 doi:10.20944/preprints202203.0224.v1

Article

The Complete Mitochondrial Genome of a Neglected Breed, the Peruvian Creole Cattle (*Bos taurus*), and Its Phylogenetic Analysis

Carlos I. Arbizu [1,*], Rubén D. Ferro-Mauricio [1], Julio C. Chávez-Galarza [1], Héctor V. Vásquez [1], Jorge L. Maicelo [1], Carlos Poemape [1], Jhony Gonzales [2], Carlos Quilcate [1] and Flor-Anita Corredor [1]

[1] Dirección de Desarrollo Tecnológico Agrario, Instituto Nacional de Innovación Agraria (INIA), Av. La Molina 1981, Lima 15024, Peru; rubenferro20@gmail.com (R.D.F-M); jcchavezgalarza@gmail.com (J.C.Ch-G.); hvasquez@inia.gob.pe (H.V.V.); jolmqt@gmail.com (J.L.M.); cpoemape@inia.gob.pe (C.P.); promegnacional@inia.gob.pe (C.Q.); ganaderia_sdiee@inia.gob.pe (F-A.C.)

[2] Laboratorio de Biología Molecular, Universidad Nacional de Frontera, Av. San Hilarión 101, Sullana 20103, Peru; jgonzales@unf.edu.pe (J.G.)

* Correspondence: carbizu@inia.gob.pe; Tel.: +51 986288181

Simple Summary: Population of Peruvian creole cattle (PCC) is decreasing mainly due to the introduction of more productive breeds in recent years. In addition, genomics of PCC is still unknown. We sequenced for the first time the complete mitochondrial genome of a Peruvian creole cattle using a next generation sequencing technique. This genome is 16,339 bp in length and consisted of 13 protein-coding genes, two ribosomal RNA genes, 22 transfer RNA genes, and a control region. Our phylogenetic assessment with other *Bos* species depicted a close relationship of PCC with African cattle mainly, bringing further evidence that they played a role in development of the PCC. This newly generated molecular tool would serve to conduct additional modern studies on this neglected breed.

Abstract: Cattle spread throughout the American continent during the colonization years, originating creole breeds that adapted to a wide range of climate conditions. Population of creole cattle in Peru is decreasing mainly due to the introduction of more productive breeds in recent years. During the last 15 years, there have been a significant progress on cattle genomics. However, little is known about the genetics of the Peruvian creole cattle (PCC) even though its importance to (i) improve productivity in the Andean region, (ii) agricultural labor, and (iii) cultural traditions. In addition, the origin and phylogenetic relationship of the PCC is still unclear. In order to promote the conservation of the PCC, we sequenced for the first time the mitochondrial genome of a creole bull from the highlands of Arequipa, which also possessed exceptional fighting skills and was employed for agricultural tasks. The total mitochondrial genome sequence is 16,339 bp in length with the base composition of 31.43 % for A, 28.64 % for T, 26.81 % for C, and 13.12 % for G. It contains 13 protein-coding genes, two ribosomal RNA genes, 22 transfer RNA genes and a control region. Among the 37 genes, 28 were positioned on the H-strand and nine were positioned on the L-strand. The most frequently used codons were CUA (Leucine), AUA (Isoleucine), AUU (Isoleucine), AUC (Isoleucine), y ACA (Threonine). Maximum likelihood reconstruction using complete mitochondrial genome sequences clearly demonstrated that the PCC is strongly related to native African breeds, giving insights into the ancestry of PCC. The annotated mitochondrial genome of PCC would serve as an important genetic data set for further breeding work and conservation strategies.

Keywords: zoogenetic resources; organelle; genomics; NGS; cattle; *Bos taurus*

1. Introduction

 doi:10.20944/preprints202203.0224.v1

Cattle is recognized as one of the most important species for livestock, economic, and cultural influence in the world [1]. The global population of cattle (*Bos taurus* and *B. indicus*) is around 1.5 billion [2], making it one of the most common livestock. The American creole cattle (*B. taurus*) presumably descends from animals that were introduced from Iberian Peninsula in the 15th century [3,4]. The origin of cattle in Peru dates back to the history of the Spanish conquest. In the year 1521 the arrival of bovines to America begins, which after a process of establishment and evolution generated cattle adapted to new environment [5]. The current PCC could be descendant of Retinta, Berrenda, Cacereña and Andaluza negra breeds [6]. Also, a level of crossbreeding among original breeds in PCC can exist. Cattle population in Peru is about 5.5 million head, of which 63% is considered creole cattle [7]. However, that percentage is not considering the level of crossbreeding PCC possess due to the introduction and use of specialized breeds in Peru during recent years. On the other hand, M. Rosenberg (UC del Sur, pers. comm.) estimates that only 5% of cattle population in Peru corresponds to creole. PCC, in difference to other exotic breeds, have developed the ability to survive in environments with food limitations and under climatic factors that heavily affect their performance. Therefore, PCC are the basis for milk and beef production in the Peruvian highlands. In some Peruvian regions, PCC are also used for agricultural purposes and cultural events. For example, cattle from Arequipa city has been specialized for bull fighting events in local celebrations. Its breeding has been committed to the cultural manifestation of Arequipa citizens [8].

Studies on PCC are mostly limited to zoometric characterization. Espinoza and Urviola [9] determined body measurements of 140 creole cattle from Puno and concluded that the PCC can be described as a small animal size with shallow chest and short rump of little amplitude that corresponds to the type of elipometric cattle. In addition, PCC presented oblique rump and narrow pin bones. Similarly, Dipas Vargas [10] analyzed 254 creole individuals from Huamanga, Ayacucho and indicated that the zoometric indices scored distinguished them as dolichocephalic (cephalic index=45.96 ± 2.98), mesomorphic (body index=87.79±4.01), brachiothoracic (thorax index=51.72±4.95), brachypelvic (pelvic index=90.68±6.53), and dolichomorphic (depth index relative to the thorax=50.81±2.26). Moreover, Montoya [11] evaluated the phaneroptic and morphometric traits of 421 PCC in Ayacucho, Cajamarca and Puno, and depicted that the creoles from Puno showed a milk orientation and had more body development than the creoles from Ayacucho, who do not correspond to a dairy or beef orientation. He concluded that the individuals showed differences in their phaneroptic and morphometric traits according to their geographic origin, which may be used in breeding programs. In a more recent study, Encina Ruiz et al. [12] characterized the zoometry of creole cattle from the Southern Amazonas region of Peru and distinguished three biotypes of creole cattle with characteristics for meat and milk production. Molecular analyses are urgently needed to reinforce all these zoometric characterizations.

