Genetic diversity and population structure of Kazakh horses (Equus caballus) inferred from mtDNA sequences

M. Geminggali1, K.K. Iskhan1, Y. Li1, A. Qi1, W. Wuningfu1, J.Y. Ding1 and A. Wunschierjiang1

1College of Animal Science, Tarim University, Alar, Xinjiang, China
2Department of Animal Science, Kazakhstan National Agrarian University, Almaty, The Republic of Kazakhstan
3Institute of Arid Ecology and Environment, Xinjiang University, Ürümqi, China
4Genomics Laboratory of Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Ürümqi, China
5College of Animal Science and Veterinary, Inner Mongolia Agricultural University, Hohhot, China

Corresponding author: A. Wunscherjiang
E-mail: wunschierjiang@semi.com

Received March 10, 2016
Accepted July 5, 2016
Published October 5, 2016
DOI http://dx.doi.org/10.4238/gmr.15048618

Copyright © 2016 The Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution ShareAlike (CC BY-SA) 4.0 License.

ABSTRACT. The Kazakh horse is an important old horse breed in Xinjiang. They have contributed greatly to the breeding and improvement of other local horse breeds, yet their genetic diversity and population structure are not well understood. In the present study, we evaluated the genetic diversity of Kazakh horses and their relationship with other horse breeds using the mtDNA D-loop region, Cyto b gene, and a DNA fragment (nps 7974-9963, containing COX3, tRNA Gly, ND3, and tRNA Arg). A total of 130 Kazakh horses from

Genetics and Molecular Research 15 (4): gmr.15048618
8 populations in China and Kazakhstan were analyzed. A total of 88 haplotypes (haplotype diversity: 0.9895) were identified, in which 3 haplotypes were shared by groups in the two countries. In a median-joining network, 6 haplogroups were found, in which most haplogroups included haplotypes from different populations. Neighbor-joining analysis revealed similar results in that haplotypes in different populations were admixed in most of the 6 clusters. In conclusion, a high level of genetic diversity was found in the Kazakh horses. However, no clear correspondence between haplogroups and geographic origin and no significant differentiation between populations in the two countries were observed. This might have resulted from the frequent contact between the two countries through the Silk Road in the past, or due to long-term outcrossing and hybridization with the introduced horses.

**Key words:** Kazakh horses; Genetic structure; Population diversity; D-loop; Cyt b

**INTRODUCTION**

An understanding of the genetic characteristics of a population is essential to implement conservation strategies that preserve genetic diversity (Hall and Bradley, 1995). Appropriate assessment and characterization of populations at the phenotypic and genotypic levels are some of the first and crucial steps in the development of conservation strategies (Plante et al., 2007; Giraz et al., 2008). Molecular data can aid in identifying animals or sets of animals that should be preserved to prevent loss of genetic diversity. Based on genetic diversity and population structure analysis of Brazilian Mangalarga Marchador horses, DeAssis et al. (2009) demonstrated the low level of genetic diversity of Mangalarga Marchador horses and recommended two conservation strategies: avoidance of crosses between highly endogamous individuals and stimulation of crosses between individuals from those regions identified as having low gene flow. In addition, genetic diversity studies of the Kiso horse indicated that the genetic diversity of the maternal lineage has been reasonably well maintained. However, since the population was small, it might be in the extinction vortex and the genetic diversity of the maternal lineage of Kiso horse would be decreasing. Therefore, key measures, such as the use of reproductive technologies and investigation of new ways to utilize the horses, should be implemented for preservation (Takasa et al., 2014).

The Kazakh horse is an old horse breed. The present Kazakh horses are thought to be the descendants of the ancestral Wusu horses of the 2nd century B.C. The development of the present Kazakh horse is the result of long-term breeding and improvement. Their development are closely related to the need for nomadic transportation along the Silk Road, wars, and horse exchanges between China and Central Asia. Kazakh horses are mainly distributed in western China, Kazakhstan, northeast Kyrgyzstan, Mongolia, and the western Altai region of Russia. In China, they are mainly found in the areas of the Tianshan Mountains, west of Junggar Basin, and the western part of the Altay Mountains. The greatest numbers of Kazakh horses are in Xinyuan County, while Nilka County has the greatest numbers of purebred horses. Kazakh horses in Zhaozhou, Turks, and Gonglia County were selected and bred to form the Illi horse. In addition to the Illi region, Boertala Mongol Autonomous Prefecture, Tacheng, the
Altay region, and Changji Hui Autonomous Prefecture also raise Kazakh horses, but they are generally fed together with Mongolian horses (Zhao, 1991).

