing to their high density and relatively uniform distribution throughout the genome, single nucleotide polymorphisms (SNPs) are ideal molecular markers for genetic mapping and population diversity assessment (Zhai et al., 2014). Currently, next-generation sequencing (NGS) technologies have been used in phylogenetic analyses to quickly collect a wide range of molecular markers for elucidating evolutionary relationships (Zhou et al., 2018). The RAD-seq has been proved to be economical and effective for understanding

the evolutionary history within and between closely related species (Kang, Ma, & He, 2017). Jones, Fan, Franchini, Schartl, and Meyer (2013) used RAD-Seq technology to analyze the phylogenetic relationship of 26 species from the Central American genus *Xiphophorus*, using a large number of SNP markers in the genome to infer the evolutionary relationship within the genus, improving the previous defects from the aspects of sequence and morphology. Zhai et al. (2014) used RAD-seq technology in chickens for SNP discovery and genotyping, and found that RAD-seq was a powerful tool for genotyping and discovering high-density genetic markers. In addition, RAD-seq can be used for deeper genetic analyses and applications, such as genome-wide association studies (GWAS), marker-assisted selection (MAS), and genome-wide selection of other organisms.

In this study, RAD-seq was used to analyze 47 samples of seven domestic Bactrian camel breeds from four regions (Inner Mongolia, Gansu, Qinghai, and Xinjiang) and to identify the SNPs of each sample. The study is beneficial to protect the genetic diversity and exploit the potential value of Bactrian camels. Camels are now more than a means of transportation. With the need of domestic animal products (camel wool, camel milk), especially the anticancer activity of camel milk (Krishnankutty et al., 2018), the genetic resources of Bactrian camels need to be redeveloped. In addition, seven kinds of domestic Bactrian camels located in different ecological zones have different characteristics and potential ecological adaptation value.

2 | MATERIALS AND METHODS

2.1 | Sample collection and DNA extraction

In this study, 47 venous blood samples of 7 breeds of domestic Bactrian agricultural camel from Inner Mongolia, Gansu, Qinghai, and Xinjiang were collected (Table 1; Figure 1) to ensure that each sample came from a different family, and there was no relationship between individuals. The experiment was approved by the National Natural Science Foundation of China (31172178), and the relevant guidelines and regulations were followed. DNA samples were extracted using a standard phenol-chloroform method (Sambrook & Russell, 2002) and were diluted to 20 ng/ μ l.

TABLE 1 Sample information for domestic Bactrian camels in China

Number	Population	Code	Sampling site	Sample size
1	Nanjiang camel	NJ	Wensu County, Xinjiang	7
2	Beijiing camel	BJ	Qinghe County, Xinjiang	7
3	Dongjiang camel	DJ	Mulei County, Xinjiang	7
4	Hexi camel	HX	Yongchang County, Gansu	5
5	Qinghai camel	QH	Mohe, Qinghai	7
6	Alashan camel	ALS	Alashan Left Banner, Inner Mongolia	7
7	Sunite camel	SNT	Sunite Right Banner, Inner Mongolia	7
Total				47



FIGURE 1 A geographical map of the sampling sites

2.2 | Library construction

RAD-seq libraries were constructed following a modified protocol (Baird et al., 2008). Briefly, genomic DNA (0.1–1 µg; from either individual or pooled samples) was digested with *EcoRI* (New England Biolabs) followed by heat inactivation of the enzyme. Individually barcoded P1 adapters were ligated onto the *EcoRI* cut site for each sample. Samples were pooled in proportionate amounts for shearing to average size of 500 bp. The library size was selected into 300- to 700-bp fragments by running on a 1% agarose gel. Libraries were blunt-end-repaired, and a 3'-adenine overhang was added to each fragment. We added a P2 adapter containing unique Illumina barcodes for each library. Libraries were amplified through 16 PCR cycles with Phusion high-fidelity DNA polymerase (New England Biolabs) and column-purified. Samples were sequenced on an Illumina HiSeq 3000 using 100 bp paired-end reads.

2.3 | Bioinformatics analysis

2.3.1 | Clean reads filtering

Quality trimming generated by the use of fastp is an essential step to generate high confidence of variant calling (Chen, Ren, et al., 2018; Chen, Zhou, Chen, & Gu, 2018). Raw reads would be processed to get high-quality clean reads according to three stringent filtering standards: (a) removing reads with $\geq 10\%$ unidentified nucleotides (N); (b) removing reads with $>50\%$ bases having phred quality scores of ≤ 20 ; and (c) removing reads aligned to the barcode adapter.

2.3.2 | SNP identification and annotation

To identify SNPs, the Burrows-Wheeler Aligner (BWA) was used to align the clean reads from each sample against the reference genome (<https://www.ncbi.nlm.nih.gov/genome/10741>) with the

settings “mem 4 -k 32 -M,” -k is the minimum seed length, and -M is an option used to mark shorter split alignment hits as secondary alignments (Li & Durbin, 2009). Variant calling was performed for all samples using the GATK's Unified Genotyper (Depristo et al., 2011). SNPs were filtered using the GATK's Variant Filtration with proper standards (-Window 4, -filter “QD < 2.0||FS > 60.0||MQ < 40.0”, -G_ filter “GQ < 20”).

2.4 | Statistical analysis

Statistically, comparing two of the multiple averages to each other is called multiple comparisons. There are many methods for multiple comparisons. In this study, multiple comparisons were made between SNPs of seven camel populations by using the least significant difference method (LSD method). The procedure of the least significant difference method is to calculate the least significant difference LSD_{α} with a significant level of α under the premise that the *F* test is significant between the treatments. We used SPSS version 18.0 (SPSS, Inc.) for statistical analyses. The difference between any two averages, such as its absolute value $\geq LSD_{\alpha}$, is significant at the level of α .

2.5 | Population genetic analysis

The population relationship based on individual sample strategies was used to analyze the differentiation, gene exchange, and evolutionary history of seven domestic camel populations. Polymorphism map construction, LD attenuation analysis, and nucleotide diversity analysis were used for camel population genetic analysis.

2.5.1 | Polymorphism map construction

For each group, in order to ensure the reliability of the population SNP, for the sites where the SNP exists, each individual in the group

needs to ensure that the site information exists, and missing phenomenon does not exist (Zhai et al., 2014). The vcftools software was used to calculate SNP statistics, the minimum read coverage for a SNP to be called is 3X, and all noncompletely missing polymorphic loci ($-\text{max-missing } 1\text{e-}06$ $-\text{non-ref-af } 1\text{e-}06$) were used for counting. A total of 6,487,849 SNPs were obtained. Based on the information of each SNP locus, the unique SNP map was constructed for Chinese camel breeds.

