Available online at sjuoz.uoz.edu.krd

journals.uoz.edu.krd

Science Journal of University of Zakho

Vol. 12, No.2, pp.176–188 April-June, 2024

p-ISSN: 2663-628X
e-ISSN: 2663-6298

BIOINFORMATICS AND MOLECULAR ANALYSES IDENTIFIED THE CONTROL REGION AS THE MOST POWERFUL MITOGENOMIC MARKER FOR DISTINGUISHING THE MAIN MATERNAL HAPLOGROUPS IN GOATS

Sarbast Ihsan Mustafa
Department of Animal Production, College of Agricultural Engineering Sciences, University of Duhok, Zakho Street 38, 42001 Duhok City, Kurdistan Region, Iraq - sarbast.ihsan@uod.ac.

Received: 19 Jan., 2023 / Accepted: 24 Mar., 2024 / Published: 15 May., 2024. https://doi.org/10.25271/sjuoz.2024.12.2.1252

ABSTRACT:

Identification of genetic markers to distinguish animals within and between species demands extensive genomic and bioinformatics investigation. Previous studies have not carefully taken into consideration the effect of mitogenomic components on the genetic differentiation of the maternal lineages in goats. As a precaution, the complete goat mitogenome was downloaded from the NCBI database and used in the current study to assess the effects of the choice of mitogenomic fragments on phylogenetic studies and to identify any potential polymorphic region by which the main maternal haplogroups of goats can be classified. Phylogenetic results confirmed that all 13 individual mitochondrial protein-coding genes and 2 ribosomal genes are not applicable to differentiate the maternal lineages. Instead, a single novel polymorphic region with a length of 756 bp within the control region was successfully amplified by newly designed primers. Both phylogenetic analysis and principal components analysis of the sequenced mitogenomic region of the mtDNA control region efficiently differentiated the main maternal haplogroups in goats. Higher numbers of polymorphic sites were found in the control region and the mitogenomic marker region. Highly significant correlations were discovered between the polymorphic sites and the length of each individual mitogenomic component. Our results demonstrate useful guidance and cautionary notes for researchers who are interested in the investigation of genetic diversity in animal species using mtDNA sequences. The bioinformatics and molecular methods used herein can be powerful in selecting a minimum amount of data using PCR amplification when the entire sequences of the mitogenome are unavailable.

KEYWORDS: Bioinformatics; Capra hircus; mtDNA marker; Mitochondrial genome; Phylogeny

1. INTRODUCTION

Archeozoological and molecular studies indicated that goats were among the first livestock species that were domesticated around 8000–11,000 years ago in the Fertile Crescent, from where goats spread throughout the world along terrestrial and maritime routes of migration and trade (Naderi *et al.*, 2007; Zeder, 2008). Goats have various, relevant, and vital roles in the world, from the emergence of human civilization to the production of meat, milk, and fiber (Lu, 2023). Accurate molecular data within- and between-breed diversity are widely accepted to be essential for optimal management of genetic resources in farm animals (Fernández *et al.*, 2011). Understanding how historical and ongoing evolutionary processes have shaped the distribution of today's goat biodiversity necessitates a knowledge of genetic diversity and population dynamics that are derived from phylogenetic studies and haplogroup classifications (Al-Araimi *et al.*, 2017a). Genetic variation and phylogenetic relationships based on the variety of maternal lineages can be used to determine the ancestry and origins of different domestic animals by studying nucleotide polymorphisms within the mitochondrial DNA (mtDNA) sequences (Kimura *et al.*, 2011; Lv *et al.*, 2015; Yang *et al.*, 2017; Mustafa *et al.*,2018; Al-Jumaili *et al.*, 2020; Mustafa *et al.*, 2022).

Mitogenomic and phylogenetic analysis using mtDNA sequences of domestic and wild goats from geographically diverse countries has classified six major maternal haplogroups (HPG) known as HPG-A, HPG-B, HPG-C, HPG-D, HPG-G, and HPG-F, in which HPGA was predominant and widely widespread in most goats worldwide (Luikart *et al.*, 2001; Sultana *et al.*, 2003; Joshi *et al.*, 2004; Sardina *et al.*, 2006; Naderi *et al.*, 2007; Naderi *et al.*, 2008; Bulut *et al.*, 2016; Al-Araimi *et al.*, 2017b; Diwedi *et al.*, 2020; Ganbold *et al.*, 2020; Mustafa *et al.*, 2022). These previous studies used laborious, expensive, and time-consuming methods to determine the patterns of genetic diversity in goats. For example, some of them amplified the complete mitogenome by using several pairs of primers, while others used whole genome sequencing data. Alternatively, with the aid and development of bioinformatics, databases, and tools, the current study aimed to estimate the suitability of protein-coding genes, ribosomal genes, and the control region (D-loop) of the complete mitogenomes of domestic and wild goats to differentiate their main maternal lineages. Additionally, it also aimed to calculate polymorphic sites within each individual mitogenomic component, thus discovering the most polymorphic region by which the main maternal haplogroups of goat mitogenomes can be distinguished. The correlation between the length of each distinct mtDNA region and its number of polymorphic sites was also investigated.

2. MATERIAL AND METHODS

2.1 Mitogenomic data from the NCBI database

In total, 29 complete mitogenomes, including 13 protein-coding genes, 2 ribosomal (rRNA), 22 transfer RNA (tRNA), and one non-coding control region, with known maternal haplogroups (HPG-A, HPG-B, HPG-C, HPG-D, HPG-G, and HPG-F) of goats from both species *Capra hircus* and *Capra aegagrus* from geographically different locations were retrieved from the National Center for Biotechnology Information (NCBI) database https://www.ncbi.nlm.nih.gov/ and used for phylogenetic and bioinformatics analyses. GenBank accession

* Corresponding author

numbers, species, maternal haplogroups, and country of complete mitogenomes of both wild and domestic goats used in the present study are provided in Appendices Table S1

2.2 Phylogenetic analysis

To examine the suitability of individual genes or regions of the mitogenome for differentiation of the main maternal lineages in goats; firstly, the complete mitogenomes from domestic and wild goats worldwide representing all known maternal haplogroups were aligned using the Geneious Prime software and used for the phylogenetic tree as a reference (Kearse *et al.*, 2012). Then, complete sequences of 13 mitochondrial protein-coding genes, 2 ribosomal genes, and the control region of the goat mitogenome were extracted separately, re-aligned, and used for a phylogenetic tree using the Geneious Prime software. In phylogenetic analysis, a neighbor-joining (NJ) method as a tree-building technique and Tamura-Nei (TN93) as an appropriate genetic distance model as implemented in the Geneious Prime software were applied. The complete mitogenome of *Capra ibex* as an outgroup was used (Hassanin *et al.*, 2009).

2.3 Primer design

Aphylogenetic examination of each individual mitogenomic component led to the selection of the control region for additional bioinformatics analysis. According to the goat mitogenome reference (NCBI Reference Sequence: NC_005044.2),a pair of primers, forward (5' CAACACAAACTTCCCACTCCAC 3') and reverse (5' AGATGGCCCTGAAGAAAGAACC 3'), flanking the most polymorphic region within the control region of the goat mitogenome, was designed using the Geneious Prime software (Kearse *et al.*, 2012). Primers begin at position 15461 and end at position 16217 within the control region of the mitogenome reference of goat (NCBI Reference Sequence: NC_005044.2). Accordingly, the 756 bp target size of the PCR products was expected. The commercial synthesis of primer sequences was carried out by Macrogen, Inc. (South Korea).

2.4 Blood and DNA samples

Five fresh whole blood samples were collected from the jugular veins of Meriz goats and collected in vacutainer tubes (VACUTEST, Italy). Using the AddPrep Genomic DNA Extraction Kit (ADD BIO Inc., South Korea), total genomic DNA was isolated from whole blood according to the manufacturer's instructions. Quality and quantity of extracted genomic DNA samples were assessed using both a Nanodrop® ND-2000 spectrophotometer (Thermo Scientific) and gel electrophoresis (Cleaver Scientific Ltd., UK).

2.5 PCR amplification

For amplification of the most polymorphic mitogenomic marker region, PCR reactions were carried out in a 40µl total volume reaction mixture containing ddH_2O (14µl), Taq master (20µl) (ADD BIO Inc., South Korea, REF: 35001), forward and reverse primers (10µM) (each 2µl). At the end, about 100-150 ng of DNA samples were added to the master mix. Mitogenomic marker region was amplified following the PCR cycling conditions of 5 minutes of initial denaturation at 95°C, followed by 35 cycles of denaturation (95°C, 0.5min), annealing (58°C, 0.5 min), and primer extension (72°C, 1 min). The final cycle included a 5-minute final extension at 72°C followed by an indefinite hold duration at 4°C. The ABI 9700 thermal cycler (Applied Biosystems) at the Department of Animal Production, College of Agricultural Engineering Sciences, University of Duhok was used for PCR amplification.

2.6 Gel electrophoresis

To examine the outcome of PCR amplification and determine the band size of the PCR products, the gel electrophoresis method was used, in which an agarose gel (1% w/v) was prepared in 1x TAE and then the Safe Gel Stain Dye was added. The gel wells were loaded with the PCR products and DNA loading buffer. Additionally, 5µl of DNA ladder was added. The gel was then run at 75–100 volts (V) for 60–90 minutes using gel electrophoresis apparatus (Cleaver Scientific Ltd., UK). Finally, the PCR products were examined with a UV transilluminator.

