

Analysis of Mitochondrial D-Loop Cloning Sequence Variations in Yak (*Bos Grunniens*)

Liu Bo

Northwest Institute of Plateau Biology, Chinese Academy of Science <https://orcid.org/0000-0002-9414-3461>

Xingwang Lu

Northwest Institute of Plateau Biology Chinese Academy of Sciences

Haiqing Wang

Northwest Institute of Plateau Biology Chinese Academy of Sciences

Xiaoyan Tao

Northwest Institute of Plateau Biology Chinese Academy of Sciences

Qien Yang

Northwest Institute of Plateau Biology Chinese Academy of Sciences

Quanwen Dou (✉ douqw@nwipb.cas.cn)

<https://orcid.org/0000-0001-5910-6270>

Research article

Keywords: Yak (*Bos grunniens*), mitochondrial D-loop, heteroplasmy

Posted Date: August 17th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-52757/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Yaks (*Bos grunniens*) are a livestock species that is highly adapted to the extremely conditions of the Qinghai-Tibet plateau. Heteroplasmy and nuclear-mitochondrial DNA have been reported in many species, but are still unknown in yaks. The diversity of D-loop sequences, derived from the sequencing of PCR products, provides widely-used maternal DNA markers for the study of yak population and biology. D-loop variations were examined by sequencing the mono clones of the PCR products from individuals, and a comparative analysis with other documented PCR-product based sequences was conducted.

Results: Twenty-two and 27 variants were detected in 25 and 38 clones in two different yak individuals, respectively. Phylogenetic analysis in each demonstrated that variants could be grouped into two distinct major clades individual. Analysis of the D-loop sequences of domesticated yaks and wild yaks (*Bos grunniens mutus*) from the gene bank, which were originally obtained by sequencing of the PCR products, also revealed the two distinct clades. Integrative analysis of all the above variants indicated two major distinct clades, consistent with the respective analysis. One common variant was revealed to be shared by the differently sourced sequences. Furthermore, consensus comparison of all clone sequences revealed 13 differential sites between the two major clades, which are still strongly conserved in sequences from the gene bank. A few recombination events were detected between individual variants, and the supposed recombinants displayed distinct phylogenetic divergence from those in the same clade.

Conclusion: Highly varied D-loop sequences were detected in the individual yaks. The detected variants were clustered into two deeply divergent clades as the haplotypes reported by other researchers employing PCR-product based D-loop sequences. This suggests either or both of heteroplasmy and nuclear-mitochondrial DNA in yaks, and also emphasizes the need for reconsidering D-loop sequences as haploid and matrilineal markers for phylogenetic analysis in yaks.

Background

Yak (*Bos grunniens*) is an unusual animal belonging to Bovinae Bos [1, 2]. Commonly, yak is discriminated as domesticated yak and wild yak (*Bos grunniens mutus*). Yaks are distributed across the Qinghai-Tibetan Plateau (QTP) and neighboring highlands, and are the only cattle that can survive in the extreme conditions of the QTP. The QTP regions of China host the greatest abundance and diversity of yak, including about 17 million domesticated yak [3] and 150 thousand wild yaks in northern Tibet and western Qinghai [4, 5]. Domesticated yaks provide the most important living resources such as food, hides, dung fuel and transport power for the local husbandry communities. Domestic history, intraspecific diversification and geographical distribution of yak lineages are of great research interest due to yak's biological and economic importance [5-11].

Mitochondrial DNA (mtDNA) is independent from nuclear DNA. Owing to their relatively simple genome structure, semi-autonomous replication, maternal heritage, and high mutation rate, the genes of mtDNA

are ideal molecular markers in phylogenetic research [12, 13]. The displacement loop region (D-loop) of mtDNA rich has the highest variation in length and sequence in the mtDNA genome. D-loop variability is universally used for research in animal phylogeography, intraspecific diversification, etc. [14]. D-loop variations in yak have been widely used for studying genetic diversity, geographic distribution and phylogenetic analysis [8, 11, 15-18].

Heteroplasmy is defined as a state in which more than one mitochondrial genotype exists in one organism. Commonly, the mitochondrial genome is regarded as maternal and haploid, but a few studies have suggested that heteroplasmy may be a common characteristic of mitochondrial genetic information in plants and animals [19-22]. Moreover, nuclear sequences of mitochondrial DNA copies (NUMTs), which comprise mitochondrial DNA integrated into the nuclear genome, were reported as a well-established phenomenon in eukaryotes [23-25]. Conventionally, D-loop variations of yak have been based on the direct sequencing of products derived from the polymerase chain reaction (PCR) using total genomic DNA as the template [8, 11, 15-18]. Thus, the possibility of heteroplasmy or NUMTs in yak may influence analysis of the variations in mitochondrial DNA. In this study, we demonstrated highly variable D-loop sequences in the individual yak genome by sequencing the D-loop PCR production cloning pool. The results suggest that heteroplasmy or NUMTs are prevalent in yak.

Methods

Experimental materials

Materials for the two breeds of domesticated yak were collected from the Qinghai LaoZhaxi Organic Agriculture and Animal Husbandry Technology Co., LTD, we called them Y1 and Y2. The bloods were individually collected in vivo, all animals were treated in accordance with good animal practice as defined by the local welfare authorities.

DNA extraction and amplification

The total genomic DNAs were extracted using a blood/tissue extraction kit (TIANamp Genomic DNA Kit (Tiangen Biotech (Beijing) Co., LTD. China). The mitochondrial D-loop sequence was amplified using primers D-loop-F 5'-GTAAAGAGCCTCACAGTAT-3' and D-loop-R 5'-TCCTGTAGCCATTGACTAT-3' following Guo et al [8]. The polymerase chain reaction (PCR) mixture of 50 ul contained template DNA 2 ul (50 ng/ul), 1.0 ul of each primer (10 ng/ul solution), 5 ul dNTPs (2 mmol/L solution), 5 ul 10*buffer, 0.5 ul Taq DNA polymerase (5 U/ul), and 35.5 ul ddH₂O. The PCR procedure used an initial denaturing step at 94°C for 1 min, 35 amplification cycles (94°C for 30 s, 49°C for 30 s and 72°C for 30 s) and a final extension at 72°C for 10 min in a Veriti Thermal Cycler (Applied Biosystem).

Library construction of D-loop PCR production and clone sequencing

The target PCR fragments of D-loop sequences were recycled and purified using a PCR purification kit following the manufacturer's instructions (Tiangen Biotech (Beijing) Co., LTD. China). The purified Cot-1

DNA fragments were ligated into the pGEM-T easy vector (Promega). The transformed *E. coli* cells (DH5 α) were identified by a blue/white screen following the manufacturer's instructions (Promega). White recombinant colonies of 20~40 individuals were randomly selected for Sanger sequencing (Sangon Biotech (Shanghai) Co., LTD. China).

Sequence analysis

The sequence chromatograms were checked by Chromas (version 2.6.6, Technelysium Pty Ltd, South Brisbane, Queensland, Australia). The alignment, base composition, and divergence of the sequences were analyzed by MEGA 6.0 [37]. The phylogenetic trees were constructed by neighbor-joining in the Kimura 2-parameter model [37]. Recombination between the D-loop variants was detected by RDP4 [27].