The typical characteristics of Kazakh horses include moderately large head, short ears, slender and slightly raised neck, high withers, narrow chest, and often knife-like hind legs. The length and height of adult horses average 150 and 152 cm and 145 and 153 cm for stallions and mares, respectively. They are well adapted to the cold climate, moderate in temperament, and grow rapidly (Gemmingguli, 2011). An adult horse (4-12 years) weighs 340 to 440 kg and the meat production rate is around 57%. A mare can produce 5 to 6 kg milk per day under grazing conditions. The meat and milk are delicious and rich in nutrients necessary for humans. They play a key role in the health and lives of local people, as well as in the development of the local economy and society (Tiemuerbai, 2014).

Over a long period of time, because close attention was not paid to breeding, the quality and quantity of Kazakh horses in China dramatically decreased, which led to their smaller physique, weight loss, and other desirable features of the breed gradually degenerating (Tiemuerbai, 2014). The number of Kazakh horses has reduced significantly in recent years with the increase in the consumption of horse meat in the Xinjiang region. Moreover, the number of purebred Kazakh horses is decreasing because the majority of Kazakh horses (such as in the ili regions) have been hybridized with introduced horses. Therefore, effective measures to conserve and improve this rare horse should be developed and implemented (Gemmingguli, 2011).

Owing to the high level of base pair substitution rate compared with nuclear DNA (Brown et al., 1978), strict maternal inheritance (Hutchison et al., 1974), and lack of recombination (Bowling et al., 2000), mtDNA is powerful for clarifying scientific issues regarding horses, such as genetic diversity in Serbia horses (Ocokoljic et al., 2013), Italian horses (Bigi et al., 2014), and Sicilian autochthonous horse breeds (Guanella et al., 2011), adaptive evolution (Ning et al., 2010), phylogenetic relationships between horse populations (Miroi et al., 2002; Pryzupa et al., 2012; Winton et al., 2013), and the origin of some horse breeds (Ciastek et al., 2010; Guanella et al., 2011; Loppeld et al., 2011; Achilli et al., 2012).

Several studies have been conducted on horses using mtDNA; however, only few studies have focused on Kazakh horses. To date, little is known about their diversity or phylogenetic relationships with other horse breeds. The aims of this study were to assess the genetic diversity and population structure of the Kazakh horses using mtDNA sequences, including the D-loop region, the cytochrome oxidase b (COI) gene, and a DNA fragment (nps 7974-9963, containing gene partial sequences for the COX1 gene, tRNA-Gly, the ND3 gene, and tRNA-Arg sequences). Phylogenetic relationships between Kazakh horses and other breeds worldwide were also analyzed.