2.7.DNA sequencing and phylogenetic analysis of PCR products

Five samples of amplified PCR products from Meriz goats were sent to Macrogen, Inc. (South Korea). The PCR products were purified and sequenced in both directions (forward and reverse). The chromatograms of the sequenced PCR products were visualized by Geneious Prime software (Kearse *et al.*, 2012), and clean PCR fragments were extracted per sample in order to generate the most polymorphic mitogenomic marker region with a length of 756 bp. Furthermore, for the phylogenetic tree, the five samples of Meriz goats representing the mitogenomic marker region were aligned with the comparable region shown in the mitogenomes of the other 29 domestic and wild goats utilized in Figure 1.

2.8.Mitogenomic pairwise and principal component analysis (PCA)

Pairwise differences in the mitogenomic marker region in terms of the number of polymorphic sites between the 34 samples, including 29 domestic and wild goats and 5 samples of Meriz goats, were estimated using the Geneious Prime software (Kearse *et al.*, 2012). The table representing the number of polymorphic sites was used as input for the TBTools program (Chen *et al.*, 2020) in order to create Principal Components Analysis (PCA) to visualize mitogenomic distances between the 34 samples representing the main maternal haplogroups in goats.

2.9 Estimation and visualization of polymorphic sites

Polymorphic sites over the consensus of the sequence alignment of individual component of mitogenome including 13 protein coding genes, 2 ribosomal genes, control region, and the mitogenomic marker region of goats were calculated using the Geneious variant finder using the find variations/SNPs feature from the annotate and predict menu within the Geneious Prime software (Kearse *et al.*, 2012). Polymorphic sites were visualized by Excel Microsoft. Statistical correlation between length of each mitogenomic component and the number of polymorphic sites were estimated using IBM SPSS's statistical software.

3. RESULTS AND DISCUSSIONS

3.1 Phylogenetic relationships based on the assessment of complete mitogenome

The current study examines the entire mitogenome and its individual genes to assess their power as a tool in the differentiation of the main maternal lineages in goats through the phylogenetics interpretation of mitogenome. Firstly, based on the alignments of 29 complete mitogenomes, the main maternal haplogroups HPG-A, HPG-B, HPG-C, HPG-D, HPG-G and HPG-F of domestic and wild goats used herein were classified through building phylogenetic trees based on a neighbour-joining (NJ) method under the appropriate parameter model using the Geneious Prime software (Kearse *et al*., 2012) (Figure 1). These results were consistent with the findings of previous studies (Colli *et al.*, 2015; Mustafa *et al*., 2022) when they found similar maternal lineages in goats using the complete mitogenome. Using nucleotide polymorphisms found in mtDNA sequences, phylogenetic relationships based on the diversity of maternal lineages can be utilized to retrace the origins and genealogy of various domestic animals (Naderi *et al.*, 2007; Lippold *et al*., 2011; Mustafa *et al*., 2018; Osman *et al*., 2021). MtDNA is

widely employed to generate phylogenetic trees at various taxonomic levels, from within species to among mammalian orders (Bruford *et al*., 2003).

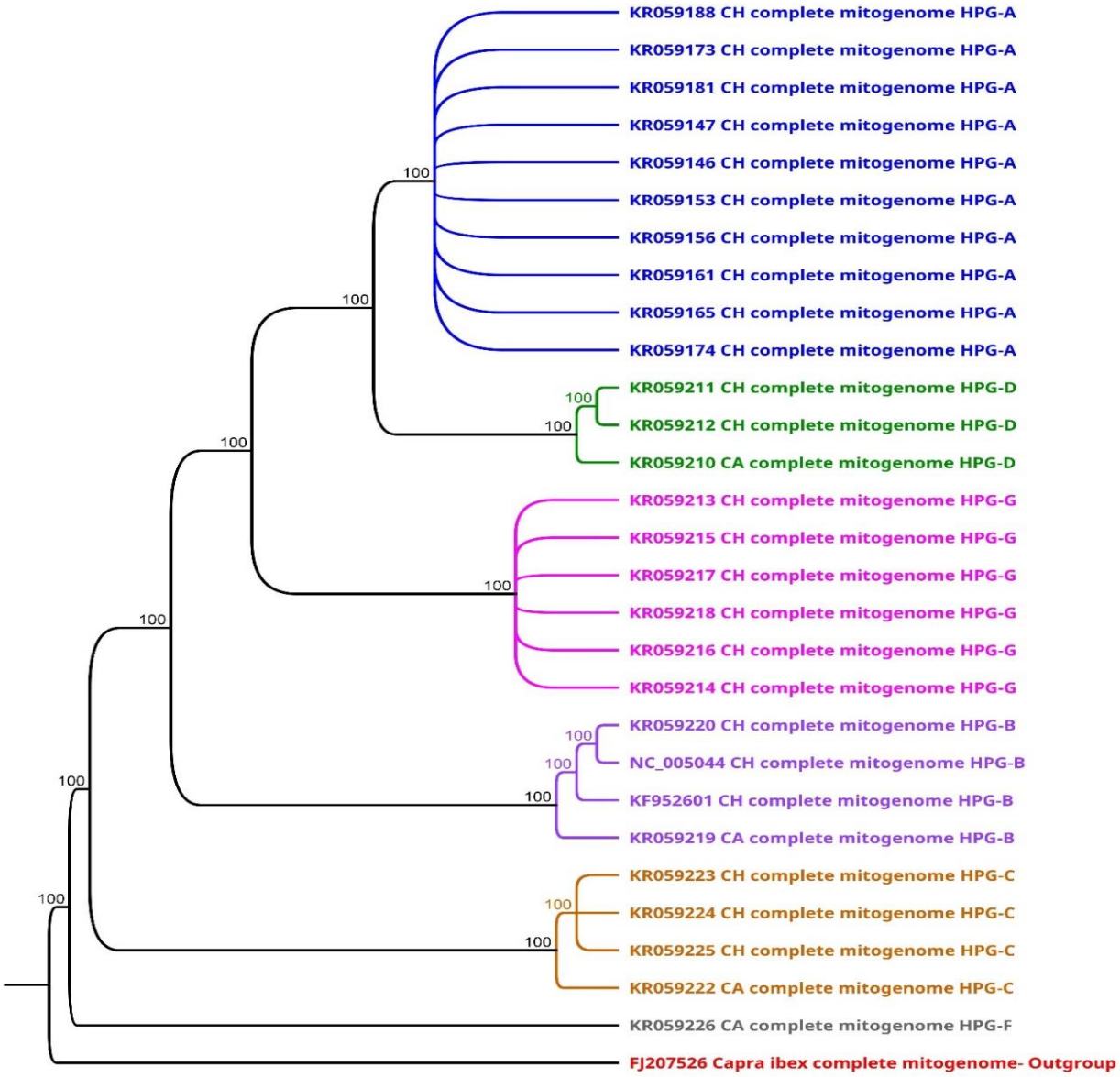

Figure 1: Phylogenetic tree of complete mitogenomes of domestic goats *Capra hircus* (CH) and wild goats *Capra aegagrus* (CA) showing the main maternal haplogroups HPG-A, HPG-B, HPG-C, HPG-D, HPG-G, and HPG-F. The mitogenome of Capra ibex was used as an outgroup. Table S1 shows GenBank numbers, haplogroups, and references for mitogenome sequences that were downloaded from NCBI.

3.2 Phylogenetic relationships based on the 13 protein coding genes, two ribosomal genes, and control region

After that, to assess the effect of 13 mitochondrial protein coding genes on the phylogenetic analysis, individual genes ND1, ND2, COX1, COX2, ATP8, ATP6, COX3, ND3, ND4L, ND4, ND5, ND6 and CYTB from the 29 aligned complete mitogenome representing all known maternal haplogroups from Figure 1 was extracted, re-aligned and used for phylogenetic trees using Geneious Prime software (Kearse *et al.*, 2012). The phylogenetic results indicated that none of them were suitable to differentiate all main maternal haplogroups (Figure 2; Appendices Figures S1-S13; Appendices Table S2). For instance, ND2, COX2, ND4 and CYTB genes were appropriate for differentiation three maternal haplogroups, while the remaining genes were fitting for discrimination only two maternal haplogroups or less (Table 1). Similar to the phylogenetic findings of protein-coding genes, the comparative alignment based on ribosomal genes (16S rRNA and 12S rRNA) specified that none of them were appropriate for classification the main maternal haplogroups (Figure 2; Appendices Figures S14 and S15; Appendices Table S2). In contrast to the results of 13 coding and 2 ribosomal genes, sequences of the mitochondrial control region were suitable for differentiation the main maternal haplogroups from each other (Figure 2; Appendices Figure S16; Appendices Table S2). Our findings are well corroborated by previous research that used the control region in which the six highly divergent haplogroups in domestic goats were identified (Vacca *et al*., 2010; Ghernouti *et al*., 2017; Nguluma *et al*., 2021; Baenyi Simon *et al*., 2022).

The control region has been widely used as a genetic marker in phylogenetic investigations of many different vertebrate classes such as reptiles (Jiang *et al.*, 2011), fish (Jamandre *et al.*, 2014), amphibians (Huang and Tu, 2016), chickens (Li *et al.*, 2019) and sheep (Wanjala *et al.*, 2021).