Results

Primary analysis of mono clone D-loop sequences.

Sanger sequencing was conducted on randomly selected 10 D-loop PCR production mono clones. The sequences alignment was carried out by Clust W in Mega 6.0. Unexpectedly, the obtained mono clone sequences did not show a 100% identity match with each other. The sequencing chromatograms were checked by Chromas, showing clean peaks and valleys in each sequence. No apparent heterozygous peaks were detected (Fig.1). This suggests that variation in D-loop sequences between the mono clones are less likely to have occurred during the sequencing procedure.

Sequencing of a low copy gene and D-loop PCR products

Since a series of experimental manipulation techniques such as DNA extraction, PCR amplification, DNA ligating, transformation, etc. was carried out before sequencing, the clone sequence mutation may have occurred artificially during the experimental procedure. Therefore, two additional controlled experiments were carried out by sequencing clones of a nuclear single copy gene α 1-globin in yak and then directly sequencing the D-loop PCR fragments. The clones of yak α 1-globin were prepared following the same procedure as the those of the D-loops, and PCR was conducted following Yuan et al. [26]. Sixteen clones were successfully sequenced, with a target length 706bp. Alignment of the sequences showed that 12 clones shared one identical haplotype, while 4 clones shared another. The first haplotype showed 100% homology with the α 1-globin sequence described by Yuan [26]. BLAST (Basic Local Alignment Search Tool) searching against the NCBI gene bank yielded a 100% query and identity with the sequences of 5 chromosomes of wild yak ([CP027093.1](#)). The second haplotype showed three mutated sites when compared with the first type, among which the mutation sites 254bp and 316bp were strongly consistent with mutation sites of the α 2-globin identified by Yuan [26], while site 263bp is a synonymous mutation. Additional sequencing was conducted on the PCR products of the D-loop. Although D-loop sequences were obtained by base calling, checking of the chromatogram demonstrated high noise. In particular, distinct heterozygous peaks were detected along the sequences, strongly suggestive of heterogeneity in the D-loop PCR product. The chromatogram for the randomly-selected 125bp-255bp region revealed three

apparent heterozygous peaks at sites 133bp, 146bp, and 171bp which showed C/T, G/A, and G/A respectively (Fig. 2). The alignment between clone sequences uncovered about 10 mutations between 125bp-255bp. Most mutations were detected in only sequenced clone, except those at sites 133bp and 171bp, the alternative bases of which were detected in more than one third of the total clones. By coincidence, the two alternative bases in 133bp and 177bp in the clone sequences were the same as the two heterozygous bases at 133bp and 177bp in the D-loop PCR product sequences (Fig. 1). Even though the heterozygous peaks were apparent at 146bp in the PCR product sequence, no mutation of this site was detected in any clones. This indicates that most of the highly frequent clone sequence mutations can also be detected by checking the heterozygous peaks of the PCR products sequences. The reliably stable sequences of the single copy gene and high frequency of mutations revealed both in clone sequences and PCR product sequences strongly suggest that the highly-variable D-loop clone sequences were obtained by an appropriated procedure.

Analysis of the D-loop sequences of clones from individual breeds

Totally, 25 clones from an individual yak (Y1) were successfully sequenced. The middle regions of the D-loop sequence (with a length of about 607bp) were truncated for further analysis to ensure the accuracy of sequences. The total clone sequences had lengths varying from 605bp to 607bp, with an average of 606.4bp and T, C, A and G contents of 31.2%, 15.8%, 27.9% and 25.0% respectively. Thirty-three Parsim-Informative sites and 30 singletons were analyzed between the sequences. The mean distance between sequences was 99.99%. Twenty-five clone sequences could be ascribed to 22 variants based on similarity, we called them Y1-V1 to Y2-V22. BLAST searching against the NCBI gene bank showed a homology of 98.5~100% with the putative D-loop sequences of the yak. One D-loop variant (Y1-V9), which was commonly shared by 3 clones, showed 100% homology with an accession (DQ139035.1) from the gene bank. Phylogenetic analysis based on neighbor-joining clearly divided all 22 variants into two clusters with a bootstrap value of 100. Cluster 1 included 13 variants of 15 clones, while Cluster 2 included 9 variants of 10 clones (Fig. 3).

To validate variations of the D-loop, we sequenced D-loop clones from another individual yak (Y2). In total, 38 clone sequences were obtained. All the sequences showed similar base compositions to those in Y1. The variable sites were identified as 26 Parsim-Informative sites and 45 singleton sites. Twenty-seven variants were identified by sequence identity in 38 clones, we called them Y2-V1 to Y2-V27. One variant (Y2-V14) was commonly shared by 11 clones, which showed 100% query coverage and identity with 15 D-loop accessions from the NCBI gene bank. Phylogenetic analysis revealed two distinct clusters, which contained 9 variants of 9 clones in Cluster 1 and 18 variants of 28 clones in Cluster 2 (Fig. 4).

Analysis of D-loop sequences from the gene bank

Seventeen and 10 sequences originated from domesticated yaks and wild yaks, respectively, were randomly downloaded from the NCBI gene bank. The same truncated regions as those in yak individuals Y1 and Y2 were used for phylogenetic analysis. The results showed 8 D-loop variants in 17 sequences from domesticated yaks. Eight variants could be divided into two distinct clusters, which contain 6 and 2

variants in each (Fig. 5). Eight variants were uncovered in 10 sequences from wild yaks. Similarly, the variants were distinctly divided into two clusters containing 3 and 5 variants in each (Fig. 6).

Analysis of the integrative D-loop sequences

The downloaded D-loop sequences originated from different yak individuals, and were reported as being obtained by PCR product sequencing. The varied D-loop sequences are conventionally regarded as representing different mitochondrial haplotypes. Both the D-loop clone variants and downloaded D-loop haplotypes were distinctly clustered as two clades. It was expected that the D-loop haplotypes obtained by PCR product sequencing could represent one D-loop variant, due to the biased amplification. Clustering was conducted on the integrative D-loop sequences including clone variants and haplotypes. The results showed that all sourced sequences could be distinctly divided as two clusters. Cluster 1 includes 13 variants of Y1, 9 variants of Y2, 6 haplotypes of domesticated yaks, and 3 haplotypes of wild yaks. Cluster 2 includes 9 variants of Y1, 18 variants of Y2, 2 haplotypes of domesticated yaks, and 5 haplotypes of wild yaks (Fig. 7). As expected, the variants and haplotypes in clusters 1 and 2 were grouped into the respective clusters found by the previous analysis (Fig. 3, 4, 5, and 6). Meanwhile, the clustering revealed an identical sequence commonly shared by variants of Y1-V9 and Y2-V24, and haplotypes of KY807492.1 and FJ548841.1.