2.5.2 | LD attenuation analysis

Linkage disequilibrium (LD) is also called as allelic association. It refers to a nonrandom association between different locus alleles within a population. After sequencing, 6,487,849 SNPs were obtained. We estimate the LD decay trend by calculating the LD coefficient (r^2) between two points in a range of sequence (typically < 5 Mb). The LD coefficient r^2 represents the correlation of two points, which varies within the interval of 0–1. If r^2 is 0, there is no correlation between the inheritance of the two locus. If r^2 is 1, it means that the site correlation is the largest. We used the “LD attenuation distance” to evaluate the speed of LD decay (Guo, 2012).

2.5.3 | Nucleotide diversity analysis

As an important indicator to measure the level of genetic variation of species, nucleotide diversity has important scientific significance for studying the genetic diversity of the species, its evolutionary history, and its systematic location. Nucleotide diversity can be calculated in the PopGenome package (Pfeifer, Wittelsbürger, Ramos-Onsins, & Lercher, 2014). Nucleotide diversity analysis is a method for calculating the genetic diversity of individuals within a population by comparing the π value changes by using SNPs to calculate the average of the differences between any two nucleotide sequences in the population.

2.6 | Population genetic relationship analysis

In the study, phylogenetic tree analysis, principal component analysis (PCA), population structure analysis, etc., were used to analyze the evolutionary relationship among different groups on different dimensions (Wang et al., 2015).

2.6.1 | Inbreeding coefficient analysis

The inbreeding coefficient refers to the correlation coefficient between two gametes forming an individual, that is, the ratio of the pair of genes from the common ancestor in all relative genes of the individual. Inbreeding coefficient was calculated using Plink 1.9 (<http://zzz.bwh.harvard.edu/plink/>). After F value of each individual was obtained, the average was obtained within the population.

2.6.2 | F_{ST} statistical analysis

Due to manual selection, there are significant differences between breeds in different regions. After removing the site with a deletion rate of 50% or more, the remaining 865,774 loci were used for the calculation of F_{ST} . F_{ST} values can be calculated from the F_{ST} statistics function in the PopGenome package to study population genetic diversities among different breeds (Pfeifer et al., 2014). It can also help to study the genetic distance among different breeds and be used to express differences in population genetic structure.

2.6.3 | Phylogenetic tree analysis

The phylogenetic tree is a branch map or tree that describes the order of evolution between groups, indicating the evolutionary relationship between groups. After SNP detection, 6,487,849 SNPs can be used to calculate the distance between the populations. The phylogenetic tree was constructed by a neighbor-joining method in the software Treebest (version 1.9.2), and a total of 1,000 replicates generated the Bootstrap values. Moreover, the phylogenetic tree was constructed by the maximum likelihood (ML) method and default parameters were set. By analyzing the phylogenetic tree constructed by the two methods, we can find the evolutionary relationship between populations.

2.6.4 | Principal component analysis

In genetics, PCA mainly converts multiple differences into a few independent factors and analyzes the differences among different groups by the degree of difference among these factors to classify different populations (Gao et al., 2017). In this study, after removing the site with a deletion rate of 50% or more, the remaining 865,774 loci were used for the calculation of PCA. We used the

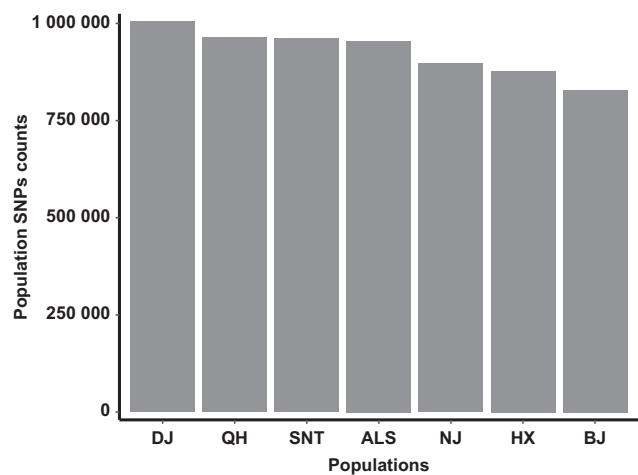


FIGURE 2 Population SNPs counts. The abscissa represents each population, and the ordinate represents number of SNPs per population. ALS, Alashan camels; BJ, Beijiang camels; DJ, Dongjiang camels; HX, Hexi camels; NJ, Nanjiang camels; QH, Qinghai camels; SNT, Sunite camels

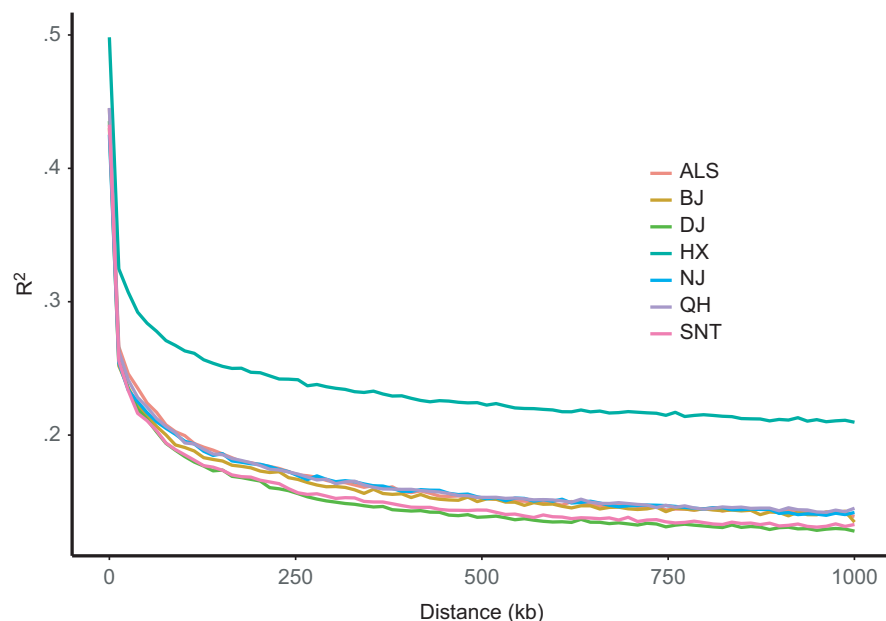


FIGURE 3 LD attenuation map of different camel populations. The abscissa represents the physical distance (bp), and the ordinate represents the LD coefficient (R^2)

software GCTA (Yang, Lee, Goddard, & Visscher, 2011) for PCA analysis based on the degree of SNP differences between individuals, obtained the principal component values of each sample, and then used the R language to draw PCA scatter plots to further cluster different breeds.