The current genomic evaluation methods have created favorable conditions to employ mitochondrial DNA as a tool for phylogenetic and biodiversity research [5]. Methods used for phylogeny verification have changed over time, from morphometric studies to molecular genetics methods. Genomic studies with the use of single nucleotide polymorphism (SNP) have been performed for the last 15 years. For instance, Hiendleder et al. [13] analyzed *B. taurus* and *B. indicus* mitochondrial genome sequences to investigate their sequence divergences and to study their taxonomic status by molecular methods. Estimated divergence times indicated that the two cattle lineages separated 1.7–2.0 million years ago. This molecular method provides new insights into intra-species taxonomy. Moreover, mitochondrial DNA studies has been carried out in American creole cattle in order to elucidate their maternal genomic origin. Lirón et al. [14] analyzed published D-loop mtDNA sequences from Creole, Iberian and African cattle breeds and identified two subclades within the African T1 haplogroup. These subclades were clearly separated between creole

 doi:10.20944/preprints202203.0224.v1

cattle from Brazil and the cattle restricted in the American region colonized by the Spaniards. They hypothesized that there were two independent sources for the origin of American creole cattle from Africa. Similarly, Bonfiglio et al. [15] agreed in an African influence for the American creole cattle. In an assessment of mtDNA from European, African, and American breeds they found six distinct sub-haplogroups. They hypothesized that seven to eight independent female lineages, belonging to haplogroup T1, were domesticated in the Near East and scattered throughout America by human migration activities. Recently, Ginja et al. [5] used mitochondrial DNA sequence data, Y-chromosome haplotype, and autosomal microsatellite markers information and showed that there was a differentiated contribution from African cattle to American creole genetic composition. Their results showed that there wass a mixed ancestry of American creole cattle and each group was genetically differentiated among each other. They also recommended that a deeply study of each creole bred would be beneficial for their conservation and use.

These molecular tools are greatly necessary in order to conduct studies for genetic distance and differentiation among cattle breeds in Peru. However, genomic tools and mitochondrial DNA studies are still in its infancy in Peru. Research is mainly focused on determining the allelic and genotypic frequency of kappa casein gene (κ-CN). Veli et al. [16] determined the allelic component of ß-lactoglobulin (BLG) in 267 creole cattle of rural communities of Ancash, Ayacucho and Puno, and indicated that the genotypic frequency of BLG^{AA} was lower to the BLG^{AB} and BLG^{BB} genotypic frequencies. In another work, Almeyda et al. [17] screened 48 creole cattle from nine villages of Bambamarca, Cajamarca and demonstrated that those individuals possessed low genotypic frequencies that favor cheese production. In a latest study, Yalta-Macedo et al. [4] identified the genetic diversity and paternal origin of Peruvian creole cattle by using seven Y-chromosome specific markers in 229 individuals from six regions of the Peruvian highlands. They depicted that PCC possessed low genetic diversity and lack of population structure. Also, their results revealed unique characteristics of PCC, and suggested they derived from the Iberian Peninsula cattle; however, the authors mentioned African cattle also had an influence on PCC.

To the best of our knowledge, this is the first time a complete mitochondrial genome of the Peruvian creole cattle was sequenced employing next-generation sequencing (NGS). We here determined the genomic features of the mitochondrial genome of the PCC and its phylogenetic relationship within the family Bovidae.

2. Materials and Methods

2.1. Sample Collection, DNA Extraction and Sequencing

A hair sample from tail was collected from a single male specimen from Andagua district, Castilla province in Arequipa (3,574 masl; -15.499548°, -72.359927°). This individual was considered an "Arequipa fighting bull" as it possessed exceptional fighting skills and was part of the traditional bullfight activity of Arequipa. It was also employed for agricultural labor.

We extracted genomic DNA with the Wizard® Genomic DNA Purification Kit (Wisconsin, USA) following the manufacturer´s instructions. The quality and quantity of genomic DNA were assessed by agarose gel electrophoresis and Qubit 2.0 Fluorometer (ThermoFisher Scientific, Waltham, MA, USA), respectively. An Illumina pair-end (2 × 150 bp) genomic library was constructed by following the standard protocol (Illumina, USA) and sequenced using an Illumina HiSeq 2500 platform by GENEWIZ (New Jersey, USA). Briefly, NEBNext® UltraTM DNA Library Prep Kit for Illumina, clustering, and sequencing reagents was used throughout the process following the manufacturer's recommendations. The genomic DNA was fragmented by acoustic shearing with a Covaris S220 instrument. Fragmented DNA was cleaned up and end repaired. Adapters were ligated after adenylation of the 3'ends followed by enrichment by limited cycle PCR. The

 doi:10.20944/preprints202203.0224.v1

DNA library was validated using a High Sensitivity D1000 ScreenTape on the Agilent TapeStation (Agilent Technologies, Palo Alto, CA, USA), and was quantified using Qubit 2.0 Fluorometer. The DNA library was also quantified by real time PCR (Applied Biosystems, Carlsbad, CA, USA). The sequencing library was clustered onto a flowcell. After clustering, the flowcell was loaded onto the Illumina instrument according to manufacturer's instructions. Image analysis and base calling were conducted by the Illumina Control Software. Raw sequence data (.bcl files) generated from the Illumina instrument was converted into fastq files using Illumina bcl2fastq 2.17 software. One mismatch was allowed for index sequence identification.

2.2. Assembly, Annotation and Sequence Analysis

Adapters and low-quality reads were removed using Trim Galore [18]. We used clean data and *Bos Taurus* (MN510465) as a reference to assemble the mitochondrial genome using GetOrganelle v1.7.2 pipeline [19], in which SPAdes v3.11.1 [20], bowtie2 v2.4.2 [21] and BLAST + v2.11 [22] were employed. The annotations of the protein-coding genes (PCGs), transfer RNAs (tRNAs) and rRNA genes from mitogenome were performed using the automatic annotators of mitochondrial genes online, Geseq in CHLOROBOX web service [23] and MITOS 2 [24], and curated manually. The tRNA secondary structure was analyzed with tRNAs-can-SE 2.0 [25]. The codon usage bias was analyzed by using MEGA X [26]. We obtained the circularized drawing of the mitochondrial genome with OGDRAW 1.3.1. [27].

2.3. Codon Usage and tRNA Analysis

Codon usage analysis was performed in Molecular Evolutionary Genetics Analysis 11 (MEGA 11) software (https://www.megasoftware.net/). For this purpose, each nucleotide sequence was cured manually, removing the STOP codon and only keeping the codons that synthesize amino acids. Then, 13 nucleotide sequences were concatenated using the Concatenate Sequence Alignment option and we evaluated the Codon Usage using RCSU option in MEGA 11. To predict the secondary structure of each tRNA, we used the tRNAscan-SE website server (http://lowelab.ucsc.edu/tRNAscan-SE/). We analyzed the sequences in FASTA format and kept the parameters by default except the Sequence Source that corresponds to Mammalian mitochondrial.

2.4. Phylogenetic Analysis

We employed 55 mitochondrial genomes of other *Bos* species available in GenBank to determine the genetic relationship of PCC (Table S1). As outgroup, we used a species of the genus Bison (*Bison bison*) from the same subfamily Bovinae. Each genome was aligned by MAFFT v7.475 [28] and with a GTR+GAMMA model of evolution, we obtained the best-scoring maximum likelihood (ML) tree, and then 1,000 nonparametric bootstrap inferences were performed with RAxML v8.2.11 [29]. The resulting trees were viewed in FigTree version 1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/).