Analysis of the consistent differential sites between clusters 1 and 2 of the D-loop sequences

The consistent differential sites of the D-loop sequences were analyzed between the 22 variants of Cluster 1 and 27 variants of Cluster 2 of the clone sequences derived from individuals Y1 and Y2 by alignment of Mega 6. The results showed 13 consistent differential sites between clusters 1 and 2 (Fig. 8). The consistent differential sites are distributed in the regions of 280bp-423bp (Table 1). Further, analysis of the downloaded sequences of the domesticated yaks showed that the 13 sites in clusters 1 and 2 are strongly conserved, with single nucleotide polymorphisms (SNP) at 281bp and 423bp of Cluster 1, and with insertion/deletion (Indel) polymorphisms at 338bp and distinct mutation at 422bp of Cluster 2 (Table 1). Analysis of the wild yaks showed that 13 differential sites in Cluster 1 are completely identical to those in the clone sequences, and most of the differential sites are conserved in Cluster 2 with Indel detected at 280bp, 316bp, 338bp, and 422bp (Table 1).

Table 1 The common differentiated sites between Cluster 1 and Cluster 2 of D-loop sequences.

Cluster	1			2			
Origin	Clones	Domesticated Yaks (gene bank)	Wild yaks (gene bank)	Clones	Domesticated Yaks (gene bank)	Wild yaks (gene bank)	
Sequence position	280	C	C	C	T	T	T/-
bp	281	C	C/T	C	T	T	T
	316	G	G	G	A	A	A/-
	338	G	G	G	A	A/-	A/-
	346	A	A	A	G	G	G
	355	T	T	T	C	C	C
	362	A	A	A	G	G	G
	363	G	G	G	A	A	G
	379	T	T	T	C	C	C
	396	T	T	T	C	C	C
	416	C	C	C	T	T	T
	422	G	G	G	A	G	A/-
	423	A	A/G	A	G	G	G

Predicting the recombination between different variants

Predicting the recombination between D-loop variants was conducted by RDP4 [27]. Two recombination events, involving Y1-V10 and Y1-V11 in Event 1 and Y1-V15 and Y1-V16 in Event 2, were identified in D-loop variants of individual yak Y1. Y1-V10 was predicted as a distinct recombinant between Y1-V6 and an unknown parent by two models. Y1-V16 was predicted as a distinct recombinant between Y1-V18 and Y1-V6 by three models. One recombination event, involving Y2-V15, Y2-V16 and Y2-V17, was identified in individual yak Y2. Y2-17 was suggested as a distinct recombinant between Y2-V1 and an unknown parent by three models. Integrative analysis of the total variants in both individuals revealed two recombination events, which involved Y1-V15, Y1-V16 and Y1-V18 in Event 1 and Y1-V10, Y1-V11, Y2-V19, Y2-V20 and Y2-V21 in Event 2. Y1-16 was suggested as a distinct recombinant between Y2-V18 and an unknown parent by five models, whereas Y1-V10 was suggested as a recombinant between Y1-V6 and Y2-V18 parent by four models (Table 2).

In total, 11 D-loop variants were identified as being involved in recombination events respective or integrative of Y1 and Y2. Retrospective inspection of the phylogenetic status of those variants in the phylogenetic trees showed that most of the recombinations involved variants that are distinctly divergent

with the others in the same clade (Fig. 3, Fig. 4 and Fig. 7). This suggests that the D-loop sequence differentiation may be promoted by the recombination between variants, especially by the recombination between variants belonging to the different clusters.

Table 2 Prediction of recombination between variants of clones in Y1 and Y2.

	Y1		Y2		Y1+Y2	
Event Number	1	2	1	1	2	
Found in	2	3	3	5	4	
Recombinants	Y1-V10	Y1-V16	Y2-V17	Y1-V16	Y1-V10	
Major parent	Y1-V6	Y1-V18	Y2-V1	Y2-V18	Y1-V6	
Minor parent	unknown	Y1-V6	unknown	unknown	Y2-V18	
Detection methods	R	□	□	□	□	□
	G	□	□	□	□	□
	B	□	□	□	□	
	M	□	□	□	□	□
	C	□	□	□	□	□
	S	□	□	□	□	□
	T	□		□	□	+

Discussion

Validation of the D-loop variation

The high frequency of D-loop variants uncovered in individual yaks was unexpected, while mitochondrial DNA is commonly regarded as having matrilineal and haploid inheritance. Thus, it is very important to exclude sequencing errors caused by the instrument, chemicals, or base calling. In this study, a single nuclear gene was cloned and sequenced using the same experimental system; identical and conserved α -globin sequences were detected in the repeated clones. Direct sequencing the PCR product revealed heterogeneity and a few alternative bases in base calling which were also detected in respective clone sequences. This suggests a low probability of artifacts from PCR amplification and the cloning procedure. Further, alignment showed 100% query coverage and identity between a few cloning and documented sequences, and highly consistent differential sites both in cloning and documented sequences. Sanger sequencing is regarded as the golden standard in sequencing, reaching 99.9% accuracy for 400-900bp length DNA fragments [28]. In this study, unique inserts in mono clones of about 1000bp were used for Sanger sequencing. About 600bp in the middle part of the sequences, which showed highly

distinct chromatogram peaks, were truncated for sequence analysis. The above evidence suggests that the sequence variations revealed in this study are not systematic or sequencing artifacts.

Heteroplasmy or NUMTs in yaks

Heteroplasmy may commonly characterize mitochondrial genetic information in plants and animals, and is the result of mitochondrial DNA recombination, small scale mutation, and paternal leakage [20]. Single point mutations or multipoint mutations, and recombination, were detected between the D-loop variants in this study. However, whether or not there is paternal leakage in yaks still needs further detailed investigation. In animals, heteroplasmy is reported to be associated with disease, aging, and family of origin [22, 29, 30]. Although many variants were detected in individuals, only one variant was commonly shared between two different samples. Since the family origin and physiological conditions of the two yaks were unclear, the origins of individual differences were uncertain. The ratio of divergent types of mtDNAs in a heteroplasmic population may be variable among siblings, between tissues and during the life span of animals and humans [31, 32], but usually one mitotype is prevalent, while the alternative one(s) are present in very low proportions [20]. In this study, two weak dominant variants (Y1-V9 and Y1-V22) with respective occupations of 12% and 8% in total clones were detected in Cluster 1 and Cluster 2 in yak Y1, while one dominant variant (Y2-V14) with an occupation of 28.9% was detected in Cluster 2 in yak Y2. The types of dominant variants and their ratio varied between the two individuals, and the ratio of dominant variants did not show a high advantage over the others.

NUMTs, nuclear sequences of mitochondrial origin, are present with variable sizes and are highly fragmented, rearranged within nuclear genomes with different degrees of homology to their mtDNA sequence fragments [33]. NUMTs can include more than one mitochondrial variant [34]. A yak reference genome is available (BosGru_v2.0 reference Annotation Release 101, NCBI). BLAST searching against the genome using D-loop sequences showed 100% query coverage and 99.34% identity with the complete mitochondrial genome, and 44%-85% query coverage and 83.2%-94.1% identity with another four unplaced genomic scaffolds. Complex concatenated mtDNA-derived sequences are possibly rearranged within the nuclear genome [24]. Repetitive distribution patterns of mitochondrial sequences in chromosomes were detected by fluorescent in-situ hybridization in sheep [35]. The high frequency of single nucleotide mutation between variants, and high variation between individuals, reflects the possibility of the repetitive distribution of D-loop sequences in nuclear genome, and that the failure to search the highly homologous D-loop sequences in Yak chromosomes may be attributed to genome assembly algorithms for repetitive sequences.