2.6.5 | Group structure analysis

Genetic structure was inferred using the program STRUCTURE (<http://web.stanford.edu/group/pritchardlab/structure.html>; to ensure the independence of the SNP marks, we used the PLINK software to filter 865,774 SNPs according to the LD intensity, the remaining 12,046 loci were used for structure analysis, -indep-pairwise 250 10 0.1, 250 kb window, the step size of 10 SNPs, r^2 is $<.1$). We predefined the number of genetic clusters from $K = 2$ to $K = 6$ (BURNIN = 5,000 times, NUMREPS = 100,000), and each K value was repeated three times. After the STRUCTURE software run was completed, we used the POPHELPER software (<https://www.ncbi.nlm.nih.gov/pubmed/26850166>) to calculate the ΔK value and the CLUMPP software (<https://www.ncbi.nlm.nih.gov/pubmed/17485429>) to merge the results of the three repetitions. After the combination, we used the POPHELPER software to illustrate it (Schraiber & Akey, 2015).

3 | RESULTS

3.1 | Sequencing basic data analysis

The ratio of high-quality clean reads was above 97.36%, and the number of reads on the alignment was mostly above 97% by RAD-seq. A total of 6,487,849 SNPs were detected. There were 2,270,747 loci (35%) in the transversion and 4,217,102 loci (65%) in the transition. The ratio of transition to transversion was nearly 2:1, the transition frequency between purine and pyrimidine was basically the same, and the transversion frequency of each type was also substantially flush.

3.2 | Population genetic analysis

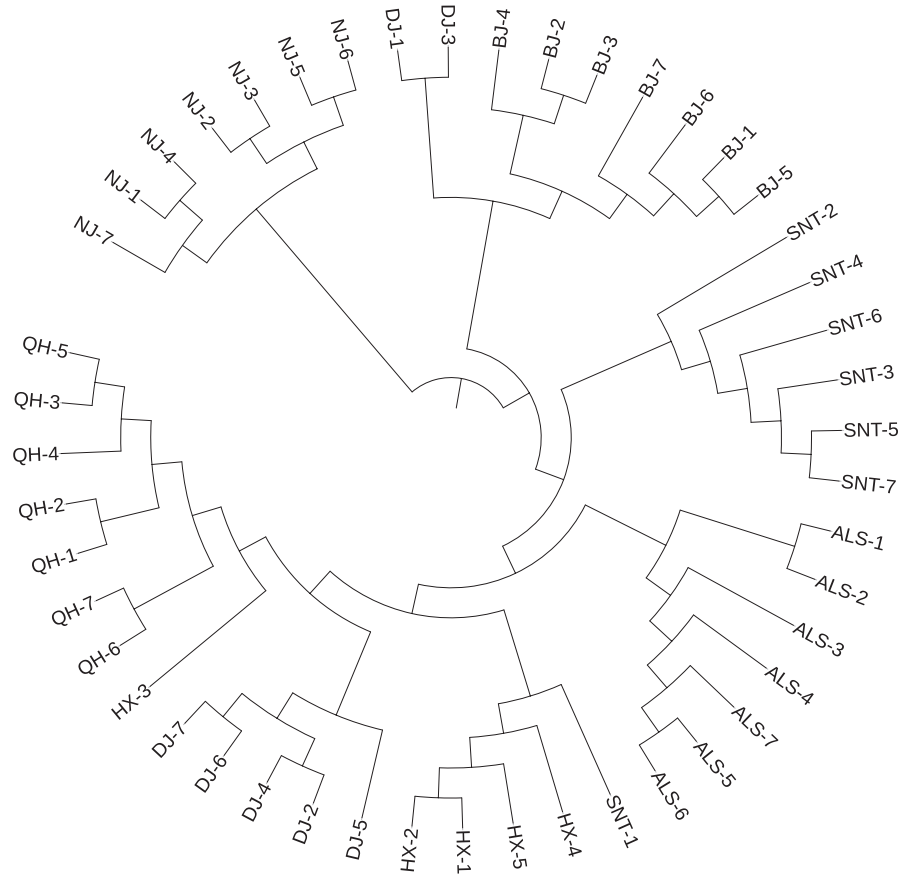
3.2.1 | Polymorphism map construction

The vcftools software was used to extract SNPs from each population, and there were differences in SNPs among different domestic camel populations. The number of SNPs in the seven domestic camel populations varied from 828,619 in the Beijiing camel to 1,004,857 in the Dongjiang camel, but the differences between the populations were not significant ($p > .05$; Figure 2). The differences of SNP in each population reflected the differences of polymorphism among different

Population	Code	Average_ F_{ST}	π	Average_ F_{IS}
Nanjiang camel	NJ	0.1971	0.0001516	0.4515
Beijiing camel	BJ	0.1991	0.0001368	0.4947
Dongjiang camel	DJ	0.1594	0.0001716	0.3959
Hexi camel	HX	0.2068	0.0001489	0.4193
Qinghai camel	QH	0.2020	0.0001573	0.3782
Alashan camel	ALS	0.1834	0.0001625	0.4361
Sunite camel	SNT	0.1742	0.0001627	0.4507

TABLE 2 F_{ST} , π , and F_{IS} values for domestic Bactrian camels in China

FIGURE 4 Phylogenetic tree construction by maximum likelihood method. ALS, Alashan camels; BJ, Beijiing camels; DJ, Dongjiang camels; HX, Hexi camels; NJ, Nanjiang camels; QH, Qinghai camels; SNT, Sunite camels



populations. The SNP map of the population was also constructed by the statistics of SNPs in different populations. It can be seen from the Figure 2 that the number of SNPs in Dongjiang camel was the highest, and the genetic diversity was highest.