3. Results

3.1. Mitochondrial genome organization

The complete assembled mitochondrial genome of the Peruvian creole cattle (*Bos taurus*) was 16,339 bp in length and consisted of 22 tRNA genes, 13 protein-coding genes, two rRNA genes, and a control region (Table 1, Figure 1). Most of the genes (28) were encoded within the heavy (H) strand, and nine genes were positioned on the L-strand. The base composition of this genome was 31.43% A, 28.64% T, 26.81% C, and 13.12% G. The D-loop region is located between the $tRNA^{Pro}$ and $tRNA^{Phe}$ genes. In the range from 1 to 40 bp, there were a total of 78 overlapping nucleotides at seven different regions. The largest overlapping region (40 bp) was emplaced between the *Atp8* and *Atp6* genes. Furthermore, the intergenic spacer (IGS) was comprised of 14 regions across the mitochondrial genome

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 March 2022 doi:10.20944/preprints202203.0224.v1

ranging from 1 to 7 bp, which summed up to a total length of 30 bp. The largest intergenic spacer (7 bp) was located between *tRNA*^{Ser2} and *tRNA*^{Asp} genes. The entire mitochondrial genome sequence was submitted to the GenBank database with accession number: OK135155. The associated Bioproject, Biosample, and SRA numbers are PRJNA763011, SAMN21419641, and SRR15883111, respectively.

Table 1. Gene organization of mitochondrial genome of Peruvian creole cattle. The (+) and (−) values correspond to intergenic nucleotides and overlapping regions between the genes, respectively.

Gene	Nucleotide Positions	Size (bp)	Strand[1]	Codon	Intergenic Spacer (bp)
tRNA^{Phe}	364-430	67	H	TTC	
12S rRNA	431-1386	956	H		
tRNA^{Val}	1387-1453	67	H	GTA	
16S rRNA	1454-3023	1570	H		
tRNA^{Leu2}	3025-3099	75	H	TTA	1
Nd1	3102-4057	956	H		2
tRNA^{Ile}	4058-4126	69	H	ATC	
tRNA^{Gln}	4124-4195	72	L	CAA	-3
tRNA^{Met}	4198-4266	69	H	ATG	2
Nd2	4267-5309	1043	H		
tRNA^{Trp}	5309-5375	67	H	TGA	-1
tRNA^{Ala}	5377-5445	69	L	GCA	1
tRNA^{Asn}	5447-5519	73	L	AAC	1
Rep_origin	5522-5552	31	H		2
tRNA^{Cys}	5552-5618	67	L	TGC	-1
tRNA^{Tyr}	5619-5686	68	L	TAC	
Cox1	5688-7232	1545	H		1
tRNA^{Ser2}	7230-7298	69	L	TCA	-3
tRNA^{Asp}	7306-7373	68	H	GAC	7
Cox2	7375-8058	684	H		1
tRNA^{Lys}	8062-8128	67	H	AAA	3
Atp8	8130-8330	201	H		1
Atp6	8291-8971	681	H		-40
Cox3	8971-9755	785	H		-1
tRNA^{Gly}	9755-9823	69	H	GGA	-1
Nd3	9821-10170	350	H		-3
tRNA^{Arg}	10171-10239	69	H	CGA	
Nd4L	10240-10536	297	H		
Nd4	10530-11907	1378	H		-7
tRNA^{His}	11908-11977	70	H	CAC	
tRNA^{Ser}	11978-12037	60	H	AGC	
tRNA^{Leu}	12039-12109	71	H	CTA	1
Nd5	12110-13930	1821	H		

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 March 2022 doi:10.20944/preprints202203.0224.v1

Nd6	13914-14441	528	L		-17
tRNA^{Glu}	14442-14510	69	L	GAA	
CytB	14515-15654	1140	H		4
tRNA^{Thr}	15658-15727	70	H		3
tRNA^{Pro}	15727-15792	66	L	CCA	-1
D-loop	15793-16339, 1-363	910	H		

[1]Strand: H (Heavy), L (Light)

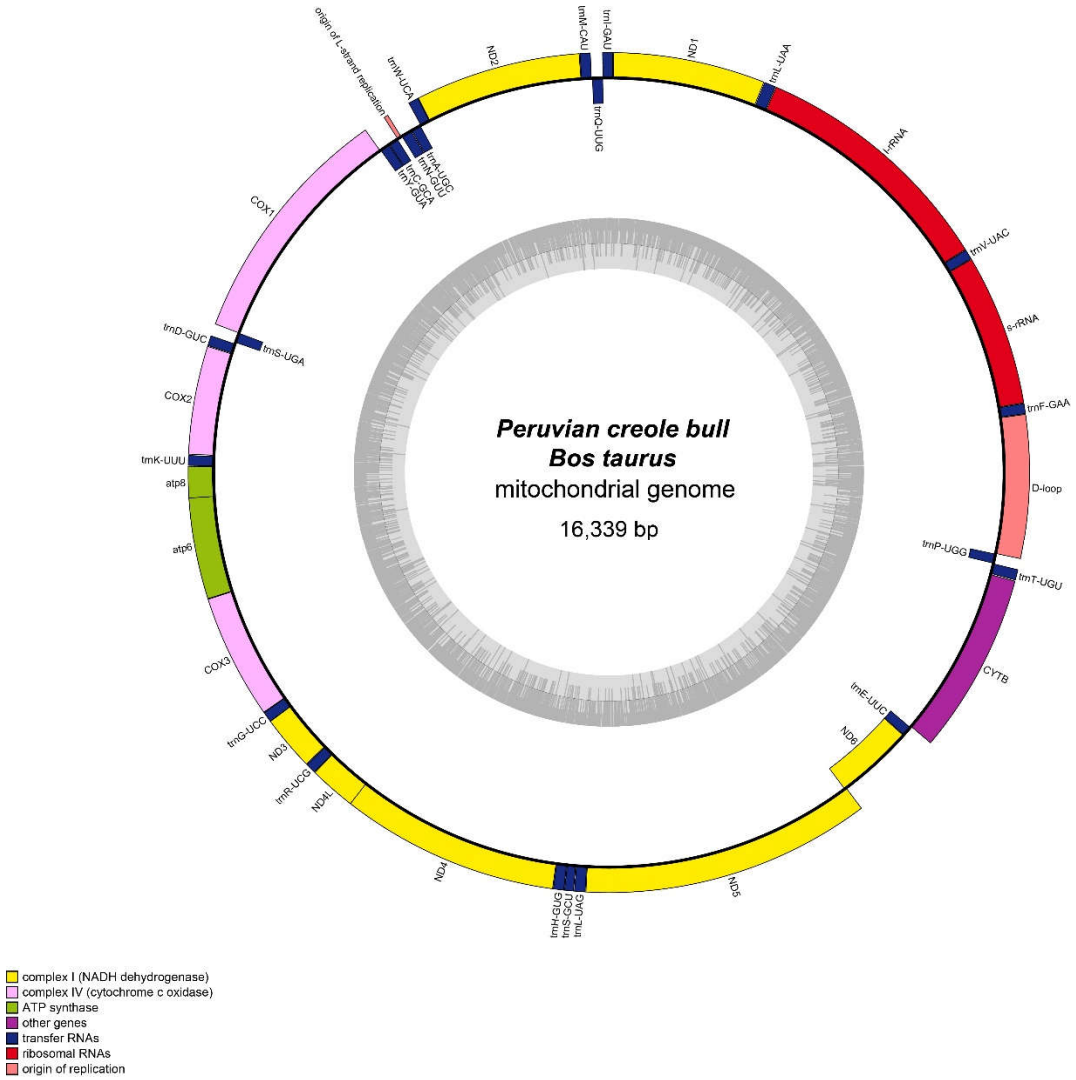

Figure 1. The mitochondrial genome map of the Peruvian creole cattle. Genes encoded by the heavy strand are shown outside the circle, while those encoded by the light strand are shown inside.