MATERIAL AND METHODS

Venous blood (2-3 mL) was drawn from the neck of each adult horse and immediately stored in EDTA-containing anticoagulation vacuum blood vessels at -20°C. A total of 130 samples were collected from 130 Kazakh horse individuals from 8 populations; 70 individuals were from 6 populations: Altai (AL), Hefeng (HE), Tuoli (TE), Bole (BL), Hermu (HM), and Manasi (MN) in China and 60 individuals were from 2 populations: Taidykorgan (TAL) and Altay (ALM) in Kazakstan. The treatment of horses followed the supporting institution’s ethical standards. In addition, 168 accessions of D-loop sequences from 25 horse breeds were retrieved from GenBank to investigate the relationships between Kazakh horses and other horse breeds (Table S1).
Total genomic DNA was extracted using a rapid blood genomic DNA extraction kit (Sangon Biotech, China). The primers used for amplification of the Cyp b gene, D-loop region, and a DNA fragment (nps 7974-9963 containing the COX3 gene, IrRNA-Gly, ND3 gene, and mitochondrial tRNA-Arg sequences, hereafter called the 7974 fragment) are listed in Table 1. PCR primers were synthesized by Sangon Biotech. PCRs included QIAGEN HotStar High fidelity polymerase (Germany). Each reaction was performed in a volume of 25 μL using 13.5 μL ddH2O, 5 μL HotStar PCR buffer, 2.5 μL forward primer (10 μM), 2.5 μL reverse primer (10 μM), 1 μL (30-50 ng) DNA sample, and 0.5 μL HotStar polymerase. DNA amplifications were performed in a GeneAmp PCR system 9700 (Applied Biosystems, USA) under the following conditions: initial denaturation at 95°C for 5 min; followed by 35 cycles of denaturation at 94°C for 30 s, annealing at 60°C for 30 s, and extension at 72°C for 1 min; and a final extension at 72°C for 10 min. PCR products were purified with a TIAN Quick Midi Purification Kit (TIANGEN, Beijing, China) and sequenced bidirectionally with forward and reverse PCR primers using an ABI PRISM® 3730 DNA sequencer by Sangon Biotech.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Forward 1</th>
<th>Prime sequence (T3)</th>
<th>Reverse 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyp</td>
<td>CCA---------CTC-TCT-TGTT-TGC-AGT-C</td>
<td>TCAATCCAGCGGCTGAATCTAACC</td>
<td></td>
</tr>
<tr>
<td>D-loop</td>
<td>GCA---------CTC-TCT-TGTT-TGC-AGT-C</td>
<td>CAGCTGACTATGACGACAGAAGAACCT</td>
<td></td>
</tr>
<tr>
<td>ND3</td>
<td>GCA---------CTC-TCT-TGTT-TGC-AGT-C</td>
<td>TGGATTGATGATATTACGAGCA</td>
<td></td>
</tr>
<tr>
<td>ND4</td>
<td>CCA---------CTC-TCT-TGTT-TGC-AGT-C</td>
<td>TCCAAAACACTCTCTCAGGACG</td>
<td></td>
</tr>
<tr>
<td>ND4L</td>
<td>CCA---------CTC-TCT-TGTT-TGC-AGT-C</td>
<td>CAGCTTATATTCTACTACGTACAC</td>
<td></td>
</tr>
<tr>
<td>ND6</td>
<td>CCA---------CTC-TCT-TGTT-TGC-AGT-C</td>
<td>TCCGAACTTTGTATACAGCTAC</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>CCA---------CTC-TCT-TGTT-TGC-AGT-C</td>
<td>TCCTTCCAGGAAGCCAGAATAT</td>
<td></td>
</tr>
</tbody>
</table>

All sequences were examined by eye in Chromas 2.4 and BioEdit version 7.0.5.3. All sequences were initially aligned using the Clustal W software (Thompson et al., 1994) with default settings. The statistical quantities for the DNA sequences, including numbers of haplotypes (Nh), haplotype diversity (Hd), and nucleotide diversity (ht), were calculated using DnaSP 5.10.1 (Librado and Rozas, 2009). The genetic distances within and between populations were calculated with the MEGA 6 software package (Tamura et al., 2013). To represent the genetic structure and differentiation of tested populations, analysis of molecular variance (AMOVA) and pairwise Fst were conducted. AMOVA and pairwise Fst were performed using the Kimura 2-parameter model with Arlequin v3.5 (Excoffier and Lischer, 2010). Cyp b, D-loop, and the 7974 fragment were evaluated independently and in combination. For the interpretation of pairwise Fst results, we followed the suggestion that refers that values between 0.0-0.05 indicates little genetic differentiation; values between 0.05 and 0.15, moderate differentiation; values between 0.15 and 0.25, great differentiation; and values above 0.25, very great genetic differentiation (Wright, 1978; Hartl and Clark, 1997). A median-joining network (MJ network) was constructed to detect haplotypic differentiation in the horse populations with the NETWORK 4.6 software (Bandelt et al., 1999), based on concatenated combined sequences. Default settings were applied. Within each haplotype in the MJ network, different color codes represented the proportions of different populations in each haplotype.

Phylogenetic analysis of the haplotypes was performed using the neighbor-joining (NJ) and Bayesian method with the MEGA 6 software package (Tamura et al., 2013) based on the Kimura 2-parameter model and the BEAST v1.7 software package (Drummond et al.,

Genetics and Molecular Research 15 (4): gmr.15048818
2012) based on the TN93+G+I model. The 50% majority-rule consensus NJ tree was obtained from 1000 bootstrap replications. For BEAST analyses, two independent runs of 10 million generations each were performed, sampling every 1000th generation. Each run was checked in Tracer v1.5 for sufficient effective sample size (over 100 as suggested by the authors) in the main relevant statistics. The two runs (log- and tree-files) were then combined with LogCombiner v1.7.4. The maximum clade credibility tree was generated from the 90% post-burn-in trees file using TreeAnnotator 1.7.4. Finally, Figure Tree 1.4.0 was used to display phylogenetic trees.