It was estimated that the two main distinct lineages separated at 100,000 years ago [8] or around 429,183 years ago [11]. This suggests the heteroplasmy may arise from paternal leakage by hybridization between two lineages, or NUMTs from mitochondrial DNA transfer by an ancient accident. The divergent time is consistent with the geological records of glaciation events experienced in the QTP [11]. Given the role of mitochondria in metabolism, variation of mitochondrial genes has been associated with phenotypes including hardiness [35]. Yaks living in the QTP are highly adaptive to extremely harsh

conditions. The relationship between variations of mitochondrial genes and environmental adaptation in yaks still needs further clarification.

Phylogenetic analysis using D-loops as maternal markers

Considering the common maternal inheritance of mitochondrial DNA, the diversity of D-loop sequences has been used to study the phylogenetic relationships in yaks correlated with geographic locations or to resolve questions regarding origin and ancestry [8, 16, 18, 36]. The variants of D-Loop sequences were repeatedly and distinctly clustered as two clades in the investigated populations of yaks in a few studies [8, 11, 16, 18, 36]. However, in this study, a few D-loop variants were uncovered in single individual, and were distinctly clustered as two clades consistent with the clustering of the other variants from the different individuals and that of the documented variants from the gene bank. The D-loop sequences for phylogenetic analysis are usually obtained by directly sequencing the PCR products, with only one representative D-loop sequence used for a single individual. The mixture of D-loop PCR products revealed in this study suggests that biased sequences can be obtained when sequencing the PCR products, and that the diversity of the mitochondrial DNA of yaks may be underestimated. Therefore, it is inferred that the two clusters of mitochondrial haplotypes identified by previous studies are primarily due to random amplification of the variants which are distinctly clustered in the genome of the single individual. The phylogenetic relationships in yaks, revealed using D-loop sequences as being of the maternal and haploid types, need to be reappraised.

Conclusion

Clone sequencing showed substantial variants of mitochondrial D-loop sequences in different yak individuals. Clone sequencing of the nuclear single copy gene and PCR-product based sequencing of D-loop sequences validated the variations of D-loop sequences in individuals. Phylogenetic analysis indicated the two deeply divergent clusters of variants identified in cloning sequencing, and also the two main clusters of the haplotypes randomly selected from documented sequences obtained by PCR-product based sequencing. Comprehensive analysis then demonstrated the consistency of the two distinct clusters in both cloned and documented D-loop sequences. Alignment revealed 13 consistently differentiated sites between the Cluster 1 and Cluster 2 sequences. Furthermore, A few recombination events were detected between different variants. The results strongly suggestive of heteroplasmy or nuclear-mitochondrial DNA, or both, in yaks. This emphasizes the need for considering NUMTs or heteroplasmy in phylogenetic analysis, and also highlights the need to improve the genome assembly of mitochondrial-related DNA sequences.

Abbreviations

DNA: deoxyribonucleic acid; PCR: polymerase chainreaction; QTP: Qinghai-Tibetan Plateau; mtDNA: Mitochondrial DNA; D-loop: Displacement loop region; NUMTs: Nuclear sequences of mitochondrial DNA copies; NCBI: National Center of Biotechnology Information; T: Thymine; G: Guanine; C: Cytosine; A:

Adenine; SNP: single nucleotide polymorphisms; BLAST: Basic Local Alignment Search Tool; NJ: neighbor-joining; Co., LTD: Company limited

Declarations

Ethics approval and consent to participate

The study was approved by animal ethic and welfare committee at Northwest Institute of Plateau Biology, Chinese Academy of Sciences. All participants have signed an Informed Consent form.

Consent for publication

Not applicable.

Availability of data and materials

The datasets supporting the conclusions of this article are included within the article.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the National Key Research and Development Program of China (2018YFD0502301), and the Major Science and Technology Project of Qinghai (2018-NK-A2). The funder had no role in the design of the study or preparation of the manuscript.

Authors' contributions

QY and QD conceived of the study, and participated in its design and coordination and helped to draft the manuscript. LB and XL carried out the molecular biology experiment, performed the data analysis and drafted the manuscript. HW participated in the design of study and the language correction. XT took part in the Sample collection. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

Authors' information

¹ Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining 810008, China.

² University of Chinese Academy of Sciences, Beijing 100049, China.

References

1. Linnaeus C. *Systema Naturae*. 12rd ed. Holmiae: Impensis direct; 1766.
2. Bohlken H. Haustiere und Zoologische Systematik. *Z für Tierzuchtug Zuchtungsbiol.* 1961;76(1-4):107-13.
3. Weiner G, Han JL, Long RJ. Yak. In: Pond W, Ullrey D, Baer C, editorts. *Encyclopedia of Animal Science*, 2nd ed. Boca Raton: CRC Press; 2011. p. 1121-24.
4. Schaller GB, Liu WL. Distribution, status, and conservation of wild yak *Bos grunniens*. *Biol Conserv.* 1996;76(1):1-8.
5. David M, Schaller GB. *Bos grunniens* and *Bos mutus* (Artiodactyla: Bovidae). *Mamm Species.* 2009;836:1-17.
6. Tu ZC, Qiu H, Zhang YP. Polymorphism in Mitochondrial DNA (mtDNA) of Yak (*Bos grunniens*). *Biochem Genet.* 2002;40(5/6):187-93.
7. Q XB, H JL, Lkhagva B, Chekarova I, Badamdorj D, Rege JEO, et al. Genetic diversity and differentiation of Mongolian and Russian yak populations. *J Anim Breed Genet.* 2005;122(2):117-26.
8. Guo S, Savolainen P, Su J, Zhang Q, Qi D, Zhou J, et al. Origin of mitochondrial DNA diversity of domestic yaks. *BMC Evol Biol.* 2006;6:73.
9. Lai SJ, Chen SY, Liu YP, Yao YG. Mitochondrial DNA sequence diversity and origin of Chinese domestic yak. *Anim Genet.* 2007;38(1):77-80.
10. Ho SY, Larson G, Edwards CJ, Heupink TH, Lakin KE, Holland PW, et al. Correlating Bayesian date estimates with climatic events and domestication using a bovine case study. *Biol Lett.* 2008;4(4):370-4.
11. Wang ZF, Shen X, Liu B, Su JP, Yonezawa T, Yu Y, et al. Phylogeographical analyses of domestic and wild yaks based on mitochondrial DNA: new data and reappraisal. *J Biogeogr.* 2010;37(12):2332-44.
12. Brown WM, Jr MG, Wilson AC. Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci U S A.* 1979;76(4):1967-71.
13. Boore JL. Animal mitochondrial genomes. *Nucleic Acids Res.* 1999;27(8):1767-80.
14. Gilbert MT, Drautz DI, Lesk AM, Ho SY, Qi J, Ratan A, et al. Intraspecific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. *Proc Natl Acad Sci U S A.* 2008;105(24):8327-32.
15. Guo SC, Liu JQ, Qi DL, Yang J, Zhao XQ. Taxonomic placement and origin of yaks: implications from analyses of mtDNA D-loop fragment sequences. *Acta Theriol Sin.* 2006;26(4):325-30. (in Chinese)
16. Guo SC, Qi DL, Chen GH, Xu SX, Zhao XQ. Genetic diversity and classification of domestic yaks inferred from mitochondrial DNA sequences. *Acta Ecol Sin.* 2008;28(9):4286-94. (in Chinese)
17. Song QQ, Zhong JC, Zhang CF, Xin JW, Ji QM, Chai ZX. Analysis on genetic diversity and phyletic evolution of mitochondrial DNA from Tibetan yaks. *Acta Theriol Sin.* 2014;34(4):356-65. (in Chinese)