3.2.2 | LD attenuation analysis

The rate of LD attenuation varied greatly among different species or among different subgroups of the same species. The faster the LD coefficient decays, the higher the genetic diversity of the population is. LD attenuation analysis was performed on seven domesticated camel populations in four regions. It can be observed from the figure that the decay rate of the LD coefficient was different in different populations, and the attenuation speed was Dongjiang camel > Sunite camel > Beijiing camel > Nanjiang camel > Qinghai camel > Alashan camel > Hexi camel (Figure 3).

3.2.3 | Nucleotide diversity analysis

Nucleotide diversity represents the genetic diversity of population to some extent. It can be seen from Table 2 that the Dongjiang camel had the highest nucleotide diversity, and the genetic diversity was highest. The nucleotide diversity of Hexi camel was low. It is consistent with the results of polymorphism map construction and LD attenuation analysis.

3.3 | Population genetic relationship analysis

3.3.1 | Inbreeding coefficient analysis

For an individual, the probability of getting a gene from the parent is 1/2, and the probability of getting a gene from the grandparent is 1/4; that is, the probability of getting the same gene from a common ancestor is reduced by 1/2 every other generation. Therefore, the closer the kinship is, the greater the inbreeding coefficient is. As can be seen from Table 2, the average value of F_{IS} between Hexi camel, Qinghai camel, and other populations was lower, indicating that the genetic distance was far from other populations.

3.3.2 | F_{ST} statistical analysis

The F_{ST} values between different populations were calculated to study the genetic diversity among different populations. It can be seen from Table 2 that the average F_{ST} value between Dongjiang camel and other camel breeds was 0.1594, which was the lowest among all the average values. It indicated that its genetic distance was close to other populations, and the genetic difference with other populations was not significant. The average value of F_{ST} between Hexi camel, Qinghai camel, and other populations was the highest, indicating that the genetic distance was far from other populations.

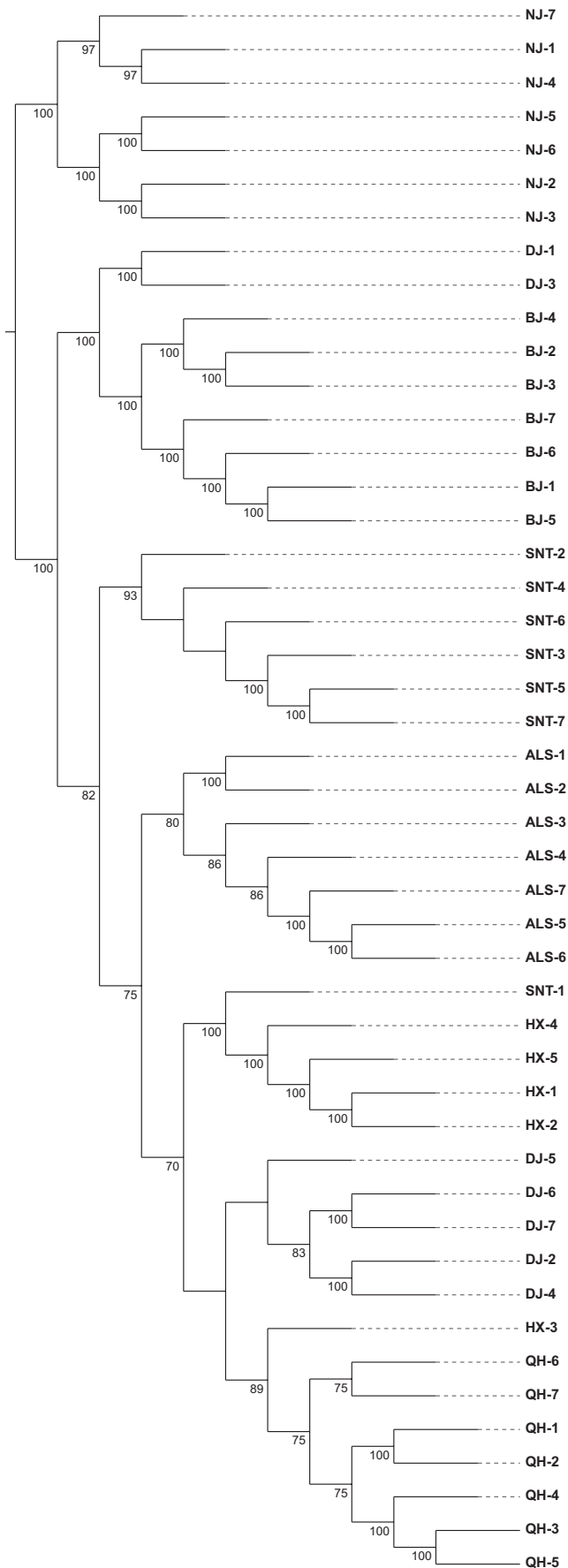


FIGURE 5 Phylogenetic tree construction of neighbor-joining (NJ) method. ALS, Alashan camels; BJ, Beijing camels; DJ, Dongjiang camels; HX, Hexi camels; NJ, Nanjiang camels; QH, Qinghai camels; SNT, Sunite camels

3.3.3 | Phylogenetic tree analysis

The filtered SNPs information was used to construct phylogenetic tree by the ML method (Figure 4) and the NJ method (Figure 5). The results showed that the seven breeds were clustered together, and the clusters were obvious, and the order of evolution was consistent. Two populations of Beijing camel and Nanjiang camel were closely related and can be regarded as a large branch. Five populations of Qinghai camel, Dongjiang camel, Hexi camel, Alashan camel, and Sunite camel can be regarded as another branch. Some individuals of Dongjiang camel and Beijing camel formed branches. It indicated that there may be gene communication between Beijing camel and Dongjiang camel.