RESULTS

Genetic diversity

Genetic diversity is usually estimated by using several parameters, e.g., the average number of nucleotide differences (k, mean number of pairwise differences in a DNA dataset of a population), number of polymorphic sites (Nps, number of variable nucleotides regarding one specific DNA region in the sample population), nucleotide diversity (π, a measure of genetic variation, defined as the average number of nucleotide differences per site between any two DNA sequences chosen randomly from the sample population), number of haplotypes (Nh), and haplotype diversity (Hd, a measure of the uniqueness of a particular haplotype in a given population) for each population or group.

For each of the D-loop, Cyt b gene, and the 7974 fragment, 130 sequences were obtained. The alignment length was 708, 1140, and 2020 bp, respectively. The nucleotide diversity (π) was 0.0220, 0.0486, and 0.0056 for the three DNA regions, respectively. The number of haplotypes and haplotype diversity for the three DNA regions were 7/0.940, and 47/0.959, respectively (Table 2).

After concatenating the Cyt b, D-loop, and the 7974 fragment, a sequence with a total of 3868 bp was obtained. Among 222 polymorphic sites, 162 were potentially phylogenetically informative. The k (gap/missing data were excluded) and π were 25.55 ± 11.3 and 0.0095. The k and π for China horse groups were 25.67 ± 11.39/0.0094 and 25.43 ± 11.31/0.0093 for Kazakhstan groups, respectively. The corrected pairwise difference between China and Kazakhstan groups was 0.02 (P > 0.01). A total of 88 haplotypes (Hd: 0.9885) were identified in which 8 haplotypes, H9, H27 and H43, were shared by the two geographic groups. In China populations, 53 unique haplotypes were found and 32 were observed in Kazakhstan populations. The K-2-P genetic distance within the 8 populations ranged from 0.005 to 0.012 and between populations from 0.008 to 0.011 (Table 3).

Distances between populations are below the diagonal and SE values are above the diagonal. Distances within a population (bold) are on the diagonal.

Network analysis and population differentiation

The MJ network resolved 6 haplogroups, which were referred to as groups A, B, C, D, E, and F (Figure 1). Group A included 29 haplotypes, 11 were from Kazakhstan, 20 from China. Haplotypes H9 and H43 were shared by China and Kazakhstan. The 19 haplotypes of B were from Kazakhstan (10) and China (9). The 24 haplotypes in C were also from China (16) and Kazakhstan (9) and 1 shared between them (H27).
Table 2: Populations tested in the study.

<table>
<thead>
<tr>
<th>Geographic Location</th>
<th>Population</th>
<th>N</th>
<th>Ns</th>
<th>Np</th>
<th>Ns</th>
<th>Np</th>
<th>Hs</th>
<th>Hs (95% CI)</th>
<th>Hs</th>
<th>Hs</th>
<th>Dp</th>
<th>Dp (95% CI)</th>
<th>Dp</th>
<th>Dp (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>China</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jilin</td>
<td>Ns</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>0.86</td>
<td>0.86</td>
<td>5.00</td>
<td>4.80-5.20</td>
<td>12</td>
<td>12</td>
<td>5.00</td>
<td>4.80-5.20</td>
<td>12</td>
</tr>
<tr>
<td>Hebei</td>
<td>Ns</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>0.86</td>
<td>0.86</td>
<td>5.00</td>
<td>4.80-5.20</td>
<td>8</td>
<td>8</td>
<td>5.00</td>
<td>4.80-5.20</td>
<td>8</td>
</tr>
<tr>
<td>Shandong</td>
<td>Ns</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>0.86</td>
<td>0.86</td>
<td>5.00</td>
<td>4.80-5.20</td>
<td>8</td>
<td>8</td>
<td>5.00</td>
<td>4.80-5.20</td>
<td>8</td>
</tr>
<tr>
<td>Jiangsu</td>
<td>Ns</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>0.86</td>
<td>0.86</td>
<td>5.00</td>
<td>4.80-5.20</td>
<td>14</td>
<td>14</td>
<td>5.00</td>
<td>4.80-5.20</td>
<td>14</td>
</tr>
<tr>
<td>Shaanxi</td>
<td>Ns</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>0.86</td>
<td>0.86</td>
<td>5.00</td>
<td>4.80-5.20</td>
<td>9</td>
<td>9</td>
<td>5.00</td>
<td>4.80-5.20</td>
<td>9</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>71</td>
<td>71</td>
<td>71</td>
<td>71</td>
<td>0.86</td>
<td>0.86</td>
<td>5.00</td>
<td>4.80-5.20</td>
<td>71</td>
<td>71</td>
<td>5.00</td>
<td>4.80-5.20</td>
<td>71</td>
</tr>
<tr>
<td>Korea</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gyeongbuk</td>
<td>Ns</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>0.86</td>
<td>0.86</td>
<td>5.00</td>
<td>4.80-5.20</td>
<td>12</td>
<td>12</td>
<td>5.00</td>
<td>4.80-5.20</td>
<td>12</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>24</td>
<td>0.86</td>
<td>0.86</td>
<td>5.00</td>
<td>4.80-5.20</td>
<td>24</td>
<td>24</td>
<td>5.00</td>
<td>4.80-5.20</td>
<td>24</td>
</tr>
</tbody>
</table>