3.2. Protein Coding Genes (PCGs) and Codon Usage

A total of 13 PCGs were encoded in the PCC mitogenome and were 11,409 bp in length, representing 69.83% of this genome. This mitochondrial genome encode 3792 amino acids. These PCGs were represented by (i) seven NADH dehydrogenase subunits, (ii) two ATPase subunits, and (iii) a cytochrome *b* gene. These PCGs are AT-biased as AT content ranges from 55.9% for *Cox3* to 68.2% for *Atp8* gene. Moreover, the length of PCGs

 doi:10.20944/preprints202203.0224.v1

varied greatly from *atp8* (201 bp) to *Nad5* (1821 bp) (Table 2). The most abundant start and stop codons were ATG and TAA, respectively. On the other hand, genes *Nd1*, *Nd2*, *Cox3*, *Nd3*, and *Nd4* presented incomplete stop codons (TA- or T--). Length of the 13 genes varied from 201 bp (*Atp8*) to 1821 bp (*Nd5*). Similarly, protein length (aa) ranged from 66 to 606 for these two genes. The five codons with the highest relative synonymous codon usage (RSCU) values in the PCGs were: CUA (2.86), CGA (2.67), UCA (2.15), ACA (1.99) and GUA (1.83) (Figure 2A). CUA (Leucine), AUA (Isoleucine), AUU (Isoleucine), AUC (Isoleucine) and ACA (Threonine) were the most frequently used codons (Figure 2B).

Table 2. Characteristics of protein coding genes identified in the mitochondrial genome of Peruvian creole cattle.

Gene	Gene Length (bp)	A + T Content (%)	Start / Stop Codon	Protein Length (aa)
Nd1	956	59.4	ATG / TA-	318
Nd2	1043	64.6	ATA / TA-	347
Cox1	1545	58.3	ATG / TAA	514
Cox2	684	61.7	ATG / TAA	227
Atp8	201	68.2	ATG / TAA	66
Atp6	681	61.5	ATG / TAA	226
Cox3	785	55.9	ATG / TA-	261
Nd3	350	58.0	ATA / TA-	116
Nd4L	297	63.9	ATG / TAA	98
Nd4	1378	60.9	ATG / T--	459
Nd5	1821	60.3	ATA / TAA	606
Nd6	528	63.1	ATG / TAA	175
CytB	1140	56.4	ATG / AGA	379
Total	11,409			3,792

3.3. Ribosomal RNA, Transfer RNA and Non-coding Regions

The total size of the two rRNA genes (*12S* and *16S*) was 2526 bp and they were flanked by $tRNA^{Phe}$ and $tRNA^{Leu2}$ (Table 1; Figure 1). We identified 22 *tRNA* genes with a total size of 1511 bp, which varied in length from 60 ($tRNA^{Ser}$) to 75 ($tRNA^{Leu2}$) bp (Table 1). The H-strand contained 14 *tRNA* genes, and eight *tRNA* genes were encoded by the L-strand. All *tRNA* genes exhibited the cloverleaf secondary structure except two of them, $tRNA^{Lys}$ and $tRNA^{Ser}$ (Figure 3). Non-coding regions comprised the origin of replication, intergenic spacers and the control region. The origin of replication possessed 31 bp and was located between $tRNA^{Asn}$ and $tRNA^{Cys}$. In addition, with a total length of 910 bp, the control region (D-loop) was flanked by $tRNA^{Pro}$ and $tRNA^{Phe}$ genes (Table 1, Figure 1).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 March 2022 doi:10.20944/preprints202203.0224.v1

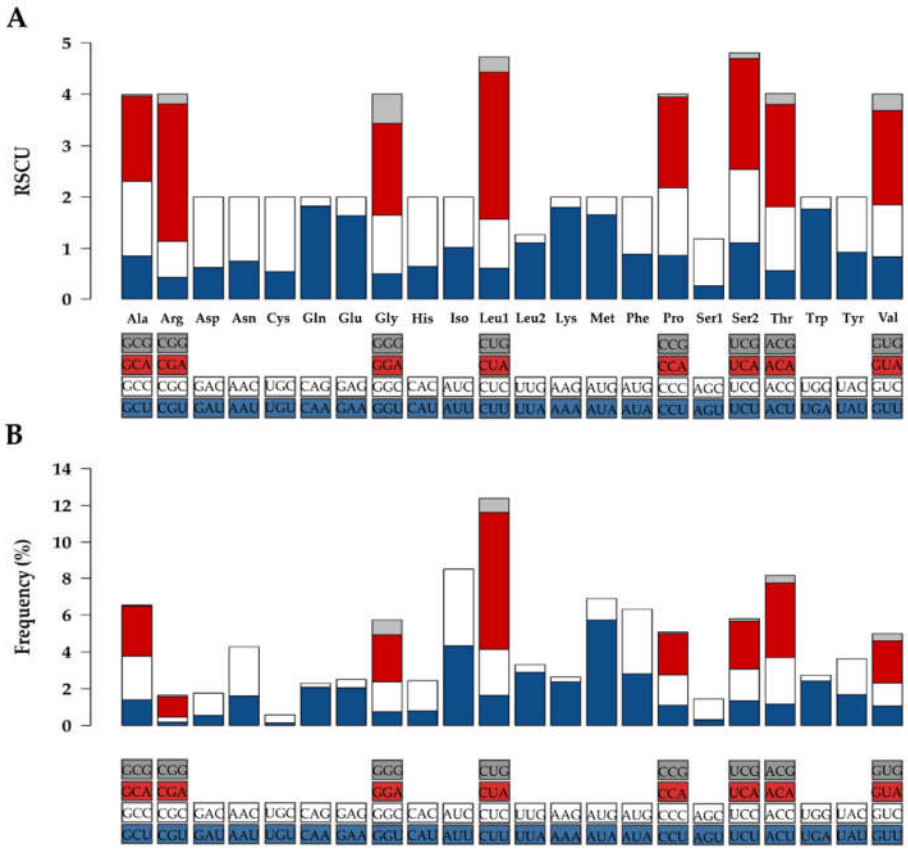

Figure 2. Codon usage of the mitochondrial genome protein coding genes of the Peruvian creole cattle. **A.** Relative synonymous codon usage (RSCU). **B.** Codon usage frequency. Codon families are plotted on the X axis and represented by different colors.