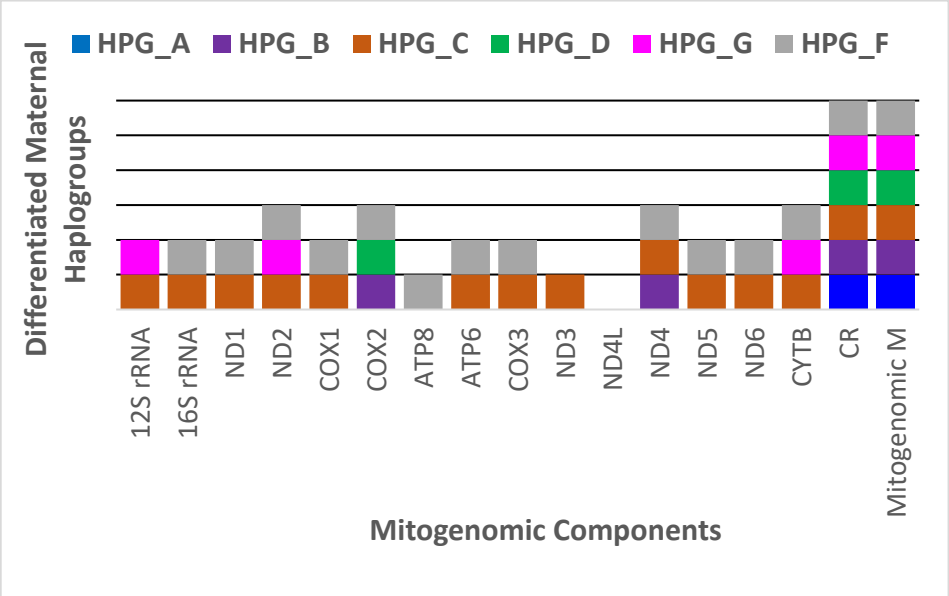

Figure 2: Differentiation of the main maternal haplogroups using a single component of the goat mitogenome. The results depicted in this figure are derived from the display of data from Appendices Table S2 using Excel.

3.3 Amplification, sequencing and phylogeny of the mitogenomic marker region

Since the goat mitogenome's control region is longer than 1 kbp and is challenging to sequence it with high-quality using Sanger sequencing technology, it is essential to identify mtDNA region shorter than 1 kbp with highly polymorphic sites. Accordingly, the sequences of 29 aligned complete mitogenomes were scanned by focusing on the control region and it has been found that highly polymorphic region within the control region could be used as a mitogenomic marker to differentiate all maternal haplogroups. As a result, a pair of primers was designed to flank the mitogenomic marker region that starts at position 15461bp and ends at position 16217bp of control region according to the mitogenome reference of goat (NCBI Reference Sequence: NC_005044.2). The most polymorphic mitogenomic marker region with a size 756bp of five genomic DNA samples of domestic Meriz goat was successfully amplified by PCR reactions using a pair of primer designed by our study. Meriz goats have been chosen as the sole model for representing goat genomic DNA in order to be used for PCR amplification and sequencing analysis of the mitogenomic marker region and to confirm our bioinformatics findings. The second reason behind the selection of genomic DNA samples from Meriz goats is that their maternal lineage (haplogroup) and entire mitogenomes have already been determined (Mustafa *et al.*, 2022).

Previous studies used huge numbers of primer to amplify multiple overlapping fragments to assemble the complete mitogenome and then used for maternal phylogeny (E *et al.*, 2016). However, in this study only a pair of primers were used to target highly informative region containing sufficient single nucleotide polymorphisms (SNPs) able to differentiate all maternal lineages in goats (Figure 2, Figure 3). In addition, Doro *et al*. (2014) used over 30 pairs of primers to amplify 33 partially overlapping mtDNA amplicons in order to assemble the whole mtDNA genome, which was then used to construct the goat phylogenetic tree. However, in our bioinformatics research, a pair of primers to identify the most informative region within the control region was developed, which represents the whole mitogenome and generates the same results. The five sequences of the PCR products of the control region representing the mitogenomic marker region of domestic Meriz goat were aligned with the corresponding region extracted from the 29 complete mitogenomes spanning the same flank region, 15461 bp to 16217 bp (756 bp in size). As a result, the 34 mtDNA control region sequences representing the mitogenomic marker region were used for phylogenetic tree to evaluate its productivity for differentiation the main maternal lineages of goats. As shown in Figure 3, the targeted region (756 bp) within the ranges 15461 bp to 16217 bp of the mtDNA control region was appropriate for distinguishing the main maternal haplogroups in goats. Similar to the phylogenetic results based on the complete mitogenome (Figure 1) and on the control region (Appendices Figure S16), the 756 bp mitogenomic region was like a powerful molecular tool to differentiate all major maternal lineages of domestic and wild goats, including the validated sequences of the mitogenomic region of Meriz goats (Figure 2; Figure 3).

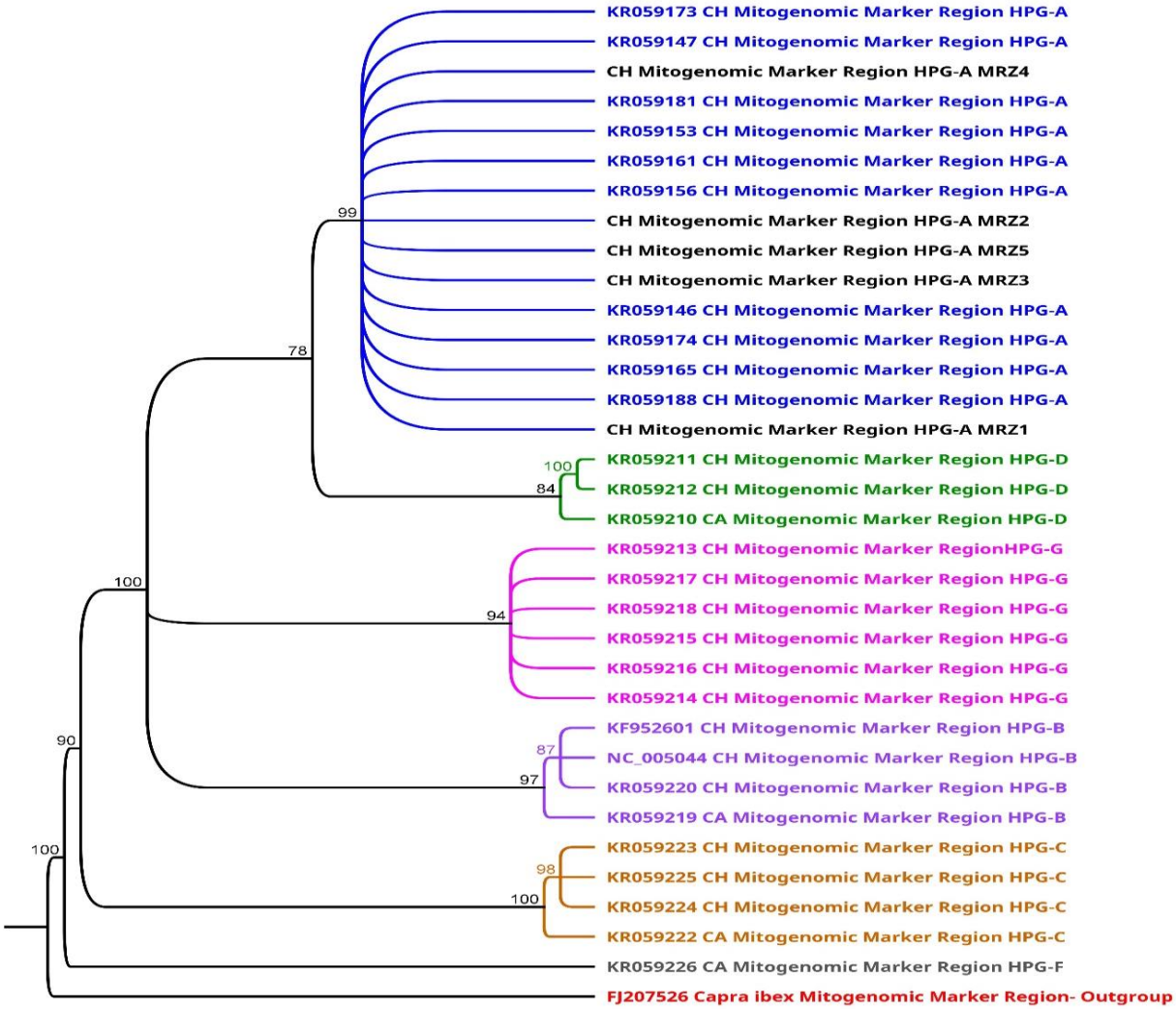

Figure 3: Phylogenetic relationships using mitogenomic marker regions segregating the main maternal haplogroups (HPG-A, HPG-B, HPG-C, HPG-D, HPG-G, and HPG-F).

Previous research examined the evolutionary history and population structure of 406 domestic goats representing 88 breeds by sequencing the hypervariable section of the mtDNA control region with a length of 481bp and their phylogeographic analysis revealed only three distinct goat lineages, HPG-A, HPG-B, and HPG-C (Luikart *et al.*, 2001). Furthermore, their study (Luikart *et al*, 2001) used external and internal primers to amplify short segment of control region not having adequate polymorphic sites to distinguish all maternal lineages, and through our analysis, it has been found thatthe sequences of their primers are not within the most conserved region of mitochondrial control region. However, in our study, a new pair of primers was designed within the most conserved regions amplifying mtDNA control region with the length 756 bp with highly polymorphic sites able to differentiate the main six maternal lineages in the goats (Figure 2, Figure 3). In another study, a fragment of the control region with a size of 639 bp, including primers, corresponded to the locations 15653 to 16250 on the complete mitogenome of reference number AF533441 and was used to analyze the genetic diversity of more than 2400 goats (Naderi *et al.*, 2007). The target region flanked by our primers is 127 bp longer than the sequenced segment of the control region (756 bp versus 639 bp) by the primer sequences of the study (Naderi *et al.*, 2007). The control region has been notably employed to characterize the genetic diversity of goats (Luikart *et al.*, 2001; Naderi *et al.*, 2007) because it is sufficiently instructed and variable over the species' geographic range and evolves at a steady pace (Bruford *et al*., 2003).