N: number of individuals in each population. Np: the number of haplotypes in each population and the haplotype diversity (Hs) are given in parentheses. Np: the number of polymorphic sites. n: nucleotide diversity. Dp: average number of nucleotide differences.
Most of the 7 haplotypes in D were from China (6), only 1 from Kazakhstan. Group E had 7 haplotypes in total with 3 from China and 4 from Kazakhstan. Haplotypes from China and Kazakhstan were admixed. There was no congruence of haplogroup distribution to populations from different geographic areas.

For AMOVA, we divided the 8 populations into 2 groups according to geographic origins. One group combined populations in China and the other included populations in Kazakhstan. The results partitioned variation in the D-loop to -1.83% ($F_{CT} = 0.018$) occurring among groups, 6.58% ($F_{ST} = 0.069$, $P < 0.01$) among populations within groups, and 94.88% ($F_{II} = 0.052$, $P < 0.01$) occurring within populations. In Cyt b, -1.51% of the variation ($F_{CT} = 0.015$) was explained by differences among groups, 6.36% ($F_{ST} = 0.063$) among populations within groups, and 95.75% ($F_{II} = 0.049$) within sampled populations. For the 7974 fragment sequences, 0.70% ($F_{CT} = 0.007$) of variation occurred among groups, 1.66% ($F_{ST} = 0.017$) among groups within populations, and 97.63 ($F_{II} = 0.024$) within populations (Table 4).

AMOVA results based on combined sequences showed that -0.91% ($F_{CT} = 0.009$) of variation occurred among groups, 5.35% ($F_{ST} = 0.053$) among populations within groups, and 95.56% ($F_{II} = 0.044$) within populations (Table 4).
Table 4. Analyses of molecular variance based on D-loop, Cyt b, and 7974 fragment sequences for the horse populations.

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Population</th>
<th>Percentage of variation explained</th>
<th>Percentage of variation explained</th>
<th>Percentage of variation explained</th>
<th>Percentage of variation explained</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>Within groups</td>
<td>Among groups</td>
<td>Among populations with groups</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.15</td>
<td>0.10</td>
<td>0.25</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White population</td>
<td>4.34</td>
<td>4.34</td>
<td>3.68</td>
<td>5.02</td>
<td>6.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>P &lt; 0.01</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Of 28 pairwise $F_{ST}$ values between the 8 populations based on combined data, 17 comparisons had $F_{ST}$ values between 0 and 0.05 showing little genetic differentiation while 8 comparisons had $F_{ST}$ values between 0.05 and 0.15 showing moderate genetic differentiation but not significant ($P > 0.01$), and 2 comparisons (AL-HE and HE-TE) had $F_{ST}$ values between 0.15 and 0.25 showing great genetic differentiation. Negative $F_{ST}$ values were recorded in some comparisons and these equate to zero $F_{ST}$ values. While most of the lowest $F_{ST}$ values were seen between populations of TE, BL, HM, and MN and TAL and ALM, the highest $F_{ST}$ values were between the HE population and three other populations (AL, TE, and HM). AL and HE comparisons showed high genetic differentiation, with $F_{ST}$ value of 0.296 (Table 5). The pairwise $F_{ST}$ values between China and Kazakhstan groups were 0.03 (P > 0.01) indicating little genetic differentiation between Kazakh horses from the two countries.