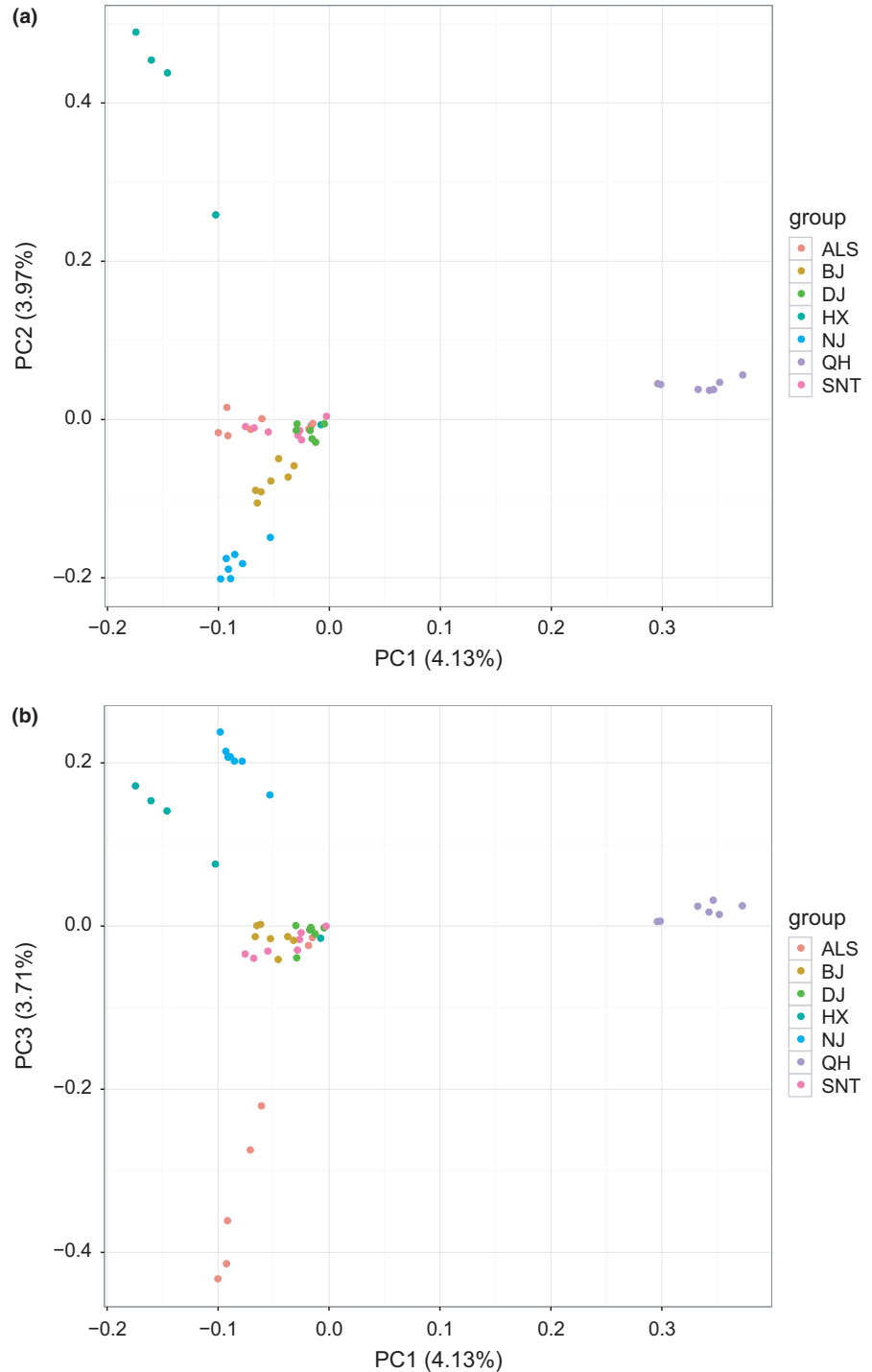
3.3.4 | Principal component analysis

The key information was extracted from millions of SNPs, and the PCA of seven breeds of domestic camels was conducted (Figure 6). On the PC1 and PC2 level, Qinghai camels were grouped together, and Hexi camels were grouped together and were far away from other groups. It was consistent with the F_{ST} statistical analysis. Two PCA maps showed that Sunite camel and Alashan camels were clustered together, and there was genetic communication between two camel breeds. They had a close genetic relationship, and it was consistent with the results of the phylogenetic tree. Dongjiang camels and Beijing camels gathered together, that showed genetic communication and relatively close genetic relationship. It was consistent with the results of the phylogenetic tree.

3.3.5 | Group structure analysis

The STRUCTURE analysis software was used to analyze the population structure of the domestic Bactrian camels and the ancestral population of each breed (Figure 7). When performing a population structure analysis for each K value, we performed cross-validation to determine which K value best fitted the true differentiation history of the population. In Figure 7, the best classification strategy for this group was $K = 5$. At $K = 3$, the Qinghai camel and the Hexi camel were clearly separated from other populations of camels, indicating that the Qinghai camel and the Hexi camel were far away from the camels of other populations, which was consistent with the results of the F_{ST} statistical analysis. At $K = 5$, the Alashan camel and the Sunite camel shared a common genetic ancestry, and there was genetic communication between two populations of camels, which was consistent with the results of the PCA. Some individuals of the Dongjiang camel shared an ancestral genetic background with the Beijing camel, which was consistent with the results of the evolutionary tree. The ancestral background of Nanjiang camel was very pure and had a genetic

FIGURE 6 Principal components analysis in domestic Bactrian camel populations (1) and (2)



ancestry. Dongjiang camel, Qinghai camel, and Sunite camel had a variety of genetic ancestry, but there was a main genetic ancestry.

4 | DISCUSSION

The genetic diversities and evolutionary origins of domestic Bactrian camels in China have been a fascinating source for decades (Chen, Ren, et al., 2018; Chen, Zhou, et al., 2018). We used RAD-seq technology to analyze the SNP loci of domestic Bactrian camels and

further carried out the population genetic research of domestic Bactrian camels. A total of 6,487,849 SNP loci were detected in all samples. Filtered SNPs were conducted for polymorphism map construction, LD decay analysis, and nucleotide diversity analysis. Additionally, the phylogenetic tree analysis, F_{ST} statistical analysis, inbreeding coefficient analysis, principal component analysis, and population structure analysis were employed to analyze the population genetic relationship. The results showed that the number of SNPs in Dongjiang camel was the highest, the LD coefficient decayed the fastest, the nucleotide diversity was the highest, and the genetic