18. Ma ZJ, Zhong JC, Han JL, Xu JT, Bai WL, Li JY, et al. Genetic diversity and demographic history of wild Yak (*Bos grunniens mutus*) inferred from mtDNA D-loop sequences. *Afr J Biotechnol.* 2010;9(46):7805-10.
19. Barr CM, Neiman M, Taylor DR. Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. *New Phytol.* 2005;168(1):39-50.
20. Kmiec B, Woloszynska M, Janska H. Heteroplasmy as a common state of mitochondrial genetic information in plants and animals. *Curr Genet.* 2006;50(3):149-59.
21. Gonzalez MDM, Ramos A, Aluja MP, Santos C. Sensitivity of mitochondrial DNA heteroplasmy detection using Next Generation Sequencing. *Mitochondrion.* 2020;50:88-93.
22. Polovina ES, Parakatselaki ME, Ladoukakis ED. Paternal leakage of mitochondrial DNA and maternal inheritance of heteroplasmy in *Drosophila* hybrids. *Sci Rep.* 2020;10:2599.
23. Hazkani-Covo E, Zeller RM, Martin W. Molecular poltergeists: mitochondrial DNA copies (numts) in sequenced nuclear genomes. *PLoS Genet.* 2010;6(2):e1000834.
24. Balciuniene J, Balciunas D. A Nuclear mtDNA Concatemer (Mega-NUMT) Could Mimic Paternal Inheritance of Mitochondrial Genome. *Front Genet.* 2019;10:518.
25. Yoshida T, Furihata HY, Kawabe A. Analysis of nuclear mitochondrial DNAs and factors affecting patterns of integration in plant species. *Genes Genet Syst.* 2017;92(1):27-33.
26. Yuan QY, Huang ZG, Xie Z, Yin FL, Zhao YH. Cloning and sequence analysis of α 1-globin gene of yak. *J Nanjing Agric Univ.* 2006;29(2):134-37. (in Chinese)
27. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* 2015;1(1):vev003.
28. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics.* 2012;13:341.
29. Zeviani M, Antozzi C. Mitochondrial disorders. *Mol Hum Reprod.* 1997;3(2):133-48.
30. Szibor M, Holtz J. Mitochondrial ageing. *Basic Res Cardiol.* 2003;98(4):210-8.
31. Sekiguchi K, Kasai K, Levin BC. Inter- and intragenerational transmission of a human mitochondrial DNA heteroplasmy among 13 maternally-related individuals and differences between and within tissues in two family members. *Mitochondrion.* 2003;2(6):401-14.
32. Jenuth JP, Peterson AC, Shoubridge EA. Tissue-specific selection for different mtDNA genotypes in heteroplasmic mice. *Nature Genet.* 1997;16(1):93-5.
33. Woischnik M, Moraes CT. Pattern of Organization of Human Mitochondrial Pseudogenes in the Nuclear Genome. *Genome Res.* 2002;12:885-93.
34. Zhang DX, Hewitt GM. Nuclear integrations: challenges for mitochondrial DNA markers. *Trends Ecol Evol.* 1996;11(6):247-51.
35. Mustafa SI, Schwarzacher T, Heslop-Harrison JS. Complete mitogenomes from Kurdistani sheep: abundant centromeric nuclear copies representing diverse ancestors. *Mitochondrial DNA A.*

2018;29(8):1180-93.

36. Zhang CF, Xu LJ, Ji QM, Xin JW, Zhong JC. Genetic diversity and evolution relationship on mtDNA D-loop in Tibetan yaks. *Acta Ecol Sin.* 2012;32(5):9. (in Chinese)
37. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol.* 2013;30(12):2725-9.

Figures

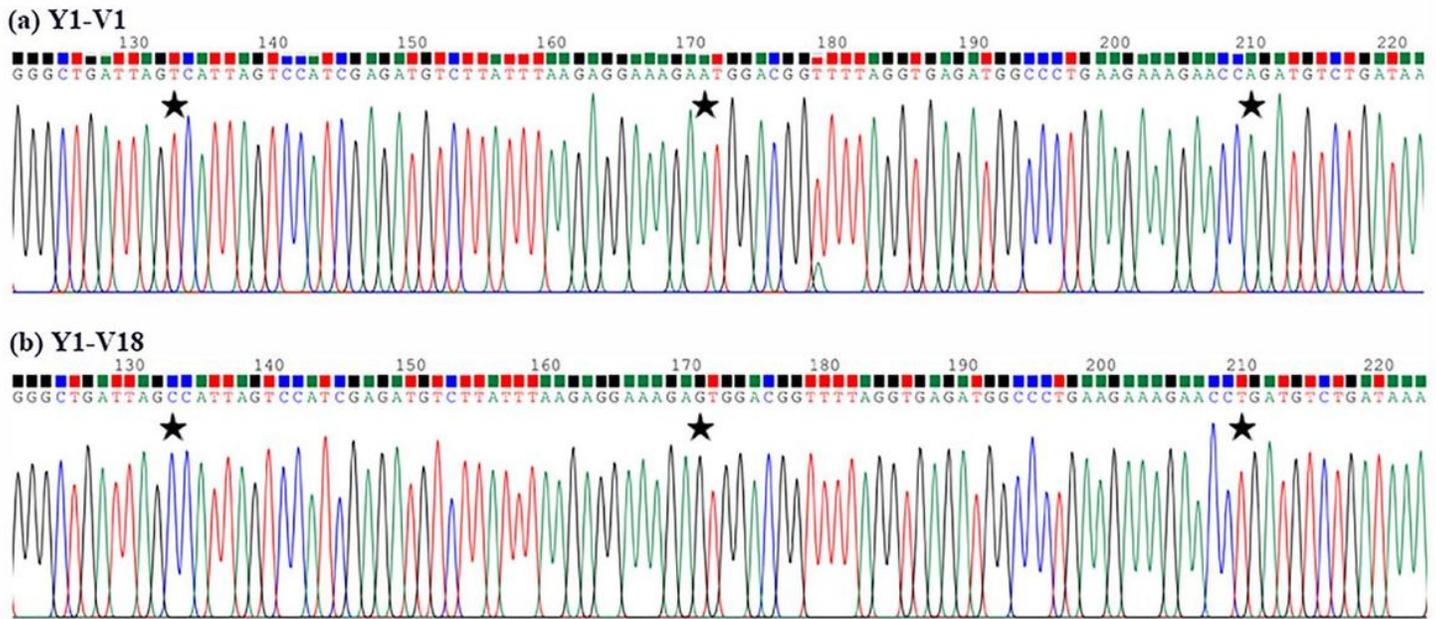


Figure 1

a Chromatogram of Sanger sequencing for the PCR clones of Y1-V1. b Chromatogram of Sanger sequencing for the PCR clones of Y1-V18. The stars symbol showing polymorphisms in base calls at the same positions between clones.