In order to verify the performance of the mitogenomic marker, pairwise differences of the mitogenomic marker region in terms of the number of polymorphic sites between the 34 samples including 29 domestic and wild goats and 5 samples of Meriz goats were estimated using the Geneious Prime software (Kearse *et al.*, 2012). TBTools (Chen *et al.*, 2020) was used to create Principal Components Analysis (PCA) in order to visualize mitogenomic distances between the 34 samples. Similar to the results of the phylogenetic trees shown in Figure 1, Figure 3, and Appendices Figure S16, PCA analysis interestingly distinguished the main maternal haplogroups (Figure 4). The results of phylogenetic tree and PCA indicated that the Meriz goats are within the maternal lineage haplogroup A(HPG-A) (Figure 3). The results are highly supported by our previous study (Mustafa *et al*., 2022) when the complete mitogenome was used and it has been found that Meriz goats belonged to the maternal haplogroup A (HPG-A). The concordance between the results of our bioinformatics analysis, the phylogenetic tree, principal components analysis, and the previous study (Mustafa *et al*., 2022) implies that the mtDNA marker from the control region of study is strong enough to alternate the entire mitogenome for maternal haplogroup identification when it sufficiently identifies that Meriz goats are within the maternal haplogroup HPG-A (Figure 3, Figure 4).

Furthermore, our results pointed out that the size of the mitogenomic marker region (756 bp) and the number of polymorphic sites in it were adequate to represent the complete mitogenome of goats due to the high polymorphisms in this

region (Figure 5). Similarly, most recently, a meta-analysis was conducted to examine the diversity and phylogenetic relationships among 399 mitogenomic control region sequences from African sheep breeds. The significant nucleotide and haplotype diversity, along with the observed huge number of haplotypes (more than 70% distinct haplotypes), point to a very high variability of the mtDNA control region (Wanjal *et al*., 2021). Variations in phylogeny estimates between mitogenomic and single-gene markers have resulted in contradicting conclusions in various occasions. Thus, selecting mitogenomic regions as the marker of choice for phylogenetic analysis often provides greater phylogenetic resolution and precision, which is very important, as we discovered herein in the goat mitogenome.

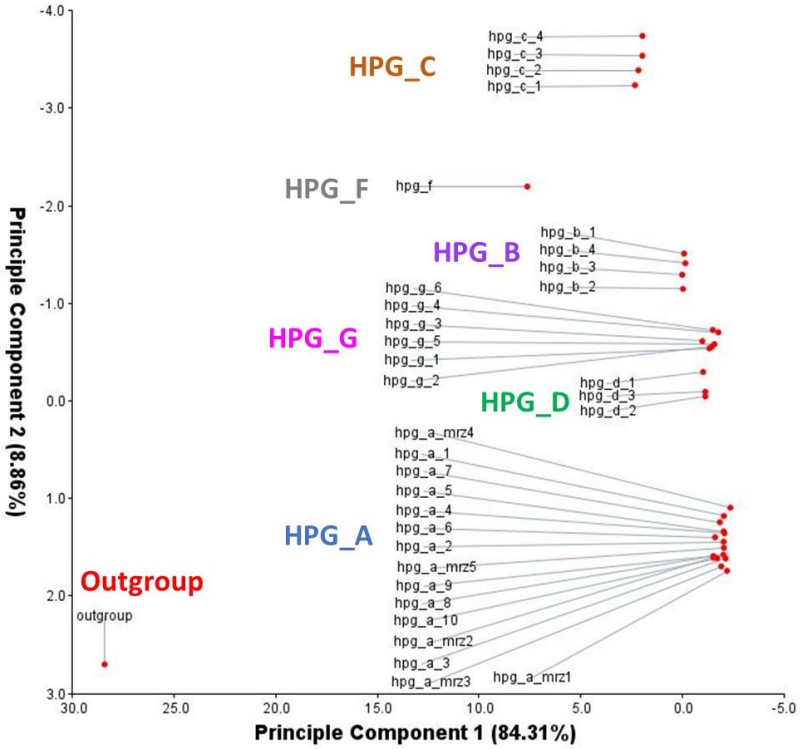

Figure 4. Principal Components Analysis (PCA) based on the pairwise distance estimated by the number of polymorphic sites in the mitogenomic marker region between the 34 samples, including 29 domestic and wild goats and 5 samples of Meriz goats.

3.4 Assessment, visualization and correlation of polymorphic sites

In the present study, polymorphic sites were estimated across the consensus of the sequence alignment of individual components of the goat mitogenome, comprising 13 protein-coding genes, 2 ribosomal genes, and the control region, using the Geneious Prime software's variant finder (Kearse *et al.*, 2012). The results, as shown in Figure 5, indicated that 12 protein-coding genes, including ND3, ATP8, ND4L, ND6, 12SrRNA, 16SrRNA, COX2, ND1, APT6, COX3, ND2, and COX1, had a number of polymorphic sites less than 40 SNPs. Three genes, such as ND4, CYTB, and ND5, had polymorphic sites with a range of 61 to 78 SNPs (Figure 5). However, the highly polymorphic regions were both the control region and the mitogenomic marker region, with 188 and 147 SNPs, respectively (Figure 5).

Luikart *et al.* (2001) conducted an extensive sampling study including 420 mtDNA control region sequences from 88 breeds and 160 variable sites were observed by which only three distinct maternal lineages, HPG-A, HPG-B, and HPG-C, were revealed in domestic and wild goats from Nigeria to Iceland, Mongolia to Malaysia, and countries from domestication center such as Jordan, Turkey, Iraq, Egypt, and Pakistan. Despite the use of large samples, other maternal haplogroups such as HPG-D, HPG-G, and HPG-F were not distinguished in their research, but other studies later did. This is convincing indication that the short mtDNA control region sequences (481 bp) employed in their phylogeographic study did not cover the potential number of polymorphic sites required to uncover all maternal lineages in goats. Thus, as a note of caution, future research on maternal haplogroups in goats should employ a longer region with more polymorphic sites, as we discovered herein. The control region can be used to track phylogeography, demographic growth, dispersal, gene flow, hybridization, and genetic drift since it exhibits exceptionally high levels of variation within species (Bruford *et al.*, 2003; Verma *et al.*, 2018).

Furthermore, based on the Spearman's rank correlation coefficient, highly significant correlations were discovered between the polymorphic sites and the length of each mitogenomic component (13 protein-coding genes, two ribosomal RNA genes, a D-loop region, and a mitogenomic marker) after applying the Shapiro normality test by IBM SPSS's statistical software (SPSS, 2019). Likewise, Naderi *et al.* (2007) observed about 336 polymorphic sites over the 558bp of the alignment of the control region fragment when they examined genetic diversity of more than 2400 individuals of goats. Polymorphic sites are important because each SNP has a value, as demonstrated by the work of Doro *et al.* (2014), who discovered that several polymorphic sites inside the HPG-A were relevant for the construction of its internal phylogeny. Yi *et al.* (2022) employed a 401-bp fragment of the mtDNA control region in their investigation of population structure, genetic diversity, and matrilineal system evolution in goats globally, which included over 190 goat breeds. Although they discovered 301 polymorphic sites within the 401 bp control region, using a longer fragment of the mtDNA control region in their analysis would result in a higher number of variable sites, giving us a

better understanding of goat domestication history and gene flow worldwide.

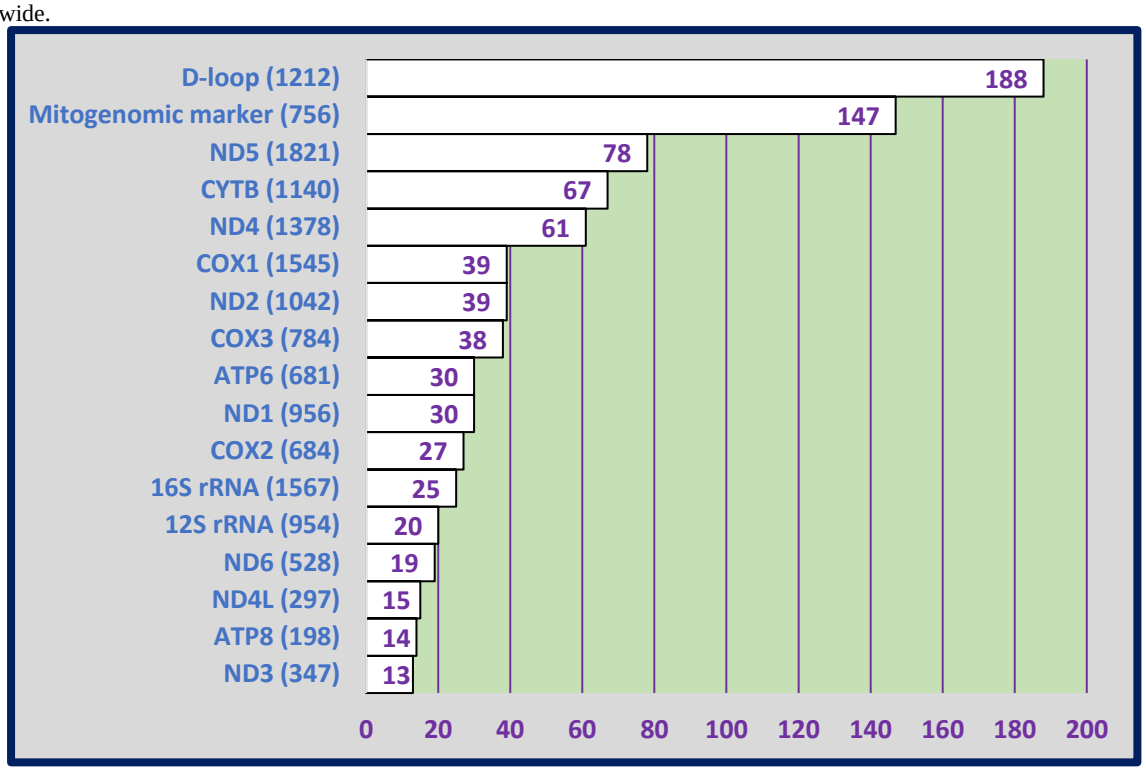

Figure 5: The number of polymorphic sites and the length of 13 protein-coding genes, 2 ribosomal RNA genes, and the D-loop region (control region).