3.4. Phylogenetic Inference

Our maximum likelihood (ML) phylogenetic tree topology showed three well-supported clades. Species *B. taurus* is placed within a monophyletic clade and is sister to a clade formed by *B. primigenius*; and sister to them, a clade formed by *B. indicus*. Members of *B. gaurus* and *B. frontalis* were placed within clade 2. In addition, *B. grunniens* and *B. mutus* form clade 3, and is sister to clades 1 and 2 (Figure 4). ML analysis revealed that within clade 1, members of *B. taurus* are placed in two subclades, 1 and 2. Peruvian creole cattle was placed in subclade 1 together with other creole cattle from Paraguay, and Mexico, and with individuals from Africa mainly. On the other hand, subclade 2 was comprised mainly of cattle from Europe that are mostly specialized for beef and milk production (Figure 4, Table S1).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 March 2022 doi:10.20944/preprints202203.0224.v1

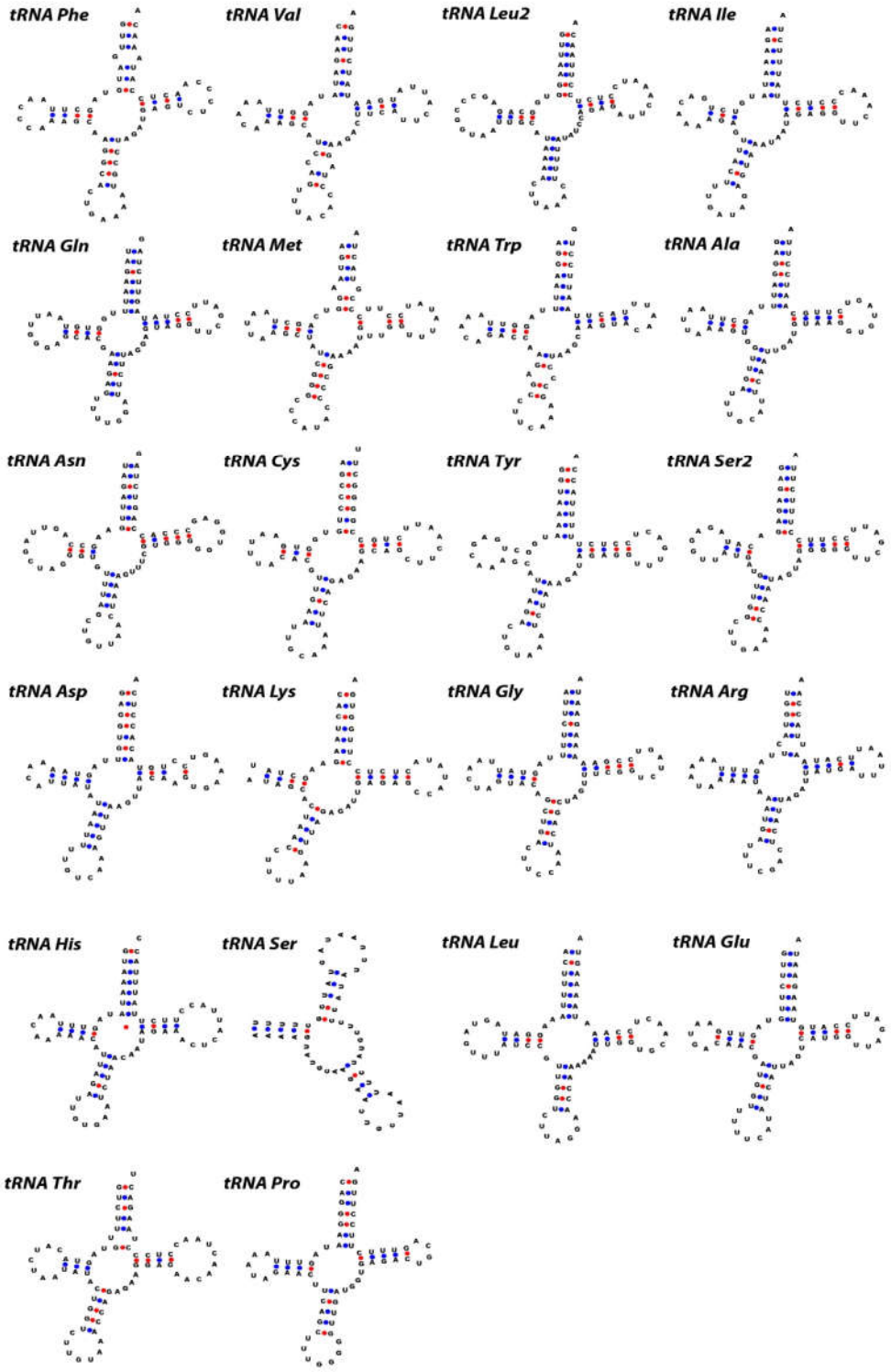

Figure 3. The predicted secondary structures of 22 transfer RNA genes of Peruvian creole cattle.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 16 March 2022 doi:10.20944/preprints202203.0224.v1

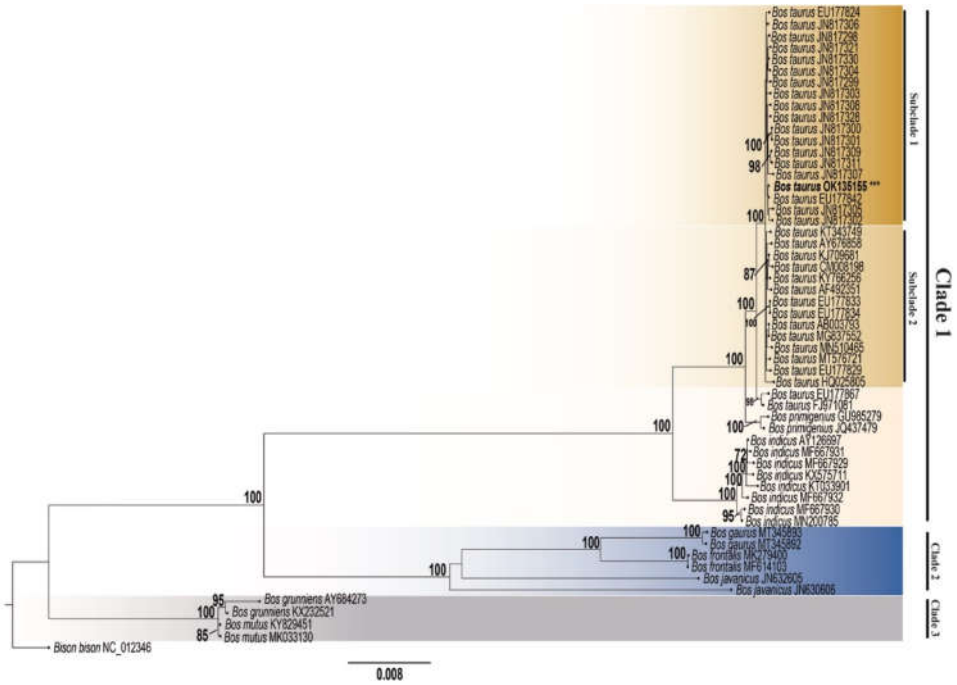

Figure 4. Maximum likelihood phylogenetic tree of *Bos* species inferred from mitochondrial genomic sequences. Bootstrap support is shown only for branches with >70% support. *Bison bison* was set as the outgroup. The PCC is designated by three asterisks.