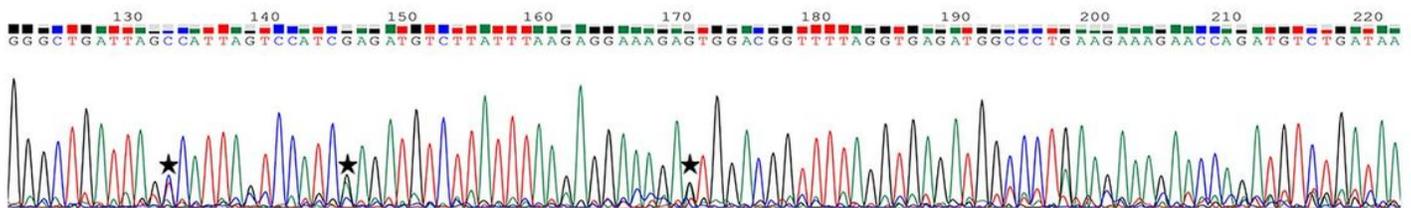


Figure 2

Chromatogram of Sanger sequencing of the PCR products spanning the same region as that in Fig. 1. The stars symbol showing heterogeneity in base calls.

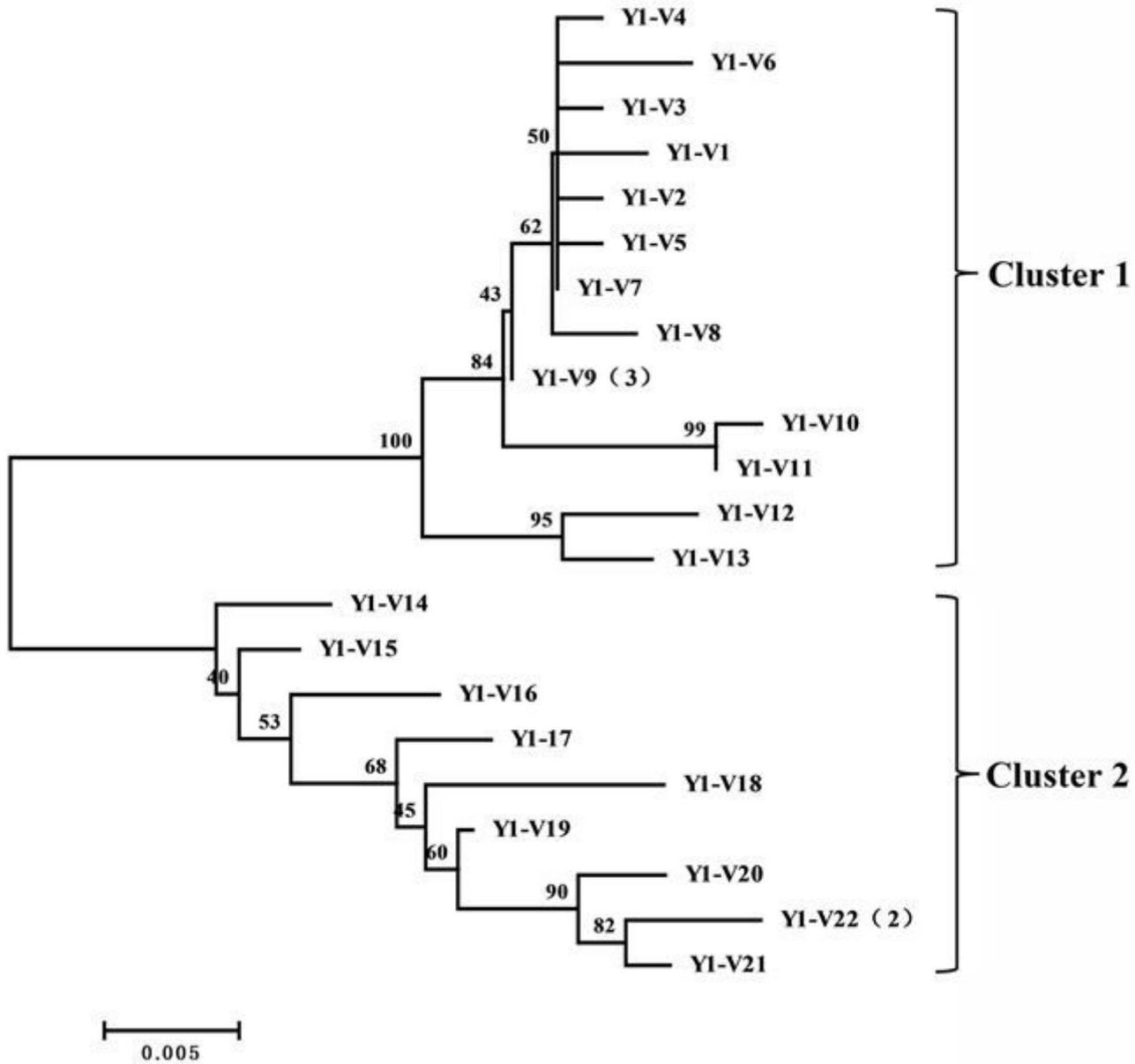


Figure 3

Phylogenetic tree of variants of cloned D-loop sequences in a yak (Y1) constructed by neighbor-joining (NJ). The number in parentheses indicates the quantity of clones.



Figure 4

Phylogenetic tree of variants of cloned D-loop sequences in another yak (Y2) constructed by neighbor-joining (NJ). The number in parentheses indicates the quantity of clones.