CONCLUSIONS

Bioinformatics and molecular analyses revealed that neither the 13 protein-coding genes nor the two ribosomal genes of mitogenome as individual subsets are suitable for distinguishing the main maternal haplogroups in goats. Alternatively, a complete mitogenome, a control region, and the most informative mitogenomic region with highly polymorphic sites with a size of 756 bp inside the control region, which were amplified using PCR utilizing newly designed primers, effectively differentiated the main maternal lineages. Moreover, significant correlations between the length of individual genes and the number of polymorphic sites were found, indicating that the region with the highest polymorphism is best suited for use as a mitogenomic marker for maternal phylogeny. The bioinformatics analysis used here can be applied to population genetics and biodiversity in other species in future research, suggesting that when examining genetic diversity, mitochondrial fragments should be chosen with caution.

Data availability statement

The partial mitochondrial control region of five samples of Meriz goats is available in GenBank at the NCBI database under accession numbers OR875928-OR875932.

REFERENCES

Al-Araimi, N.A., Al-Atiyat, R.M., Gaafar, O.M., Vasconcelos, R., Luzuriaga-Neira, A., Eisa, M.O., Amir, N., Benaissa, M.H., Alfaris, A.A., Aljumaah, R.S., Elnakhla, S.M., Salem, M.M.I., Ishag, I.A., El Khasmi, M. and Beja-Pereira, A. (2017a). Maternal genetic diversity and phylogeography of native Arabian goats, *Livestock Science*, 206, 88–94. Available at: https://doi.org/https://doi.org/10.1016/j.livsci.2017.09.017.

Al-Araimi, N.A., Gaafar, O.M., Costa, V., Neira, A.L., Al-Atiyat, R.M. and Beja-Pereira, A. (2017b). Genetic origin of goat populations in Oman revealed by mitochondrial DNA analysis., *PloS one*, 12(12), p. e0190235. Available at: https://doi.org/10.1371/journal.pone.0190235.

Al-Jumaili, A.S., Boudali, S.F., Kebede, A., Al-Bayatti, S.A., Essa, A.A., Ahbara, A., Aljumaah, R.S., Alatiyat, R.M., Mwacharo, J.M., Bjørnstad, G., Naqvi, A.N., Gaouar, S.B.S. and Hanotte, O. (2020). The maternal origin of indigenous domestic chicken from the Middle East, the north and the horn of Africa, *BMC Genetics*, 21(1), 1–16. Available at: https://doi.org/10.1186/s12863-020-0830-0.

Baenyi Simon, P., Owino Junga, J., Tarekegn, G.M., Machuka, E., Tiambo, C.K., Kabange, D., Musale M. Dieudinné, K., Vumilia Kizungu, R., Ochieng, J.W. and Pelle, R. (2022). Haplotype analysis of the mitochondrial DNA d-loop region reveals the maternal origin and historical dynamics among the indigenous goat populations in east and west of the Democratic Republic of Congo. *Ecology and Evolution*, 12(3), p.e8713. Available at: https://doi.org/10.1002/ece3.8713.

Bruford, M.W., Bradley, D.G. and Luikart, G. (2003). DNA markers reveal the complexity of livestock domestication, *Nature Reviews Genetics*, 4(11), 900–910. Available at: https://doi.org/10.1038/nrg1203.

Bulut, Z., Kurar, E., Ozsensoy, Y., Altunok, V. and Nizamlioglu, M. (2016). Genetic Diversity of Eight Domestic Goat Populations Raised in Turkey., *BioMed research international*, 2016, p. 2830394. Available at: https://doi.org/10.1155/2016/2830394.

Chen, C., Chen, H., Zhang, Y., Thomas, H.R., Frank, M.H., He, Y. and Xia, R. (2020) TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data, *Molecular Plant*, 13(8), 1194–1202. Available at: https://doi.org/https://doi.org/10.1016/j.molp.2020.06.009.

Colli, L., Lancioni, H., Cardinali, I., Olivieri, A., Capodiferro, M.R., Pellecchia, M., Rzepus, M., Zamani, W., Naderi,

S., Gandini, F., Vahidi, S.M.F., Agha, S., Randi, E., Battaglia, V., Sardina, M.T., Portolano, B., Rezaei, H.R., Lymberakis, P., Boyer, F., Coissac, E., Pompanon, F., Taberlet, P., Ajmone Marsan, P. and Achilli, A. (2015). Whole mitochondrial genomes unveil the impact of domestication on goat matrilineal variability, *BMC Genomics*, 16(1), p. 1115. Available at: https://doi.org/10.1186/s12864-015-2342-2.

Diwedi, J., Singh, A.W., Ahlawat, S., Sharma, R., Arora, R., Sharma, H., Raja, K.N., Verma, N.K. and Tantia, M.S. (2020). Comprehensive analysis of mitochondrial DNA based genetic diversity in Indian goats., *Gene*, 756, p. 144910. Available at: https://doi.org/10.1016/j.gene.2020.144910.

Doro, M.G., Piras, D., Leoni, G.G., Casu, G., Vaccargiu, S., Parracciani, D., Naitana, S., Pirastu, M. and Novelletto, A. (2014). Phylogeny and patterns of diversity of goat mtDNA haplogroup A revealed by resequencing complete mitogenomes. *PLoS One*, 9(4), p.e95969. Available at:https://doi.org/10.1371/journal.pone.0095969.

E, G.-X., Huang, Y.-F., Liu, N., Zhao, Y.-J., He, J.-N., Na, R.-S., Zhao, Z.-Q., Jiang, C.-D., Zhang, J.-H., Ma, Y.-H., Chen, L.-P., Qiu, X.-Y., Sun, Y.-W., Zeng, Y., Sun, Y.-Z., Yu, C.-H. and Wei, S.-Y. (2016). Characteristics of the mitochondrial genome of four native goats in China (Capra hircus), *Mitochondrial DNA. Part A, DNA mapping, sequencing, and analysis*, 27(5), 3308–3309. Available at: https://doi.org/10.3109/19401736.2015.1015016.

Fernández, J., Meuwissen, T.H.E., Toro, M.A. and Mäki-Tanila, A. (2011). Management of genetic diversity in small farm animal populations, *Animal*, 5(11), 1684–1698. Available at: https://doi.org/10.1017/S1751731111000930.

Ganbold, O., Lee, S.-H., Paek, W.K., Munkhbayar, M., Seo, D., Manjula, P., Khujuu, T., Purevee, E. and Lee, J.H. (2020). Mitochondrial DNA variation and phylogeography of native Mongolian goats., *Asian-Australasian journal of animal sciences*, 33(6), 902–912. Available at: https://doi.org/10.5713/ajas.19.0396.

Ghernouti N, Bodinier N, Ranebi M, Maftah D, Petit D, Gaouar S B S. (2017). Control region of mtDNA identifies three migration events of sheep breeds in Algeria. *Small Ruminant Research*, 155, 66–71. Available at: https://doi.org/10.1016/j.smallrumres.2017.09.003.

Hassanin, A., Ropiquet, A., Couloux, A. and Cruaud, C. (2009) Evolution of the mitochondrial genome in mammals living at high altitude: new insights from a study of the tribe Caprini (Bovidae, Antilopinae)., *Journal of molecular evolution*, 68(4), 293–310. Available at: https://doi.org/10.1007/s00239-009-9208-7.

Huang, Z.H. and Tu, F.Y. (2016). Characterization and evolution of the mitochondrial DNA control region in Ranidae and their phylogenetic relationship., *Genetics and molecular research : GMR*, 15(3). Available at: https://doi.org/10.4238/gmr.15038491.

Jamandre, B.W., Durand, J.-D. and Tzeng, W.-N. (2014). High Sequence Variations in Mitochondrial DNA Control Region among Worldwide Populations of Flathead Mullet *Mugil cephalus*, *International Journal of Zoology*. Edited by T. Iliffe, 2014, p. 564105. Available at: https://doi.org/10.1155/2014/564105.

Jiang, Y., Nie, L.W., Huang, Z.F., Jing, W.X., Wang, L., Liu, L. and Dai, X.T. (2011). Comparison of complete mitochondrial DNA control regions among five Asian freshwater turtle species and their phylogenetic relationships., *Genetics and molecular research : GMR*, 10(3), 1545–1557. Available at: https://doi.org/10.4238/vol10-3gmr1205.