4. Discussion

During the last five centuries, creole cattle went through multiple random crosses, conferring not only zoometric characteristics diversity, but also adaptive traits that allowed them to survive in different environmental conditions [30]. Due to the introduction of specialized breeds, now the Peruvian creole cattle (PCC) are mainly distributed in rural communities in southern Peru (Ayacucho, Arequipa, Huancavelica and Apurimac) (M. Rosenberg, UC del Sur, pers. comm.) and they represent a significant protein source, agricultural labor force, and are part of cultural traditions. Due to the reduction of sequencing costs, livestock geneticists have sequenced over 2,500 cattle genomes [31]. Nevertheless, PCC is still a neglected breed as they have been poorly studied. Thus, in the present work, we sequenced and characterized for the first time the PPC mitochondrial genome, which was 16,339 bp in length. This size is in agreement with mitogenomes of other individuals of *Bos taurus* (16,338 bp), *B. indicus* (16,339 bp), *B. frontalis* (16,346 bp), *B. grunniens* (16,324 bp) and *B. gaurus* (16,345 bp) [13,32–34]. Our result confirmed that mitogenomes in the Bovinae subfamily are very similar in size. In addition, gene order and structure were similar to previously reported *B. taurus* mitogenomes [35,36]. The AT and GC content was 60.07% and 39.93%, respectively, revealing a nucleotide composition bias of the PCC mitogenome towards adenine and thymine, which is commonly observed in other mammals [32,37,38].

Indices of codon usage bias are used to measure differences in the occurrence of codon usage that indicate the evolutionary patterns of their genome [39]. Start and stop codons ATG and TAA, respectively were the most abundant in PCC mitogenome, which is in agreement with the mitochondrial genome of other mammals [38,40,41]. The presence of incomplete stop codons is a common phenomenon found in mitogenomes of other vertebrates [42–44]. These codons might be further completed by the posttranscriptional

 doi:10.20944/preprints202203.0224.v1

poly-adenylation of the 3′-end of the mRNA [45,46]. Codon usage bias is a known phenomenon that occurs in a wide variety of organisms. Reporting codon use bias for the first time in PCC provides us useful information about mutation frequency, GC composition gene expression level, and abundance of tRNA. Codon usage can be determined based on relative synonymous codon usage (RCSU) index, which is calculated as the ratio of the observed frequency of a codon to the expected frequency of that codon, assuming uniform codon usage [47]. The synonymous codons with RSCU values of 1 indicate no codon usage bias for that amino acid and the codons are chosen equally or randomly. RSCU values above and below 1 indicate positive and negative codon usage bias, respectively [47,48].

Non-coding regions are known to contain controlling elements for replication and transcription [49], and include L-strand origin of replication, intergenic spacers and control region [50]. The organization of the origin of replication of the PCC mitogenome is similar to other mammal species [38,51,52]. The length of the control region of the PCC mitochondrial genomes was 910 bp, a finding that agrees with previous work on other *Bos* mitogenomes [32,33].

In order to understand the phylogenetic relationship of the Peruvian creole cattle, a maximum likelihood (ML) phylogenetic tree was constructed along with other cattle breeds and *Bos* species. Our entire mitochondrial genome analysis provided a highly supported topology of the *Bos* genus, as reported by Kamalakkannan et al. [33]. ML phylogenetic tree showed three clades, and clade 1 was further divided in two subclades. The PCC is distributed in subclade 1 together with cattle from Paraguay, Mexico and Africa mainly. Interestingly, the PCC + an Italian breed (Cinisara) were sister to a clade formed by two African indicine breeds, Boran and Arsi. Porter et al [53] indicated that Boran breed underwent selection and improvement with European taurine cattle in the early twentieth century. In addition, Ginja et al. [5] used autosomal microsatellites to demonstrate that American creole cattle occupied an intermediate position between African and European breeds. However, they did not used individuals from Peru. Therefore, it is very likely that African cattle played a role in the development of PCC, as showed by Yalta-Macedo [4] with Y-chromosome specific markers. This may be explained by the influential role that African cattle have had for the development of Iberian breeds or imported cattle directly from Africa to Peru.

Mitogenomes possess limitations such as maternal inheritance, accelerated rates of substitution, introgression, neutrality and effective population size [54], variation in rates of gene loss accompanied by functional transfer to the nucleus, and in rated of genome rearrangement [55]. Therefore, the phylogenetic results of the PCC in this work need to be validated by other nuclear markers. Our next step is to continue developing molecular tools for the PCC. Further research is needed aiming at identifying the putative genes of the PCC that confer resilience in adverse climate conditions and high nutritional conversion with limited food sources. These molecular tools may shed light on developing a modern cattle breeding program in Peru by exploiting the alleles that possess this neglected breed. Moreover, a sustainable local management and implementation of conservation programs of the PCC is urgently needed.

5. Conclusions

Here, we first reported the complete mitochondrial genome of a neglected breed, the Peruvian creole cattle. Our results showed that the organization and characteristics of this mitogenome is in agreement with the mitochondrial genomes of other mammals, possessing 16,339 bp in length and 13 protein-coding genes, two ribosomal RNA genes, 22 transfer RNA genes and a control region. Moreover, a total of 28 genes were positioned on the H-strand, and 9 on the L-strand. Phylogenetic analysis showed PCC is closely related to native African breeds. However, more work is needed employing nuclear markers and additional individuals of creole cattle from Peru. We hope this work will help pave

the way toward establishing conservation policies and conducting more modern research on the PCC.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1, Table S1: Species, breed, origin, accession code and clade assignment of the 56 individuals employed in this study.

Author Contributions: Conceptualization, C.I.A., C. Q. and F.-A.C; Data curation, C.I.A.; Formal analysis, C.I.A., R.D.F.-M. and J.C.Ch.-G.; Funding acquisition, H.V.V., J.L.M., J.G. and C.Q.; Investigation, F.-A.C.; Methodology, R.D.F.-M. and J.C.Ch.-G.; Project administration, J.L.M and C.P.; Supervision, H.V.V., J.L.M and F.-A.C.; Writing – original draft, C.I.A. and F.-A.C.; Writing – review & editing, C.I.A., R.D.F.-M., J.C.Ch.-G., H.V.V., J.L.M, C.P., J.G., C.Q. and F.-A.C.

Funding: This research was funded by the following two research projects: "Creación del servicio de agricultura de precisión en los Departamentos de Lambayeque, Huancavelica, Ucayali y San Martín 4 Departamentos" and "Mejoramiento de la disponibilidad de material genético de ganado bovino con alto valor a nivel nacional 7 departamentos" of the Ministry of Agrarian Development and Irrigation (MIDAGRI) of the Peruvian Government, with grant number CUI 2449640 and CUI 2432072, respectively. C.I.A. was supported by PP0068 "Reducción de la vulnerabilidad y atención de emergencias por desastres".

Institutional Review Board Statement: The sample collection from cattle specimen was conducted in accordance to the Peruvian National Law No. 30407: "Animal Protection and Welfare."

Informed Consent Statement: Written informed consent was obtained from the owner of the bull studied here.

Data Availability Statement: All data generated during this study are included in this published article.