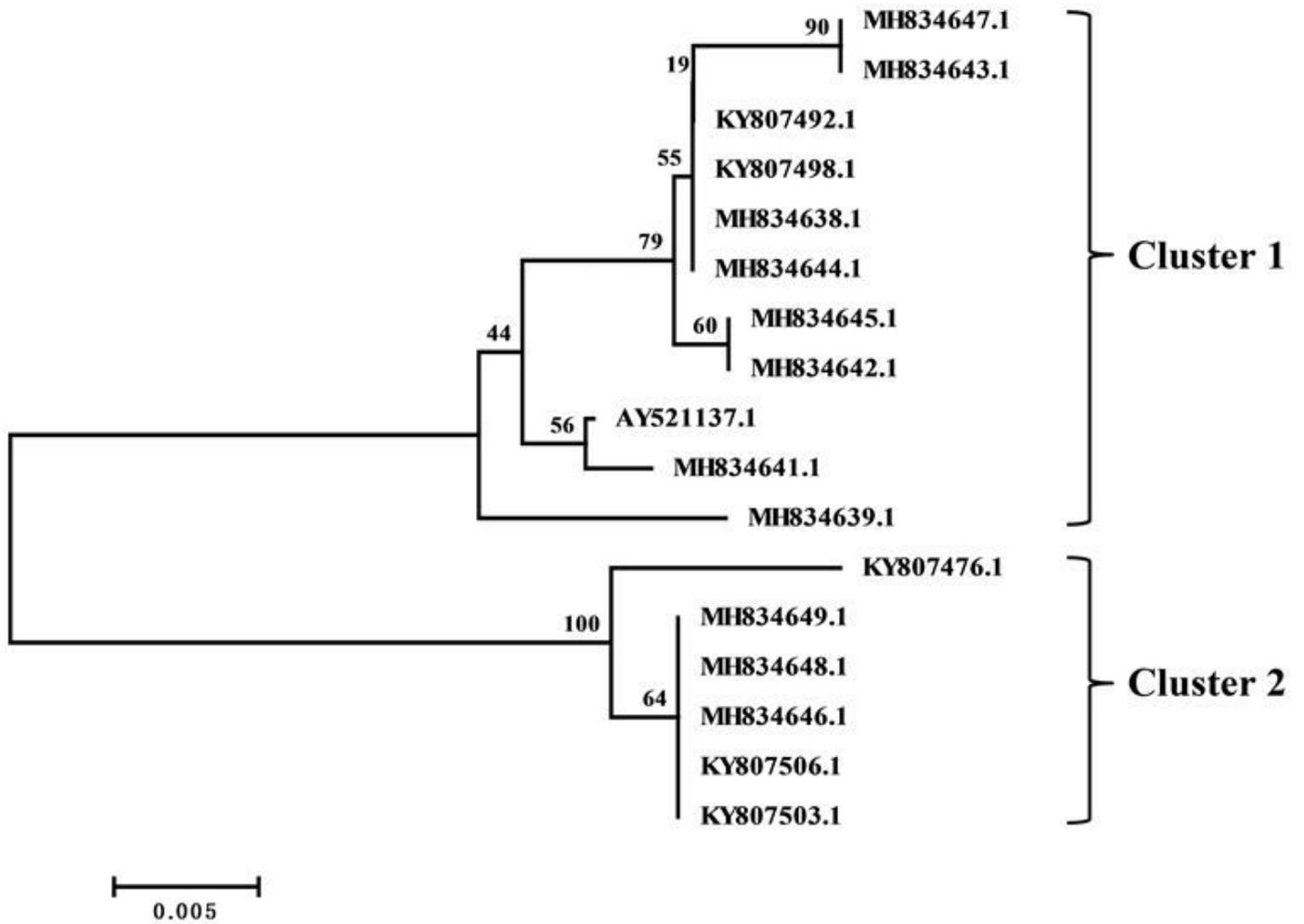


Figure 5

Phylogenetic tree of haplotypes of D-loop sequences from the gene bank in domesticated yaks constructed by neighbor-joining (NJ).

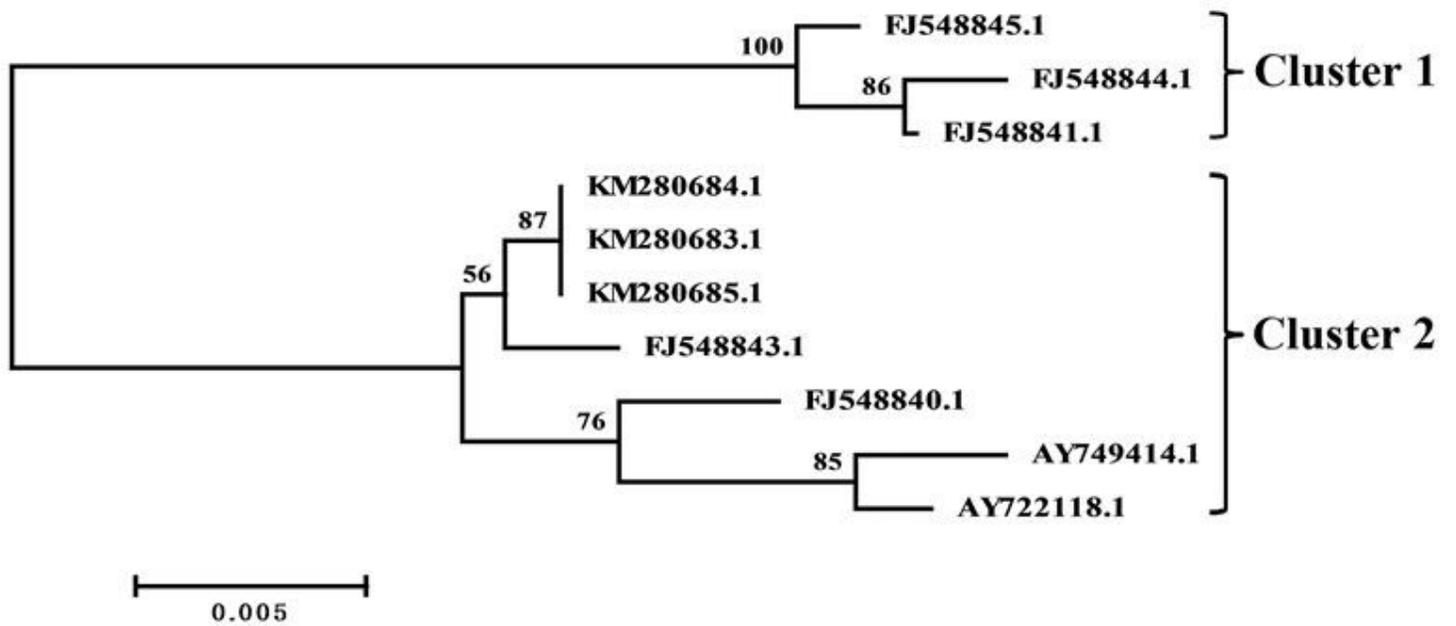


Figure 6

Phylogenetic tree of haplotypes of D-loop sequences from the gene bank in wild yaks constructed by neighbor-joining (NJ).