Joshi, M.B., Rout, P.K., Mandal, A.K., Tyler-Smith, C., Singh, L. and Thangaraj, K. (2004). Phylogeography and origin of Indian domestic goats., *Molecular biology and evolution*, 21(3), 454–462. Available at: https://doi.org/10.1093/molbev/msh038.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P. and Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data, *Bioinformatics*, 28(12), 1647–1649. Available at: https://doi.org/10.1093/bioinformatics/bts199.

Kimura, B., Marshall, F.B., Chen, S., Rosenbom, S., Moehlman, P.D., Tuross, N., Sabin, R.C., Peters, J., Barich, B., Yohannes, H., Kebede, F., Teclai, R., Beja-Pereira, A. and Mulligan, C.J. (2011). Ancient DNA from Nubian and Somali wild ass provides insights into donkey ancestry and domestication, *Proceedings of the Royal Society B: Biological Sciences*, 278(1702), 50–57. Available at: https://doi.org/10.1098/rspb.2010.0708.

Li, Q., Zhang, P., Li, M., Li, L., Hu, Y., Mishra, S.K., Guo, A., Li, G., Li, D. and Duan, Y. (2019). Genetic diversity and relationship of Dulong chickens using mitochondrial DNA control region., *Mitochondrial DNA. Part B, Resources*, 5(1), 275–280. Available at: https://doi.org/10.1080/23802359.2019.1700837.

Lippold, S., Matzke, N.J., Reissmann, M. and Hofreiter, M. (2011). Whole mitochondrial genome sequencing of domestic horses reveals incorporation of extensive wild horse diversity during domestication, *BMC Evolutionary Biology*, 11(1), p. 328. Available at: https://doi.org/10.1186/1471-2148-11-328.

Lu, C.D. (2023). The role of goats in the world: Society, science, and sustainability, *Small Ruminant Research*, 227, p. 107056. Available at: https://doi.org/https://doi.org/10.1016/j.smallrumres.2023.107056.

Luikart, G., Gielly, L., Excoffier, L., Vigne, J.-D., Bouvet, J. and Taberlet, P. (2001). Multiple maternal origins and weak phylogeographic structure in domestic goats, *Proceedings of the National Academy of Sciences*, 98(10), 5927–5932. Available at: https://doi.org/10.1073/pnas.091591198.

Lv, F.-H., Peng, W.-F., Yang, J., Zhao, Y.-X., Li, W.-R., Liu, M.-J., Ma, Y.-H., Zhao, Q.-J., Yang, G.-L., Wang, F., Li, J.-Q., Liu, Y.-G., Shen, Z.-Q., Zhao, S.-G., Hehua, Ee., Gorkhali, N.A., Farhad Vahidi, S.M., Muladno, M., Naqvi, A.N., Tabell, J., Iso-Touru, T., Bruford, M.W., Kantanen, J., Han, J.-L. and Li, M.-H. (2015). Mitogenomic Meta-Analysis Identifies Two Phases of Migration in the History of Eastern Eurasian Sheep, *Molecular Biology and Evolution*, 32(10), 2515–2533. Available at: https://doi.org/10.1093/molbev/msv139.

Mustafa, S.I., Heslop-Harrison, J.S. and Schwarzacher, T. (2022) The complete mitochondrial genome from Iraqi Meriz goats and the maternal lineage using whole genome sequencing data, *Iranian Journal of Applied Animal Science*, 12(2), 321–328.

Mustafa, S.I., Schwarzacher, T. and Heslop-Harrison, J.S. (2018) Complete mitogenomes from Kurdistani sheep: abundant centromeric nuclear copies representing diverse ancestors, *Mitochondrial DNA Part A: DNA Mapping, Sequencing, and Analysis*, 29(8), 1180–1193. Available at: https://doi.org/10.1080/24701394.2018.1431226.

Naderi, S., Rezaei, H.-R., Taberlet, P., Zundel, S., Rafat, S.-A., Naghash, H.-R., el-Barody, M.A.A., Ertugrul, O, Pompanon, F, and and Econogene Consortium (2007).

Large-scale mitochondrial DNA analysis of the domestic goat reveals six haplogroups with high diversity., *PloS one*, 2(10), p. e1012. Available at: https://doi.org/10.1371/journal.pone.0001012.

Naderi, S., Rezaei, H.R., Pompanon, F., Blum, M.G.B., Negrini, R., Naghash, H.R., Balkiz, Ö., Mashkour, M., Gaggiotti, O.E., Ajmone-Marsan, P., Kence, A., Vigne, J.D. and Taberlet, P. (2008) The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals, *Proceedings of the National Academy of Sciences of the United States of America*, 105(46), 17659–17664. Available at: https://doi.org/10.1073/pnas.0804782105.

Nguluma, A., Kyallo, M., Tarekegn, G.M., Loina, R., Nziku, Z., Chenyambuga, S. and Pelle, R. (2021). Mitochondrial DNA D-loop sequence analysis reveals high variation and multiple maternal origins of indigenous Tanzanian goat populations. *Ecology and Evolution*, 11(22), pp.15961-15971. Available at: https://doi.org/10.1002/ece3.8265.

Osman, S.A.-M., Nishibori, M. and Yonezawa, T. (2021). Complete mitochondrial genome sequence of Tosa-Jidori sheds light on the origin and evolution of Japanese native chickens., *Animal bioscience*, 34(6), 941–948. Available at: https://doi.org/10.5713/ajas.19.0932.

Sardina, M.T., Ballester, M., Marmi, J., Finocchiaro, R., van Kaam, J.B.C.H.M., Portolano, B. and Folch, J.M. (2006) Phylogenetic analysis of Sicilian goats reveals a new mtDNA lineage., *Animal genetics*, 37(4), 376–378. Available at: https://doi.org/10.1111/j.1365-2052.2006.01451.x.

Sultana, S., Mannen, H. and Tsuji, S. (2003). Mitochondrial DNA diversity of Pakistani goats., *Animal genetics*, 34(6), 417–421. Available at: https://doi.org/10.1046/j.0268-9146.2003.01040.x.

Verma, K., Sharma, S., Sharma, A., Dalal, J. and Bhardwaj, T. (2018) Data on haplotype diversity in the hypervariable region I, II and III of mtDNA amongst the Brahmin population of Haryana, *Data in Brief*, 17, 305–313. Available at: https://doi.org/10.1016/j.dib.2018.01.011.

Vacca, G.M., Daga, C., Pazzola, M., Carcangiu, V., Dettori, M.L. and Cozzi, M.C. (2010). D-loop sequence mitochondrial DNA variability of Sarda goat and other goat breeds and populations reared in the Mediterranean area. *Journal of Animal Breeding and Genetics*, 127(5), pp.352-360. Available at: https://doi.org/10.1111/j.1439-0388.2010.00863.x.

Wanjala, G., Bagi, Z. and Kusza, S. (2021) Meta-Analysis of Mitochondrial DNA Control Region Diversity to Shed Light on Phylogenetic Relationship and Demographic History of African Sheep (Ovis aries) Breeds., *Biology*, 10(8). Available at: https://doi.org/10.3390/biology10080762.

Yang, H., Wang, G., Wang, M., Ma, Y., Yin, T., Fan, R., Wu, H., Zhong, L., Irwin, D.M., Zhai, W. and Zhang, Y. (2017). The origin of chow chows in the light of the East Asian breeds, *BMC Genomics*, 18(1), 1–13. Available at: https://doi.org/10.1186/s12864-017-3525-9.

Yi, G.U.O., Ying, G.O.N.G., Yong-meng, H.E., Bai-gao, Y.A.N.G., Wei-yi, Z.H.A.N.G., Bo-er, C.H.E.N., Yong-fu, H.U.A.N.G., Yong-ju, Z.H.A.O., Dan-ping, Z.H.A.N.G., Yue-hui, M.A. and Ming-xing, C.H.U. (2022). Investigation of mitochondrial DNA genetic diversity and phylogeny of goats worldwide. *Journal of Integrative Agriculture*, 21(6), pp.1830-1837. Available at: https://doi.org/10.1016/S2095-3119(21)63882-0.

Zeder, M.A. (2008). Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact, *Proceedings of the National Academy of Sciences of the United States of America*, 105(33), 11597–11604. Available at: https://doi.org/10.1073/pnas.0801317105.

Appendices:

Table S1. GenBank accession numbers, goat species, maternal haplogroups, and country of complete mitogenomes used in the bioinformatics analysis of present study.

#	GenBank ID	Species	Maternal Haplogroups	Country
1	KR059146	C. hircus	A	Albania
2	KR059147	C. hircus	A	Switzerland
3	KR059153	C. hircus	A	Jordan
4	KR059156	C. hircus	A	Spain
5	KR059161	C. hircus	A	Turkey
6	KR059165	C. hircus	A	Italy
7	KR059173	C. hircus	A	Spain
8	KR059174	C. hircus	A	Romania
9	KR059181	C. hircus	A	Switzerland
10	KR059188	C. hircus	A	Egypt
11	KR059210	C. aegagrus	D	Iran
12	KR059211	C. hircus	D	Kyrgyzstan
13	KR059212	C. hircus	D	Kyrgyzstan
14	KR059213	C. hircus	G	Iran
15	KR059214	C. hircus	G	Turkey
16	KR059215	C. hircus	G	Turkey
17	KR059216	C. hircus	G	Iran
18	KR059217	C. hircus	G	Iran
19	KR059218	C. hircus	G	Iran
20	KR059219	C. aegagrus	B	Iran
21	KF952601	C. hircus	B	China
22	NC_005044	C. hircus	B	Vietnam
23	KR059220	C. hircus	B	Malaysia
24	KR059222	C. aegagrus	C	Iran
25	KR059223	C. hircus	C	Switzerland
26	KR059224	C. hircus	C	Switzerland
27	KR059225	C. hircus	C	Spain
28	KR059226	C. aegagrus	F	Iran
29	FJ207526	Capra ibex	Outgroup	European countries

Table S2. Differentiation of the main maternal haplogroups using single component of goat mitogenome. The results presented in this table are interpreted from the phylogenetic results of Appendices Figures S1-S16.