Table 5. Pairwise $F_{ST}$ values among populations.

<table>
<thead>
<tr>
<th>Population*</th>
<th>AL</th>
<th>HE</th>
<th>TE</th>
<th>BL</th>
<th>HM</th>
<th>MN</th>
<th>TAL</th>
<th>ALM</th>
</tr>
</thead>
<tbody>
<tr>
<td>AE</td>
<td>0.20**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HE</td>
<td>0.002</td>
<td>0.13**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TE</td>
<td>0.114</td>
<td>0.097</td>
<td>0.055</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BL</td>
<td>0.547</td>
<td>0.561</td>
<td>0.591</td>
<td>0.096</td>
<td>0.086</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HM</td>
<td>0.008</td>
<td>0.132</td>
<td>0.158</td>
<td>0.020</td>
<td>0.036</td>
<td>0.032</td>
<td>0.053</td>
<td>0.054</td>
</tr>
<tr>
<td>MN</td>
<td>0.012</td>
<td>0.016</td>
<td>0.016</td>
<td>0.016</td>
<td>0.014</td>
<td>0.018</td>
<td>0.067</td>
<td>0.067</td>
</tr>
<tr>
<td>TAL</td>
<td>0.118</td>
<td>0.126</td>
<td>0.034</td>
<td>0.026</td>
<td>0.036</td>
<td>0.014</td>
<td>0.016</td>
<td>0.016</td>
</tr>
<tr>
<td>ALM</td>
<td>0.135</td>
<td>0.120</td>
<td>0.016</td>
<td>0.016</td>
<td>0.016</td>
<td>0.033</td>
<td>0.053</td>
<td>0.054</td>
</tr>
</tbody>
</table>

Negative values equal to zero. **P < 0.01.

Phylogenetic analysis

The haplotypes of Kazak horses are grouped into six major lineages (Figure 2). Four of the six lineages (A, B, C, and D) corresponded to the four haplogroups (A, B, C, and D) shown in the MJ network (Figure 1). The other two lineages (E and F) corresponded to haplogroup E (Figure 1). Phylogenetic inference based on combined sequences supported the monophyly of most lineages (Figure 2). Five lineages (A to E) included haplotypes from horses in Kazakhstan and China. Haplotypes H9 and H43 were shared by horses in China and Kazakhstan. Lineage F only contained two haplotypes from China populations. The phylogenetic analyses showed that haplotypes from China and Kazakhstan populations were admixed. Five main clades (A-D and F, Figure 2) correspond to the haplogroups (A-D and F, Figure 1). Haplogroup E (Figure 1) contains not only clade E (Figure 2) but also some haplotypes that are not united in any of the above six clades, such as H7, H73, H26, and H64.

Genetics and Molecular Research 15 (4): gm15048618
Figure 2. Consensus neighbor-joining tree of the 88 haplotypes. The tree is based upon 1000 bootstrap replicates. The reference *Equus asinus* sequence NC0017798.1 is an outgroup. Bootstrap values are shown as percentages above branches. The individuals with each haplotype are represented by colored circles indicating different populations. Different lineages (A–F) are indicated.
Relationships between Kazakh and other horse breeds

Apart from Kazakh breed, 25 horse breeds, Abaga, Akhal-Teke, Arabian, Baise, Belgian, Caspian, Cleveland Bay, Clydesdale, Cheju, Debao, Ecomor, Friesian, Garmaco, Haflinger, Jetfa, Lustiano, Mongolian, Noriker, Prewalski, SanDie, Shetland, Soroniai, Thoroughbred, Wuzharqai, and Yunnan, were used in the analyses. Within the Kazakh populations, 12 haplotypes were also found in individuals from a number of breeds, including Akhal-Teke, Cheju, Debao, Caspian, SanDie, Haflinger, Noriker, Arabian, and Soroniai (Figure 3). Some of the Kazakh horse haplotypes were either embedded or clustered with other breeds of horses (Figure 4).

Figure 3. Phylogenetic relationships between the Kazakh horse breed and other horse breeds. BEAST phylogenetic tree was constructed with the D-loop region of 243 horse individuals including the Kazakh and 25 other breeds. Taxa names of Kazakh horses are indicated in red. Equus asinus (NC001788.1) was used as the outgroup.