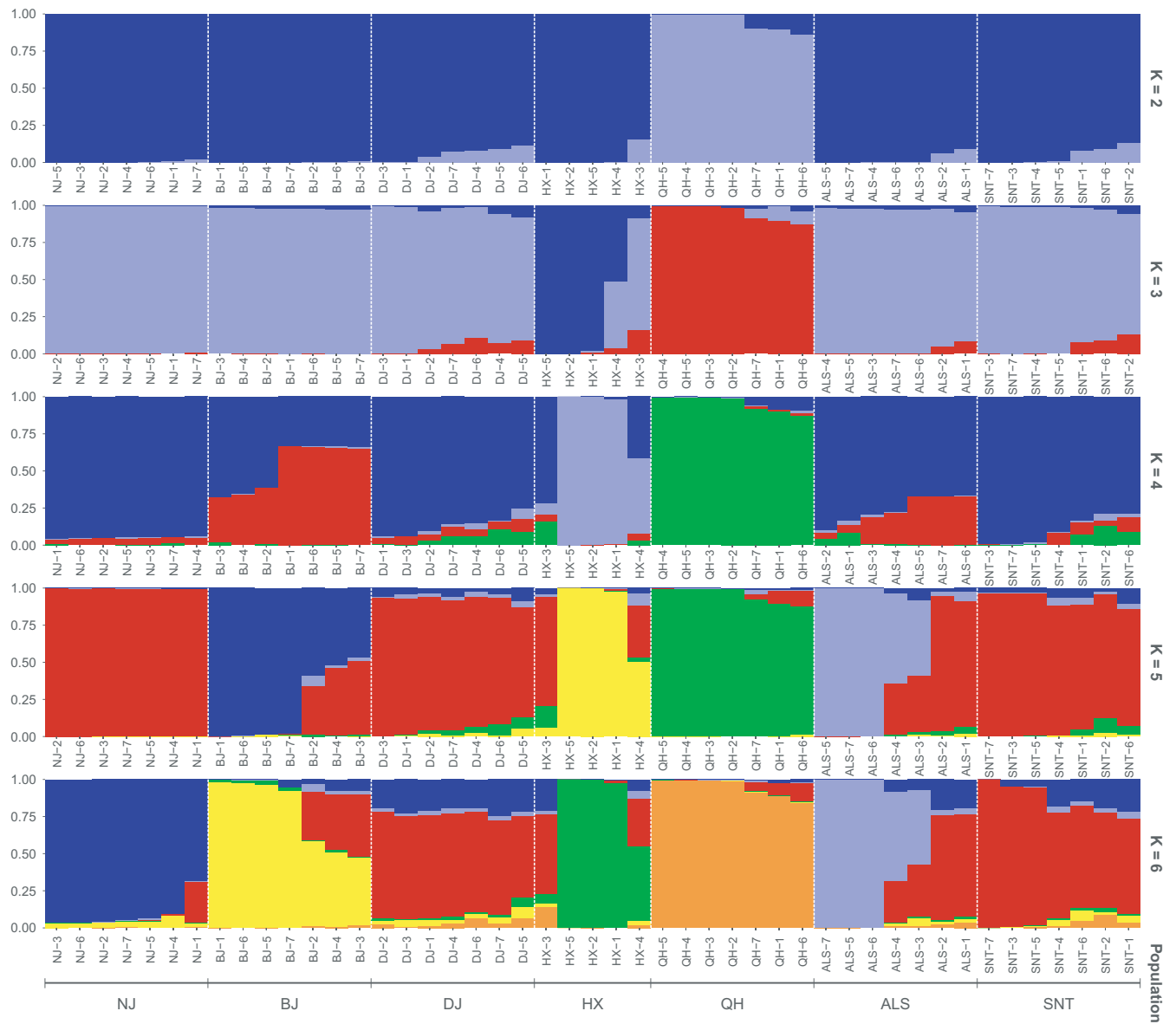


FIGURE 7 Groups structure clustering figure in domestic Bactrian camel populations. ALS, Alashan camels; BJ, Beijing camels; DJ, Dongjiang camels; HX, Hexi camels; NJ, Nanjiang camels; QH, Qinghai camels; SNT, Sunite camels

diversity was the highest, which may be due to the preservation of genetic diversity of ancestors (Abdulla et al., 2009). From the PCA map and the phylogenetic tree map, we can see that the Dongjiang camels clustered in two different branches. It indicates that Dongjiang camel may be mixed with other populations of Bactrian camels, resulting in higher genetic diversity of Dongjiang camel. In addition, the seven breeds of domesticated Bactrian camels were distinctly grouped and there was extensive genetic communication between the breeds. Nanjiang camel and Beijing camel gathered together, and the other five Bactrian camel groups (including Dongjiang camel, Hexi camel, Qinghai camel, Alashan camel, and Sunite camel) gathered together. It is consistent with the existing literature (Bai et al., 2017; Wang, Ren, Chen, Wang, & Wang, 2010). Since some previous studies on Chinese Bactrian camels (Gao, Wang, He, Chen, & Meng, 2009; Han, Ochieng, Lkhagva, & Hanotte, 2004) did not involve

Xinjiang groups, the selected groups were all in the second branch of this study, so it was concluded that there was no obvious genetic differentiation in the domestic Bactrian camel in China (He et al., 2012). Geographically, the Altai Mountains in the northern part of Xinjiang, the Tianshan Mountains, and the Taklimakan Desert in central Xinjiang have formed a natural barrier that isolates the Nanjiang camel and the Beijing camel from the other five groups. In another branch, the Sunite camel had a far-reaching genetic relationship with other groups, which may be related to its geographical distribution. In the south of Xilin Gol League, where the Sunite camel is distributed, the Yinshan Mountains form a natural barrier that blocks communication between the group and other groups (Bai et al., 2017). The Alashan camel and the Sunite camel had genetic communication, and the genetic relationship was relatively close. This may be related to the fact that they belong to the Inner Mongolia camel

and are geographically close. In the phylogenetic tree analysis, the genetic relationship among Qinghai camel, Hexi camel, and Alashan camel was relatively close, which may be related to its geographical distribution. They are all distributed near Qilian Mountain.