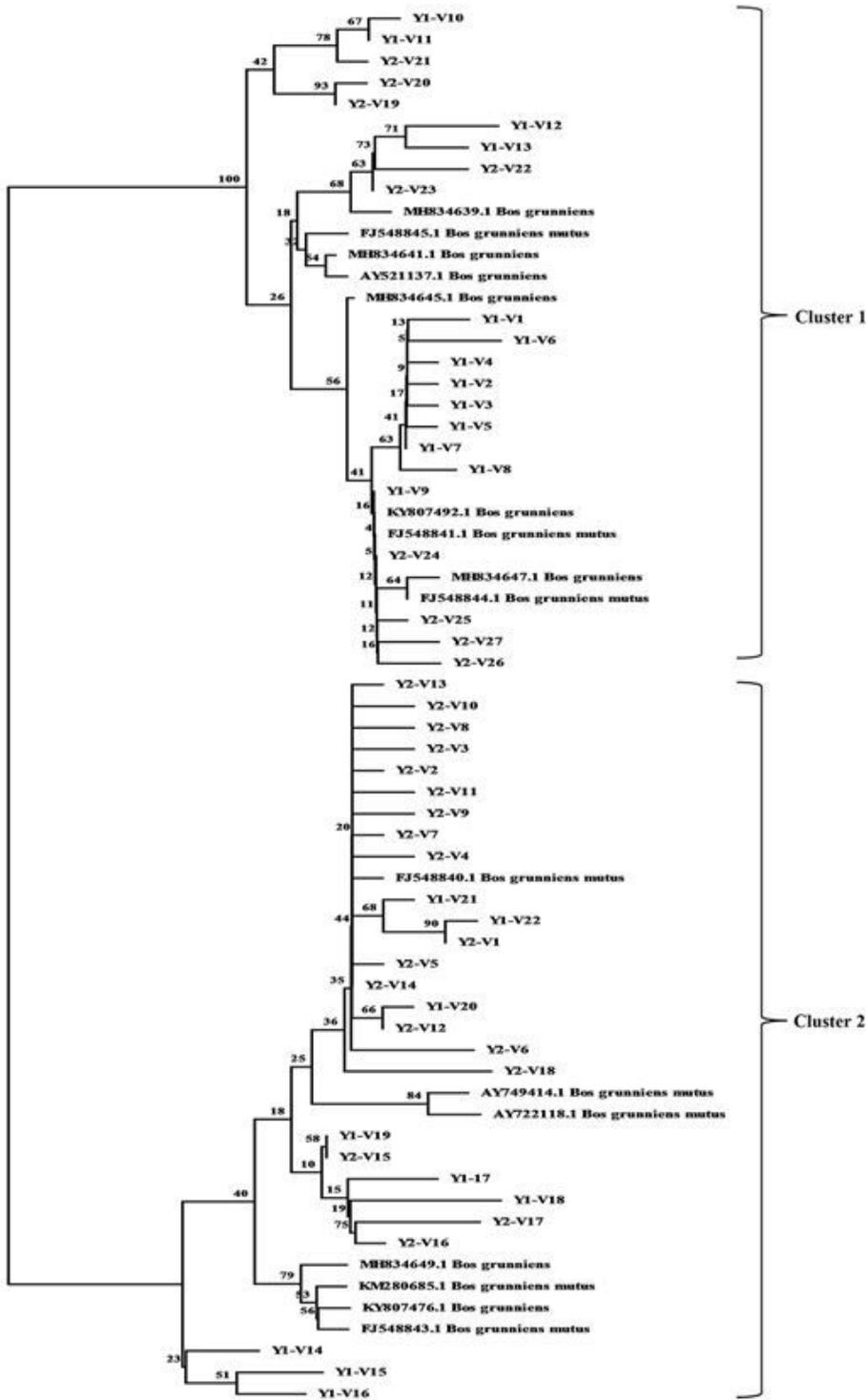


Figure 7

Phylogenetic tree of integrative variants and haplotypes of D-loop sequences of yaks constructed by neighbor-joining (NJ).

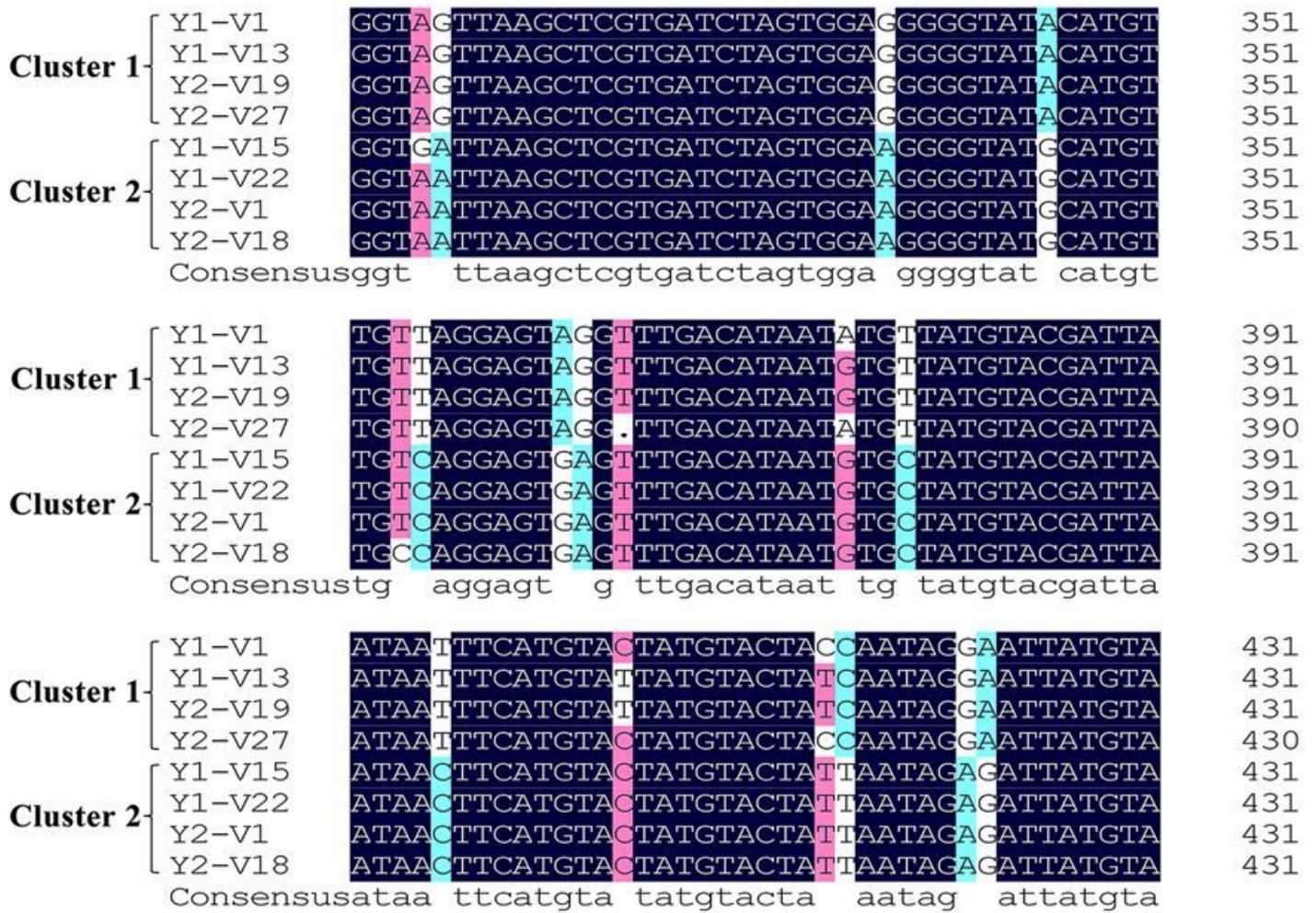


Figure 8

Representative common differentiated sites between Cluster 1 and Cluster 2 of Yak mitochondrial D-loop sequences (from 312bp-431bp).