Genetics and Molecular Research 15 (4): gmr.15048618
DISCUSSION

Genetic diversity plays an important role in the survival and adaptability of a species (Frankham, 2005). Variation in the population's gene pool provides variable traits among the individuals of that population. These variable traits can be selected via natural selection, ultimately leading to an adaptive change in the population, allowing it to survive in a changed environment. If a population of a species has a highly diverse gene pool, individuals of that population will possess more variety of traits, thereby more traits for natural selection will act upon to select the fittest individuals to survive (Pullin, 2002). Genetic diversity is essential for a species to evolve. With very little genetic variation within a species, successful reproduction becomes increasingly difficult, and offspring are more likely to have issues resulting from inbreeding.

Some wild species, especially horse breeds, are subjected to small populations and inbreeding; thus, these species are facing the loss of genetic diversity. The Kazakh horse breed is an original local variety formed by long-term manipulation by people of all nationalities in Xinjiang. They are important to the local people, not only because they are a source of food (meat and milk), but an efficient tool in grazing or transport, and are used for horse breeding, but also they are of cultural and entertainment interest, playing an important role in the national culture, painting, music, dance, and horse racing.

Because of the lack of improvement, increase in consumption of horse meat, and excess hybridization with introduced horse breeds, the quality and quantity of Kazakh horses in China decreased dramatically, especially purebred Kazakh horses. It has been suggested that effective measures should be taken to protect and improve the Kazakh horses (Geninggali, 2011). Accurate assessment of genetic diversity and structure is crucial to design and implement conservation strategies in natural and domesticated species. To date, there have been no studies on population structure and genetic diversity of Kazakh horses. Therefore, we
applied mitochondrial DNA data to investigate the genetic diversity of the Kazakh horse breed and its relationships with some other horse breeds.

A high level of genetic variability was observed in the Kazakh horses. Haplotype diversity among Kazakh horses was 0.9895, similar to that estimated in other horses, such as the Barblian horse (0.974), Italian heavy draught horses (0.991), Haflinger horses (0.926), Lipizzan horses (0.941), Macedonian horses (0.991), Murgec horses (0.964), Ventusso horses (0.900) (Bigi et al., 2014), Thoroughbred horses (1.00), Arab horses (0.96), Barb horses (1.00), and Andalusian horses (1.00) (Guatellia et al., 2011). Among the haplotypes found in Kazakh horses, only three (H9, H27, and H43) are shared by horses in China and Kazakhstan. A higher haplotype diversity was found in Kazakh horse populations in China than in Kazakhstan. No congruence of haplogroup to a population’s geographic origin was observed. Most of the haplogroups in the MJ network (Figure 1) contained haplotypes found both in China and Kazakhstan populations and no haplogroup included haplotypes from a single population. The six clades in the NJ tree largely corresponded with the haplogroups in the MJ network.

However, the haplotypes were admixed and none of the phylogeographic clusters corresponded to populations of geographic areas or groups of the two countries. The pattern of broadly mixed haplotypes might suggest a recent history of hybridization or multiple origins of the Kazakh horse. *E. caballus* is thought to have multiple origins of domestication along with a strong bias for the recruitment of mares over stallions from the wild (Vilà et al., 2001; Jansen et al., 2002; Walther et al., 2003; Ciesiak et al., 2010).

Pairwise FST values between the populations and groups were estimated. Most populations show little differentiation with FST < 0.05 (P > 0.01) or moderate but not significant differentiation (FST > 0.05, P < 0.01). However, great differentiation was found between AL and HE populations, and HE and TE populations, both with FST > 0.15 (P < 0.01). Overall, little genetic differentiation was found between Kazakh horse groups from the two countries as the FST values between them was 0.02 (P > 0.01). AMOVA showed no subdivision between the two groups. None of the variance (0.991%) is accounted for among the two groups, while a large fraction of variation (95.56%) was found within populations. Genetic distances also support that differentiation between Kazakh horses in China and Kazakhstan is very low. X-2-P distance between groups (1.0%) was not larger than distances within groups (1.0% each). Our results show that there is no clear pattern of differentiation among Kazakh horse populations from different geographic locations. This pattern indicates a close genetic relationship between the two horses from the two countries. This result might be because the two countries have frequent communication through the Silk Road in the past. In addition, long-term outcrossing and hybridization with introduced horses may be a factor. One study on maternal genetics of Kazakh horses revealed that the Kazakh horses have a rich maternal background, and the Chinese Kazakh horses and Kazakh horse have close genetic relationships (Qi et al., 2014). Similar results are seen in other horse breeds, such as the Arabian horse (Khanhong and Cothren, 2013), the Cheju horse (Yang et al., 2002), and the Hispano-Breton heavy horse (Pérez-Gutiérrez et al., 2008). This can be explained by separate and geographically diverse populations participating in the domestication process or by domestic horses having multiple origins (Lister et al., 1998; Vilà et al., 2001; Jansen et al., 2002; Khanhong and Cothren, 2013).