In the study, RAD-seq was first applied on camels. RAD-seq technology includes DNA sequence data from multiple loci in the genome at a lower cost, simpler library preparation process, no prior knowledge of any locus DNA sequences, management of data from simplified genomes, and well-developed pipelines for data treatment and analysis to retrieve orthologous sites for comparison (Eaton & Ree, 2013; McCormack, Hird, Zellmer, Carstens, & Brumfield, 2013; Rubin, Ree, & Moreau, 2012; Zhou et al., 2018). So far, various studies have been carried out on the origin, domestication, and genetic evolution of camels at home and abroad. Zhang (2014) studied the mitochondrial Cytb gene and D-loop sequence of the domestic Bactrian camel in China. Phylogenetic tree and network analysis indicated that all domestic Bactrian camels were clustered into one group and belonged to a single maternal origin. The mitochondrial genome on large-scale, wide-spread, domesticated Bactrian camels (including Mongolia, Russia, and China) was studied. It indicated that most of the mitochondrial haplotypes of domestic Bactrian camels were shared, supporting the claim that the domestic Bactrian camels belonged to the single maternal origin (Ming et al., 2017). At present, the earliest evidence for the domestication of Bactrian camels is concentrated in the northeast of Ilang and the neighboring southern Turkmenistan, especially in the Kopet Dagmountain region. Bactrian camel bones were found in the second cultural layer of the Anau site on the northern edge of the mountain range, dating back to 3500–3000 BC (Meadow & Zeder, 1978). However, the earliest archeological evidence of domestic Bactrian camels in China was the buck cemetery of Luntai group, which can be estimated as early as 800BC. Combined with the materials of the similar Nomhong ruins, it can be clearly assumed that there were domestic Bactrian camels in the northwestern part of China in the late Western Zhou Dynasty (Zhang & Luo, 2014). Domestic Bactrian camels are less likely to be domesticated in China and are more likely to be introduced directly from abroad (Bulliet, 1975). The most likely route is starting at Turkmenistan, passing successively through the south coast of the Aral Sea, today's Kazakhstan, Kyrgyzstan, the north and south of the Tianshan Mountains into the northern Xinjiang, and then getting to the Hexi corridor (Zhang & Luo, 2014). In addition, most experts have devoted themselves to the study of genomic DNA of Bactrian camels using microsatellites to reveal the genetic evolution of Bactrian camels distributed in different regions, providing basic research materials for the protection and utilization of Bactrian camels (Gao et al., 2017).

Male-specific region of Y chromosome in mammals does not recombine with X chromosome during meiosis. It has low mutation rate and is easy to form specific haploid. It follows strict paternal inheritance, so it is an important genetic resource for studying paternal genetic diversity (Cao et al., 2018). Chen, Ren, et al. (2018) and Chen, Zhou, et al. (2018) studied single nucleotide polymorphisms, microsatellites, and copy number variation (CNV) on the Y chromosome

of Chinese domesticated Bactrian camels to understand Y chromosome polymorphism of Chinese domesticated Bactrian camels. The polymorphic markers of the male-specific portion of the Y chromosome (MSY) provide useful information for tracking male genealogies. Based on single nucleotide variants, a Y-phylogenetic network with seven haplotypes found that wild and domestic Bactrian camels were obviously divided into two distinct groups (Felkel et al., 2019). Therefore, we need high-throughput sequencing to analyze the camel's Y chromosome to discover polymorphic markers on the Y chromosome in the next step, further enriching our results. In addition, we will actively seek cooperation with foreign countries in the next step of research, expanding camel blood samples along the Silk Road to further reveal the phylogenetic relationship of camels.

ACKNOWLEDGMENTS

This study was supported by the National Natural Science Foundation of China, 31172178.

CONFLICT OF INTEREST

None declared.

AUTHOR CONTRIBUTIONS

Zhanjun Ren and Chenmiao Liu designed the study and wrote the manuscript, Huiling Chen collected the samples, and Chengdong Zhang and Xuejiao Yang contributed to the data analysis. All authors read and approved the manuscript.

DATA AVAILABILITY STATEMENT

Domestic Bactrian camel DNA sequences: All raw RAD-seq reads data can be accessed at NCBI SRA. Bioproject # PRJNA522647, Biosample #s SAMN10948548–SAMN10948594.

ORCID

Zhanjun Ren  <https://orcid.org/0000-0001-5607-172X>

REFERENCES

- Abdulla, M. A., Ahmed, I., Assawamakin, A., Bhak, J., Brahmachari, S. K., Calacal, G. C., ... Zilfalil, B. A. (2009). Mapping human genetic diversity in Asia. *Science*, 326, 1541–1545. <https://doi.org/10.1126/science.1177074>
- Bai, L., Zhou, J. W., Li, X. H., & Feng, D. Z. (2017). Microsatellite genetic diversity analysis of six Bactrian camel populations in China. *China Animal Husbandry & Veterinary Medicine*, 44(6), 1734–1745. <https://doi.org/10.16431/j.cnki.1671-7236.2017.06.023>
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., ... Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, 3(10), e3376. <https://doi.org/10.1371/journal.pone.0003376>