	Differentiation of maternal haplogroups					
Mitogenomic components	HPG-A	HPG-B	HPG-C	HPG-D	HPG-G	HPG-F
12S rRNA	-	-	✓	-	✓	-
16S rRNA	-	-	✓	-	-	✓
ND1	-	-	✓	-	-	✓
ND2	-	-	✓	-	✓	✓
COX1	-	-	✓	-	-	✓
COX2	-	✓	-	✓	-	✓
ATP8	-	-	-	-	-	✓
ATP6	-	-	✓	-	-	✓
COX3	-	-	✓	-	-	✓
ND3	-	-	✓	-	-	-
ND4L	-	-	-	-	-	-
ND4	-	✓	✓	-	-	✓
ND5	-	-	✓	-	-	✓
ND6	-	-	✓	-	-	✓
CYTB	-	-	✓	-	✓	✓
Control region (D-loop)	✓	✓	✓	✓	✓	✓
Mitogenomic marker	✓	✓	✓	✓	✓	✓

Note:
✓ = Applicable
- = Not Applicable

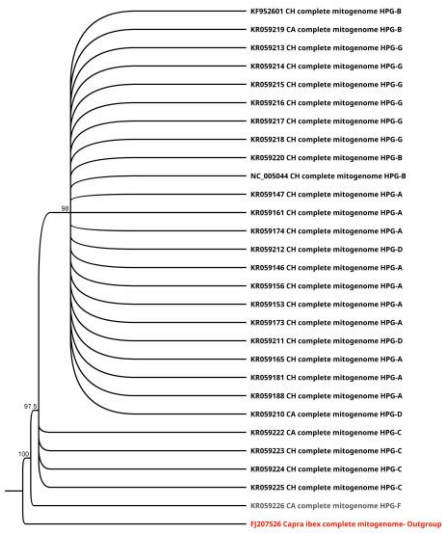

Figure S1. Differentiation of the main maternal haplogroups of domestic goats *Capra hircus* (CH) and wild goats *Capra aegagrus* (CA) based on the phylogeny of ND1 gene.

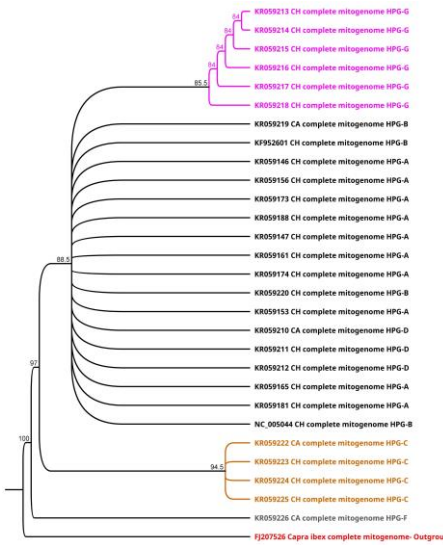

Figure S2. Differentiation of the main maternal haplogroups of domestic goats *Capra hircus* (CH) and wild goats *Capra aegagrus* (CA) based on the phylogeny of ND2 gene.

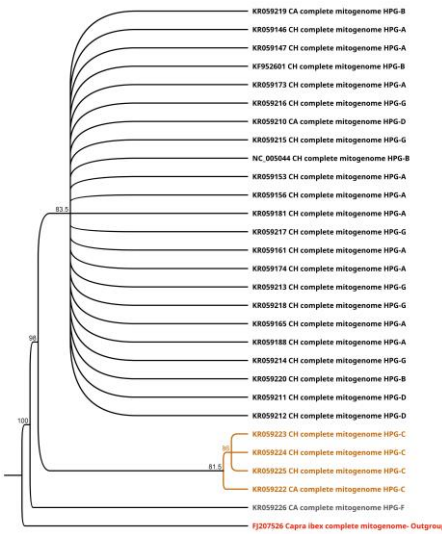

Figure S3. Differentiation of the main maternal haplogroups of domestic goats *Capra hircus* (CH) and wild goats *Capra aegagrus* (CA) based on the phylogeny of COX1 gene.

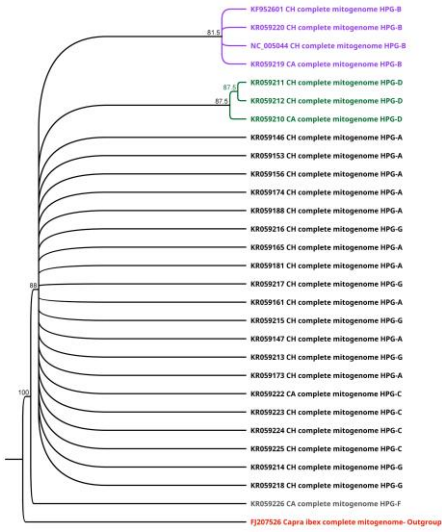

Figure S4. Differentiation of the main maternal haplogroups of domestic goats *Capra hircus* (CH) and wild goats *Capra aegagrus* (CA) based on the phylogeny of COX2 gene.

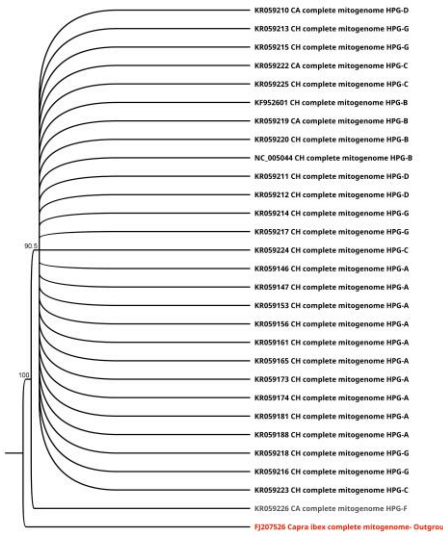

Figure S5. Differentiation of the main maternal haplogroups of domestic goats *Capra hircus* (CH) and wild goats *Capra aegagrus* (CA) based on the phylogeny of ATP8 gene.

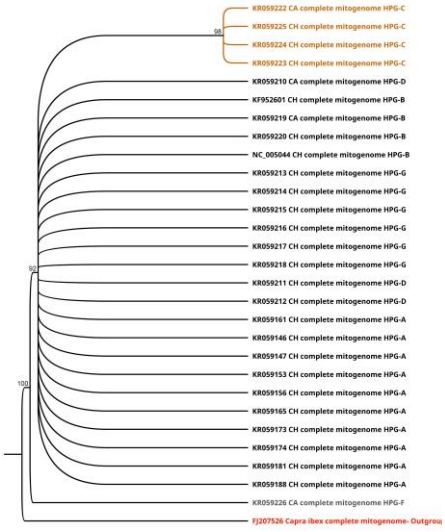

Figure S6. Differentiation of the main maternal haplogroups of domestic goats *Capra hircus* (CH) and wild goats *Capra aegagrus* (CA) based on the phylogeny of ATP6 gene.

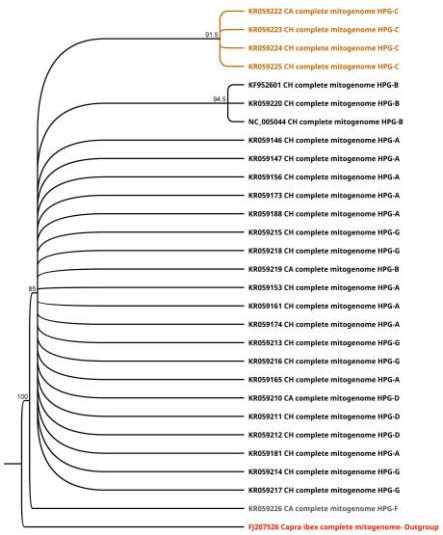

Figure S7. Differentiation of the main maternal haplogroups of domestic goats *Capra hircus* (CH) and wild goats *Capra aegagrus* (CA) based on the phylogeny of COX3 gene.

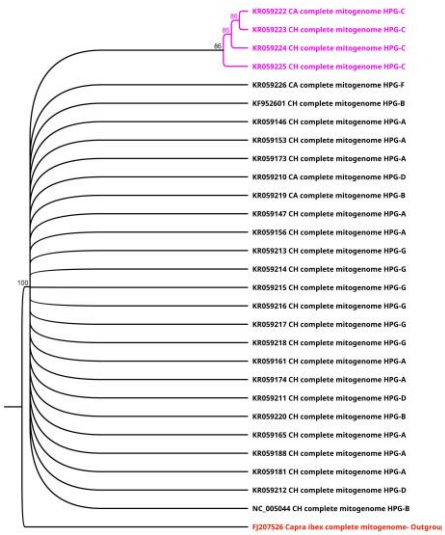

Figure S8. Differentiation of the main maternal haplogroups of domestic goats *Capra hircus* (CH) and wild goats *Capra aegagrus* (CA) based on the phylogeny of ND3 gene.