Acknowledgments: We express our special appreciation and thank to Family Ojeda from Arequipa who kindly provided hair samples of their bull for this work. Also, we would like to thank Ivan Ucharima, Maria Angélica Puyo, Cristina Aybar, and Erick Rodriguez for supporting the logistic activities in the laboratory. Finally, the authors thank the Bioinformatics High-performance Computing server of Universidad Nacional Agraria la Molina for providing resources for data analysis.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Herrero, M.; Grace, D.; Njuki, J.; Johnson, N.; Enahoro, D.; Silvestri, S.; Rufino, M.C. The Roles of Livestock in Developing Countries. *Animal* **2013**, *7*, 3–18.
2. McLeod, A. World Livestock 2011 - Livestock in Food Security. In *World livestock 2011 - livestock in food security*; Food and Agriculture Organization of the United Nations (FAO), 2011.
3. Martínez, A.M.; Gama, L.T.; Cañón, J.; Ginja, C.; Delgado, J. V.; Dunner, S.; Landi, V.; Martín-Burriel, I.; Penedo, M.C.T.; Rodellar, C.; et al. Genetic Footprints of Iberian Cattle in America 500 Years after the Arrival of Columbus. *PLoS One* **2012**, *7*, 1–13.
4. Yalta-Macedo, C.E.; Veli, E.A.; Díaz, G.R.; Vallejo-Trujillo, A. Paternal Ancestry of Peruvian Creole Cattle Inferred from Y-Chromosome Analysis. *Livest. Sci.* **2021**, *244*, 104376.
5. Ginja, C.; Gama, L.T.; Cortés, O.; Burriel, I.M.; Vega-Pla, J.L.; Penedo, C.; Sponenberg, P.; Cañón, J.; Sanz, A.; do Egito, A.A.; et al. The Genetic Ancestry of American Creole Cattle Inferred from Uniparental and Autosomal Genetic Markers. *Sci. Reports 2019 91* **2019**, *9*, 1–16.
6. Villalobos-Cortés, A.; Martínez, A.; Vega-Pla, J.L.; Landi, V.; Quiroz, J.; Martínez, R.; López, R.M.; Sponenberg, P.; Armstrong, E.; Zambrano, D.; et al. Relationships between Panamanians and Some Creole Cattle Landraces in Latin America. *Pesqui. Agropecu. Bras.* **2012**, *47*, 1637–1646.

7. Instituto Nacional de Estadística e Informática IV Censo Nacional Agropecuario 2012 Available online: http://censos.inei.gob.pe/Cenagro/redatam/# (accessed on 14 March 2022).
8. Motta Zamalloa, E. El Astero de Plata. *Anthropologica* **1988**, *6*, 259–283.
9. Espinoza, R.; Urviola, G. Biometría y Constantes Clínicas Del Bovino Criollo En El Centro de Investigación y Producción Chuquibambilla de Puno (Perú). *Arch. Zootec.* **2005**, *54*, 233–236.
10. Dipas Vargas, E.S. Zoometría e Índices Corporales Del Vacuno Criollo En El Matadero de Quicapata de La Provincia de Huamanga, a 2720 m.s.n.m. Ayacucho - 2014, Universidad Nacional San Cristóbal de Huamanga, 2015.
11. Montoya, M.J. Caracterización Faneróptica y Morfométrica Del Vacuno Criollo En Ayacucho, Puno y Cajamarca, 2016.
12. Ruiz, R.E.; Saucedo-uriarte, J.A.; Portocarrero-villegas, S.M.; Quispe-ccasa, H.A.; Cayo-colca, I.S. Zoometric Characterization of Creole Cows from the Southern Amazon Region of Peru. *Diversity* **2021**, *13*.
13. Hiendleder, S.; Lewalski, H.; Janke, A. Complete Mitochondrial Genomes of *Bos taurus* and *Bos indicus* Provide New Insights into Intra-Species Variation, Taxonomy and Domestication. *Cytogenet Genome Res* **2008**, *120*, 150–156.
14. Lirón, J.P.; Bravi, C.M.; Mirol, P.M.; Peral-García, P.; Giovambattista, G. African Matrilineages in American Creole Cattle: Evidence of Two Independent Continental Sources. *Anim. Genet.* **2006**, *37*, 379–382.
15. Bonfiglio, S.; Ginja, C.; Gaetano, A. De; Achilli, A.; Olivieri, A.; Colli, L.; Tesfaye, K.; Agha, S.H.; Gama, L.T.; Cattonaro, F.; et al. Origin and Spread of Bos Taurus: New Clues from Mitochondrial Genomes Belonging to Haplogroup T1. *PLoS One* **2012**, *7*, e38601.
16. Veli, E.A.; Rivas Seoane, E.; Rivas Palma, V.; Aquino, Y.; Estrada, R. Variabilidad Genética Del Gen de Beta Lactoglobulina En Bovinos Criollos de Perú. *Arch. Zootec.* **2008**, *57*, 341–344.
17. Almeyda, R.M.; Rosadio, A.R.; Maturrano, H.L. Genotypes of Kappa-Casein Gen in Creole Cattle of Bambamarca District (Cajamarca, Peru). *Rev. Investig. Vet. del Peru* **2016**, *27*, 82–90.
18. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.J.* **2013**, *27*, 10-12.
19. Jin, J.J.; Yu, W. Bin; Yang, J.B.; Song, Y.; Depamphilis, C.W.; Yi, T.S.; Li, D.Z. GetOrganelle: A Fast and Versatile Toolkit for Accurate de Novo Assembly of Organelle Genomes. *Genome Biol.* **2020**, *21*, 1–31.
20. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477.
21. Langmead, B.; Salzberg, S.L. Fast Gapped-Read Alignment with Bowtie 2. *Nat. Methods* **2012**, *9*, 357–359.
22. Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K.; Madden, T.L. BLAST+: Architecture and Applications. *BMC Bioinformatics* **2009**, *10*, 1–9.
23. Tillich, M.; Lehwark, P.; Pellizzer, T.; Ulbricht-Jones, E.S.; Fischer, A.; Bock, R.; Greiner, S. GeSeq - Versatile and Accurate Annotation of Organelle Genomes. *Nucleic Acids Res.* **2017**, *45*, W6–W11.
24. Donath, A.; Jühling, F.; Al-Arab, M.; Bernhart, S.H.; Reinhardt, F.; Stadler, P.F.; Middendorf, M.; Bernt, M. Improved Annotation of Protein-Coding Genes Boundaries in Metazoan Mitochondrial Genomes. *Nucleic Acids Res.* **2019**, *47*, 10543–10552.
25. Chan, P.P.; Lin, B.Y.; Mak, A.J.; Lowe, T.M. TRNAscan-SE 2.0: Improved Detection and Functional Classification of Transfer RNA Genes. *Nucleic Acids Res.* **2021**, *49*, 9077–9096.
26. Kumar, S.; Stecher, G.; Li, M.; Knyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549.
27. Greiner, S.; Lehwark, P.; Bock, R. OrganellarGenomeDRAW (OGDRAW) Version 1.3.1: Expanded Toolkit for the Graphical Visualization of Organellar Genomes. *Nucleic Acids Res.* **2019**, *47*, W59–W64.
28. Katoh, K.; Standley, D.M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780.