- Bulliet, R. W. (1975). *The camel and the wheel*. Cambridge, MA: Harvard University Press.
- Cao, X. T., Pen, S. W., Zhang, J., Li, F. D., Li, G., Li, W., ... Le, X. P. (2018). Screening of Y-Chromosome specific primers and SNPs in sheep. *Chinese Agricultural Science*, 51(15), 170–186. <https://doi.org/10.3864/j.issn.0578-1752.2018.15.014>
- Chen, H. L., Ren, Z. J., Zhao, J. P., Zhang, C. D., & Yang, X. J. (2018). Y-chromosome polymorphisms of the domestic Bactrian camel in China. *Journal of Genetics*, 97(1), 3–10. <https://doi.org/10.1007/s12041-017-0852-1>
- Chen, S. F., Zhou, Y. Q., Chen, Y. R., & Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17), i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
- DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., ... Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43(5), 491–498. <https://doi.org/10.1038/ng.806>
- Eaton, D. A. R., & Ree, R. H. (2013). Inferring phylogeny and introgression using RAD-seq data: An example from flowering plants (Pedicularis: Orobanchaceae). *Systematic Biology*, 62(5), 689–706. <https://doi.org/10.1093/sysbio/syt032>
- Felkel, S., Wallner, B., Chuluunbat, B., Yadamsuren, A., Faye, B., Brem, G., ... Burger, P. A. (2019). A first Y-Chromosomal haplotype network to investigate male-driven population dynamics in domestic and wild bactrian camels. *Frontiers in Genetics*, 10, 423. <https://doi.org/10.3389/fgene.2019.00423>
- Gao, H. W., Wang, J., He, J. X., Chen, L. Y., & Meng, H. (2009). Analysis on the genetic diversity and evolution in the *Camelu bactrianus* using SSR markers. *Journal of Shanghai Jiaotong University (Agricultural Science)*, 27(2), 89–95.
- Gao, Y., Yin, S. I., Wu, L., Dai, D., Wang, H., Liu, C., & Tang, L. (2017). Genetic diversity and structure of wild and cultivated *Amorphophallus paeoniifolius* populations in southwestern China as revealed by RAD-seq. *Scientific Reports*, 7(1), 14183. <https://doi.org/10.1038/s41598-017-14738-6>
- Guo, C. Y. (2012). *Association analysis of genetic diversity of Jianghuai Soybean breeding germplasm and QTL of breeding target traits*. The thesis submitted to Nanjing Agriculture University.
- Han, J. L., Ochieng, J. W., Lkhagva, B., & Hanotte, O. (2004). Genetic diversity and relationship of domestic Bactrian camels (*Camelus bactrianus*) in China and Mongolia. *Journal of Camel Practice and Research*, 11(2), 97–99.
- He, X. H. (2011). *Genetic diversity, phylogenetic and mtDNA heterogeneity of major Bactrian camel populations in China*. The thesis submitted to Chinese Academy of Agricultural Sciences. Retrieved from <http://kns.cnki.net/KCMS/detail/detail.aspx?FileName=1011159174.nh&DbName=CDFD2011>
- He, X. H., Han, X. L., Guan, W. J., Tian, K. C., Zhang, W. & Ma, Y. H. (2012). Analysis of genetic variation and genetic relationship among populations of 10 Bactrian camel populations using microsatellite markers. *Biodiversity*, 20(2), 199–206. <https://doi.org/10.3724/SP.J.1003.2012.08159>
- He, X. H., Han, X. L., & Ma, Y. H. (2009). Advances in research on genetic diversity of Bactrian camel. *Journal of Livestock Ecology*, 30(4), 9–13.
- Jones, J. C., Fan, S. H., Franchini, P., Schartl, M., & Meyer, A. (2013). The evolutionary history of Xiphophorus fish and their sexually selected sword: A genome-wide approach using restriction site-associated DNA sequencing. *Molecular Ecology*, 22(11), 2986–3001. <https://doi.org/10.1111/mec.12269>
- Kang, J. L., Ma, X. H., & He, S. P. (2017). Population genetics analysis of the Nuijiang catfish *Creteuchiloglanis macropterus* through a genome-wide single nucleotide polymorphisms resource generated by RAD-seq. *Scientific Reports*, 7(1), 2813. <https://doi.org/10.1038/s41598-017-02853-3>
- Krishnakutty, R., Iskandarani, A., Therachiyil, L., Uddin, S., Azizi, F., Kulinski, M., ... Mohammad, R. M. (2018). Anticancer activity of camel milk via induction of autophagic death in human colorectal and breast cancer cells. *Asian Pacific Journal of Cancer Prevention*, 19(12), 3501–3509. <https://doi.org/10.31557/APJCP.2018.19.12.3501>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- McCormack, J. E., Hird, S. M., Zellmer, A. J., Carstens, B. C., & Brumfield, R. T. (2013). Applications of next-generation sequencing to phylogeography and phylogenetics. *Molecular Phylogenetics and Evolution*, 66(2), 526–538. <https://doi.org/10.1016/j.ympev.2011.12.007>
- Meadow, R. H., & Zeder, M. A. (1978). Approaches to faunal analysis in the Middle East. *Peabody Museum Bulletin*, 2, 91–103.
- Ming, L., Yi, L., Sa, R., Wang, Z. X., Wang, Z., & Ji, R. (2017). Genetic diversity and phylogeographic structure of Bactrian camels shown by mitochondrial sequence variations. *Animal Genetics*, 48(2), 217–220. <https://doi.org/10.1111/age.12511>
- Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E., & Lercher, M. J. (2014). PopGenome: An efficient Swiss army knife for population genomic analyses in R. *Molecular Biology and Evolution*, 31(7), 1929–1936. <https://doi.org/10.1093/molbev/msu136>
- Rubin, B. E. R., Ree, R. H., & Moreau, C. S. (2012). Inferring phylogenies from RAD sequence data. *PLoS ONE*, 7(4), e33394. <https://doi.org/10.1371/journal.pone.0033394>
- Sambrook, J., & Russell, D. W. (2002). *Translated by Huang, PT 444 molecular cloning a laboratory manual* (p. 445). Beijing, China: Science Press.
- Schraiber, J. G., & Akey, J. M. (2015). Methods and models for unravelling human evolutionary history. *Nature Reviews Genetics*, 16(12), 727–740. <https://doi.org/10.1038/nrg4005>
- Wang, L., Ren, Z. J., Chen, J., Wang, Y. X., & Wang, H. L. (2010). Genetic diversity of domesticated Bactrian camels in five regions of China. *Acta Agriculturae Boreali-occidentalis Sinica*, 19(11), 18–23.
- Wang, M.-S., Li, Y., Peng, M.-S., Zhong, L. I., Wang, Z.-J., Li, Q.-Y., ... Zhang, Y.-P. (2015). Genomic analyses reveal potential independent adaptation to high altitude in Tibetan Chickens. *Molecular Biology and Evolution*, 32(7), 1880–1889. <https://doi.org/10.1093/molbev/msv071>
- Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, 88(1), 76–82. <https://doi.org/10.1016/j.ajhg.2010.11.011>
- Zhai, Z., Zhao, W., He, C., Yang, K., Tang, L., Liu, S., ... Meng, H. E. (2014). SNP discovery and genotyping using restriction-site-associated DNA sequencing in chickens. *Animal Genetics*, 46(2), 216–219. <https://doi.org/10.1111/age.12250>
- Zhang, C. D. (2014). *Molecular characteristics and phylogenetic studies of paternal and maternal origins of Chinese camels*. The thesis submitted to Northwest A&F University. Retrieved from <http://kns.cnki.net/KCMS/detail/detail.aspx?FileName=1014440289.nh&DbName=CMFD2015>
- Zhang, X. Y., & Luo, Y. B. (2014). Archaeological observation on the origin of Chinese camel domestication. *Ancient and Modern Agriculture*, 1, 47–55.
- Zhou, W., Ji, X., Obata, S., Pais, A., Dong, Y., Peet, R., & Xiang, Q.-Y. (2018). Resolving relationships and phylogeographic history of the *Nyssa sylvatica* complex using data from RAD-seq and species distribution modeling. *Molecular Phylogenetics and Evolution*, 126, 1–16. <https://doi.org/10.1016/j.ympev.2018.04.001>

How to cite this article: Liu C, Chen H, Ren Z, Zhang C, Yang X. Population genetic analysis of the domestic Bactrian camel in China by RAD-seq. *Ecol Evol*. 2019;9:11232–11242. <https://doi.org/10.1002/ece3.5624>