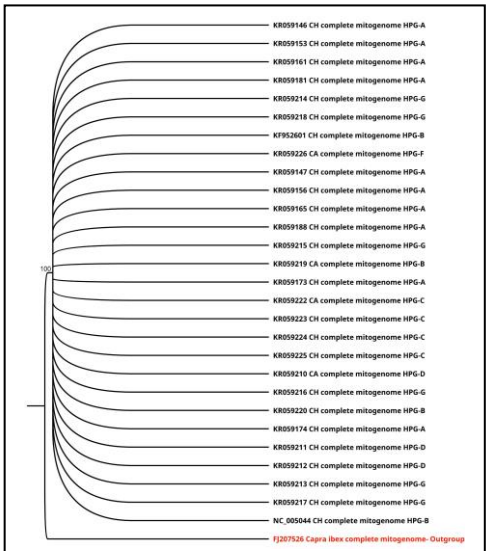

Figure S9. Differentiation of the main maternal haplogroups of domestic goats *Capra hircus* (CH) and wild goats *Capra aegagrus* (CA) based on the phylogeny of ND4L gene.

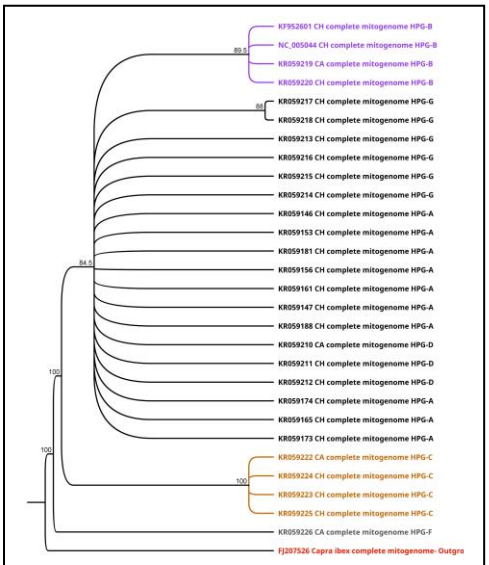

Figure S10. Differentiation of the main maternal haplogroups of domestic goats *Capra hircus* (CH) and wild goats *Capra aegagrus* (CA) based on the phylogeny of ND4 gene.

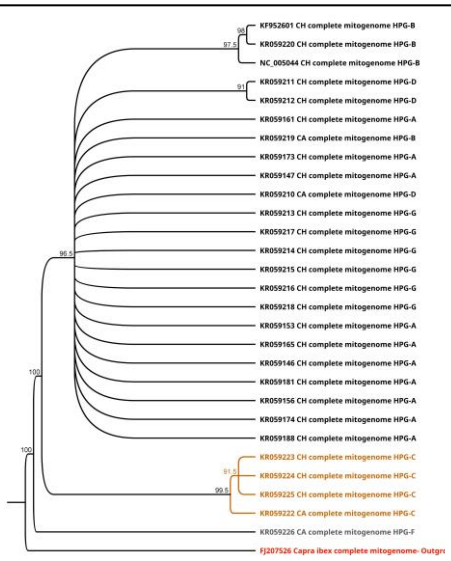

Figure S11. Differentiation of the main maternal haplogroups of domestic goats *Capra hircus* (CH) and wild goats *Capra aegagrus* (CA) based on the phylogeny of ND5 gene.

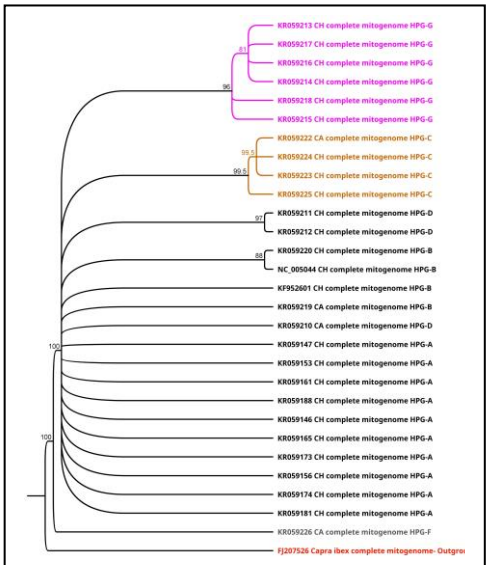

Figure S12. Differentiation of the main maternal haplogroups of domestic goats *Capra hircus* (CH) and wild goats *Capra aegagrus* (CA) based on the phylogeny of CYTB gene.

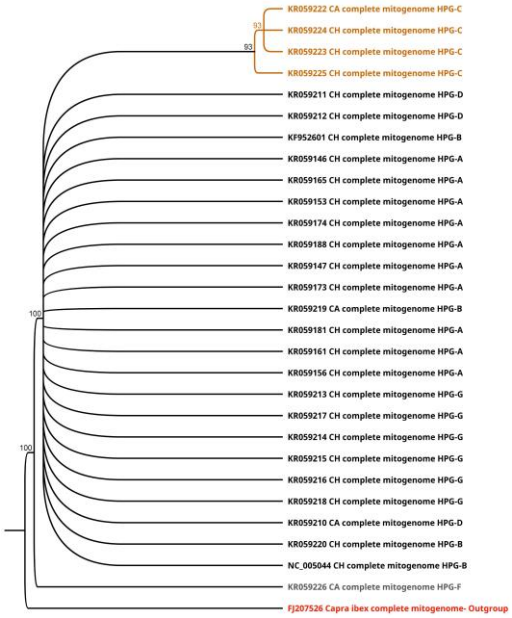

Figure S13. Differentiation of the main maternal haplogroups of domestic goats *Capra hircus* (CH) and wild goats *Capra aegagrus* (CA) based on the phylogeny of ND6 gene.

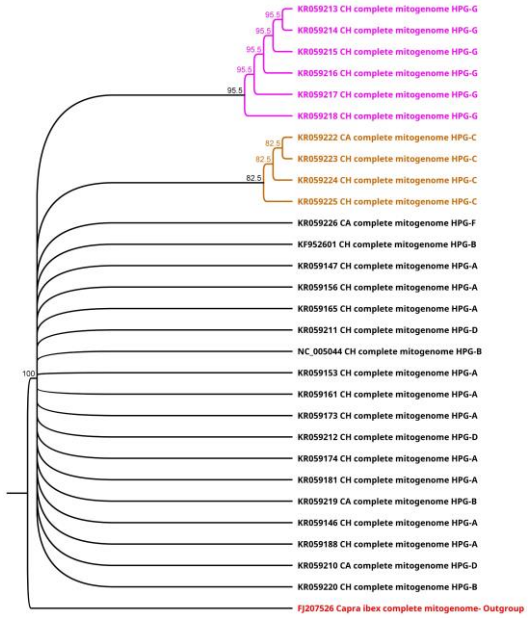

Figure S14. Differentiation of the main maternal haplogroups of domestic goats *Capra hircus* (CH) and wild goats *Capra aegagrus* (CA) based on the phylogeny of 12S rRNA gene.

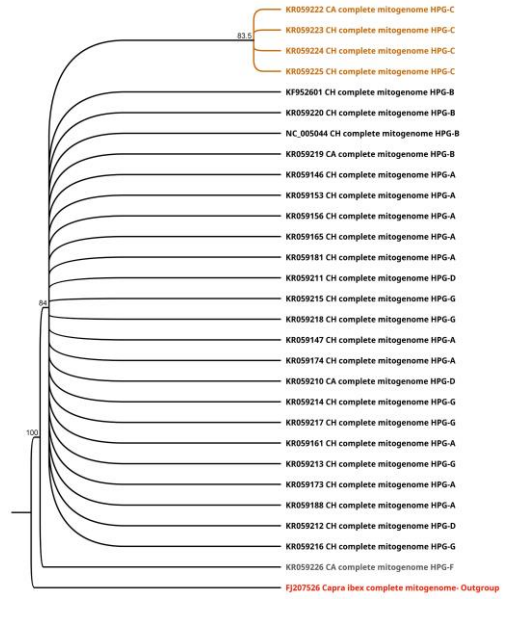

Figure S15. Differentiation of the main maternal haplogroups of domestic goats *Capra hircus* (CH) and wild goats *Capra aegagrus* (CA) based on the phylogeny of 16S rRNA gene.

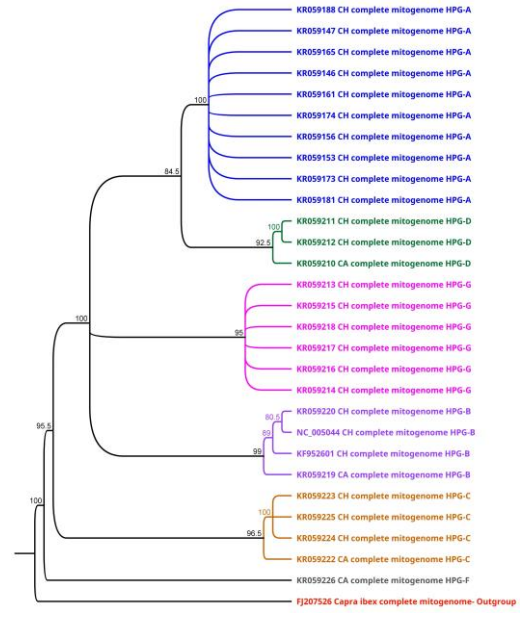

Figure S16. Differentiation of the main maternal haplogroups of domestic goats *Capra hircus* (CH) and wild goats *Capra aegagrus* (CA) based on the phylogeny of control region.