29. Stamatakis, A. RAxML Version 8: A Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinforma. Appl.* **2014**, *30*, 1312–1313.
30. Ginja, C.; Penedo, M.C.T.; Melucci, L.; Quiroz, J.; Martínez López, O.R.; Revidatti, M.A.; Martínez-Martínez, A.; Delgado, J. V.; Gama, L.T. Origins and Genetic Diversity of New World Creole Cattle: Inferences from Mitochondrial and Y Chromosome Polymorphisms. *Anim. Genet.* **2010**, *41*, 128–141.
31. Georges, M.; Charlier, C.; Hayes, B. Harnessing Genomic Information for Livestock Improvement. *Nat. Rev. Genet.* **2019**, *20*, 135–156.
32. Prabhu, V.R.; Arjun, M.S.; Bhavana, K.; Kamalakkannan, R.; Nagarajan, M. Complete Mitochondrial Genome of Indian Mithun, Bos Frontalis and Its Phylogenetic Implications. *Mol. Biol. Rep.* **2019**, *46*, 2561–2566.
33. Kamalakkannan, R.; Bhavana, K.; Prabhu, V.R.; Sureshgopi, D.; Singha, H.S.; Nagarajan, M. The Complete Mitochondrial Genome of Indian Gaur, Bos Gaurus and Its Phylogenetic Implications. *Sci. Rep.* **2020**, *10*, 1–11.
34. Chu, M.; Wu, X.; Liang, C.; Pei, J.; Ding, X.; Guo, X.; Bao, P.; Yan, P. The Complete Sequence of Mitochondrial Genome of Polled Yak (Bos Grunniens). *Mitochondrial DNA. Part A, DNA mapping, Seq. Anal.* **2016**, *27*, 2032–2033.
35. Wang, X.; Zhang, Y.Q.; He, D.C.; Yang, X.M.; Li, B.; Wang, D.C.; Guang, J.; Xu, F.; Li, J.Y.; Gao, X.; et al. The Complete Mitochondrial Genome of Bos Taurus Coreanae (Korean Native Cattle). *Mitochondrial DNA* **2016**, *27*, 120–121.
36. Liu, S.-J.; Lv, J.-Z.; Tan, Z.-Y.; Ge, X.-Y. The Complete Mitochondrial Genome of Uruguayan Native Cattle (*Bos taurus*). *Mitochondrial DNA Part B Resour.* **2020**, *5*, 443-444.
37. Gutiérrez, J.; Lamelas, L.; Aleix-Mata, G.; Arroyo, M.; Marchal, J.A.; Palomeque, T.; Lorite, P.; Sánchez, A. Complete Mitochondrial Genome of the Iberian Mole *Talpa Occidentalis* (Talpidae, Insectivora) and Comparison with *Talpa Europaea*. *Genetica* **2018**, *146*, 415–423.
38. Lamelas, L.; Aleix-Mata, G.; Rovatsos, M.; Marchal, J.A.; Palomeque, T.; Lorite, P.; Sánchez, A. Complete Mitochondrial Genome of Three Species of the Genus *Microtus* (Arvicolinae, Rodentia). **2020**, *10*, 2130.
39. Paul, P.; Malakar, A.K.; Chakraborty, S. Codon Usage Vis-a-Vis Start and Stop Codon Context Analysis of Three Dicot Species. *J. Genet.* **2018**, *97*, 97–107.
40. Xu, Y.; Huang, X.; Hu, Y.; Tu, F. Description of the Mitogenome of Gansu Mole (*Scapanulus oweni*). *Mitochondrial DNA. Part A, DNA mapping, Seq. Anal.* **2016**, *27*, 2083–2084.
41. Kim, N.H.; Lim, S.J.; Chae, H.M.; Park, Y.C. Complete Mitochondrial Genome of the Amur Hedgehog *Erinaceus amurensis* (Erinaceidae) and Higher Phylogeny of the Family Erinaceidae. *Genet. Mol. Res.* **2017**, *16*, 1–8.
42. Hou, W.R.; Chen, Y.; Wu, X.; Hu, J.C.; Peng, Z.S.; Yang, J.; Tang, Z.X.; Zhou, C.Q.; Li, Y.M.; Yang, S.K.; et al. A Complete Mitochondrial Genome Sequence of Asian Black Bear Sichuan Subspecies (*Ursus Thibetanus Mupinensis*). *Int. J. Biol. Sci.* **2007**, *3*, 85–90.
43. Oh, D.J.; Kim, J.Y.; Lee, J.A.; Yoon, W.J.; Park, S.Y.; Jung, Y.H. Complete Mitochondrial Genome of the Rabbitfish *Siganus fuscescens* (Perciformes, Siganidae). *DNA Seq. - J. DNA Seq. Mapp.* **2007**, *18*, 295–301.
44. Prabhu, V.R.; Singha, H.S.; Kumar, R.G.; Gopalakrishnan, A.; Nagarajan, M. Characterization of the Complete Mitochondrial Genome of *Barilius malabaricus* and Its Phylogenetic Implications. *Genomics* **2020**, *112*, 2154–2163.
45. Ojala, D.; Montoya, J.; Attardi, G. tRNA Punctuation Model of RNA Processing in Human Mitochondria. *Nature* **1981**, *290*, 470–474.
46. Cha, S.Y.; Yoon, H.J.; Lee, E.M.; Yoon, M.H.; Hwang, J.S.; Jin, B.R.; Han, Y.S.; Kim, I. The Complete Nucleotide Sequence and Gene Organization of the Mitochondrial Genome of the Bumblebee, Bombus Ignitus (Hymenoptera: Apidae). *Gene* **2007**, *392*, 206–220.
47. Sharp, P.M.; Tuohy, T.M.; Mosurski, K.R. Codon Usage in Yeast: Cluster Analysis Dearly Differentiates Highly and Lowly Expressed Genes. *Nucleic Acids Res.* **1986**, *14*, 5125–5143.
48. Mazumdar, P.; Binti Othman, R.Y.; Mebus, K.; Ramakrishnan, N.; Ann Harikrishna, J. Codon Usage and Codon Pair Patterns

in Non-Grass Monocot Genomes. *Ann. Bot.* **2017**, *120*, 893–909.

49. Boore, J.L. Animal Mitochondrial Genomes. *Nucleic Acids Res.* **1999**, *27*, 1767–1780.
50. Fernández-Silva, P.; Enriquez, J.A.; Montoya, J. Replication and Transcription of Mammalian Mitochondrial DNA. *Exp. Physiol.* **2003**, *88*, 41–56.
51. Hao, H.; Liu, S.; Zhang, X.; Chen, W.; Song, Z.; Peng, H.; Liu, Y.; Yue, B. Complete Mitochondrial Genome of a New Vole *Proedromys liangshanensis* (Rodentia: Cricetidae) and Phylogenetic Analysis with Related Species: Are There Implications for the Validity of the Genus *Proedromys*? *Mitochondrial DNA* **2011**, *22*, 28–34.
52. Kim, J.Y.; Park, Y.C. Gene Organization and Characterization of the Complete Mitogenome of *Hypsugo alaschanicus* (Chiroptera: Vespertilionidae). *Genet. Mol. Res.* **2015**, *14*, 16325–16331, doi:10.4238/2015.December.8.24.
53. Porter, V.; Alderson, L.; Hall, S.J.; Sponenberg, D.P. *Mason's World Encyclopedia of Livestock Breeds and Breeding, 2 Volume Pack*; CABI, 2016;
54. Rubinoff, D.; Holland, B.S. Between Two Extremes: Mitochondrial DNA Is Neither the Panacea nor the Nemesis of Phylogenetic and Taxonomic Inference. *Syst. Biol.* **2005**, *54*, 952–961.
55. Cole, L.W.; Guo, W.; Mower, J.P.; Palmer, J.D. High and Variable Rates of Repeat-Mediated Mitochondrial Genome Rearrangement in a Genus of Plants. *Mol. Biol. Evol.* **2018**, *35*, 2773–2785.