In contrast with little differentiation in genetic variation and population structure, Kazakh horses in Kazakhstan and China have some obvious differences in their phenotype. After long periods of selective breeding, Kazakh horses in Kazakhstan have some better physical characteristics and performance than those in China. Individuals from Kazakhstan
are generally larger than those from China; for example, stallions are 500-600 kg and mares are 400-450 kg for Kazakhstan individuals, whereas stallions are 320-400 kg and mares are 260-350 kg for China horses. The average milk production of Kazakh horses in Kazakhstan is 12-15 L per day and only 3-8 L in China. The incongruence between genotypes and phenotype can result from limited or unbalanced sampling. In the present study, we found that some haplotypes are unique in Kazakhstan populations. These genetic resources can be used for horse breeding or improvement of Kazakh horses in China in the future.

Genetic relationships between Kazakh and other breeds

Our results reveal that some of the Kazakh haplotypes are either shared or have close relationship with other kinds of horses. Within the Kazakh populations, 12 haplotypes were found in individuals from a number of breeds, including the Akhal-Teke, Cheju, Dehao, Caspian, SanHe, Haflinger, Noriker, Arabian, and Sormia (Figure 3), perhaps indicating that the Kazakh horse shares a common maternal ancestry with several different types of equines or there might have been gene flow among these horse breeds in the past. Further, a BEAST phylogenetic tree was constructed with the same sequence data. Phylogenetic analyses show that Kazakh horse mitochondrial sequences are widespread and distributed in many different clusters in the tree (Figure 3). Kazakh sequences cluster with many of the breeds, such as Akhal-Teke, Cheju, Dehao, Caspian, SanNe, Haflinger, Noriker, Arabian, Sormia, Mongolian, and Thoroughbred, to form several clades. Similarly, little congruence of haplogroup distribution to breeds or geographic areas has been found as demonstrated in Cheju horses (Yang et al., 2002), Arabian horses (Khanshur and Cothran, 2013), German draught horses (Aberle et al., 2007), and Italian horses (Bigi et al., 2014). Achilli et al. (2012) analyzed 83 mitochondrial genomes from modern horses across Asia, Europe, the Middle East, and the Americas. Their data revealed 18 major haplogroups (Â-R). Most of the 17 haplogroups identified in domestic breeds are spread over different geographic areas. This suggests that mitochondrial lineages are not powerful for identifying horse breeds because their diversification is thought to have occurred prior to domestication (Vilà et al., 2001; Jansen et al., 2002) or the existence of multiple maternal lines in the analyzed autochthonous breeds.

The shared haplotypes between Kazakh and other horse breeds and admixture between these breeds in phylogenetic analysis demonstrate past gene flow by hybridization with different horse breeds. The Kazakh horse breed is thought to have been influenced as early as the 5th century B.C. by many breeds, such as Mongolian, Karabair, Arabian, and Akhal-Teke. In the late 20th century, Kazakh horses were crossed with the Thoroughbred, Orlov Trotter, and Don (Dmitriev and Ernst, 1989) breeds. The development and success of the Kazakh horse breed is closely related to the nomadic transportation along the Silk Road and wars in these areas, as well as the vast lush grasslands and horse exchange between Xinjiang and Central Asia over the past several thousand years.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

Research supported by the International Cooperation Project of the Xinjiang Production and Construction Corps (Grant #2011BC007).

Genetics and Molecular Research 15 (4): gmr.15048618
REFERENCES


Genetics and Molecular Research 15 (4): gmr.15048618
Genetic diversity and population structure of Kazakh horses


Supplementary material

*Table S1*: D-loop sequences downloaded from Genbank and accession numbers.

---

*Genetics and Molecular Research* 15 (4): gmr.1548618