Data in Brief 57 (2024) 111142

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Characterization of the complete mitogenome data of collared peccary, *Dicotyles tajacu* (Linnaeus, 1758) (Suina: Tayassuidae) from Ucayali, Peru

Julio César Chávez-Galarza[a,*], Victor Manuel Arévalo-Rojas[b], Gladys Luz Garay Livia[b], Rubén Dario Ferro-Mauricio[a], Daniel Vecco[c,d], Agustin Cerna-Mendoza[c], Aldi Alida Guerra-Teixeira[e], Dora Henriques[f], Fredy Fabian Domínguez[g], Wilder Macedo-Córdova[e], Miguel Alexis Llanto-López[h]

[a] *Departamento Académico de Ciencias Básicas y Afines, Facultad de Ingeniería, Universidad Nacional de Barranca, Av. Toribio de Luzuriaga N °376, Mz J, Urbanización La Florida, Barranca, Perú*
[b] *Departamento Académico de Ingeniería, Facultad de Ingeniería, Universidad Nacional de Barranca, Av. Toribio de Luzuriaga N °376, Mz J, Urbanización La Florida, Barranca, Perú*
[c] *Departamento Académico Agrosilvo Pastoril, Facultad de Ciencias Agrarias, Universidad Nacional de San Martin, Ciudad Universitaria, Jr. Amorarca N° 334, Morales, Perú*
[d] *Centro URKU, Jirón Prolongación Alerta km 2.3 Sector Bocatoma del río Shilcayo, Tarapoto, Perú*
[e] *Departamento Académico de Ciencias Básicas y Pecuarias, Facultad de Zootecnia, Universidad Nacional de la Amazonia Peruana, Calle Francisco Bardales 804, Barrio Moralillo, Yurimaguas, Perú*
[f] *Centro de Investigação em Montanha, Instituto Politécnico de Bragança, Campus Santa Apolónia, 5300-253 Bragança, Portugal*
[g] *Escuela Académico Profesional de Medicina Veterinaria, Facultad de Ciencias Agrarias, Universidad Nacional de San Martin, Jr. Maynas N° 177, Tarapoto, Perú*
[h] *Parque Natural de Pucallpa, Autoridad Regional Ambiental de Ucayali, JC3V+P9F, José Balta, 25002 Pucallpa, Perú*

ARTICLE INFO

Article history:
Received 13 September 2024
Revised 18 October 2024
Accepted 11 November 2024
Available online 20 November 2024

Dataset link: Mitochondrial genome of Dicotyles tajacu (Original data)

ABSTRACT

The collared peccary (*Dicotyles tajacu* Linnaeus, 1758) is a vital resource for the subsistence and economy of the Amazonian inhabitants. Despite its importance, there is a notable lack of genetic information on Peruvian collared peccary populations. This study presents the complete mitogenome of *D.*

* Corresponding author.
E-mail address: jchavez@unab.edu.pe (J.C. Chávez-Galarza).

https://doi.org/10.1016/j.dib.2024.111142

Keywords:
Mitochondrial genome
Assembly
Annotation
Phylogeny

tajacu from Peru, obtained by Illumina sequencing. The mitochondrial genome spans 16,836 bp and has a nucleotide composition of 34.2 % A, 25.5 % T, 13.5 % G, and 26.8 % C, with a GC content of 40.3 %. The genome includes 13 protein-coding genes, 22 tRNA genes, two rRNA genes, and one control region. Phylogenetic analysis of protein-coding genes indicates that Peruvian *D. tajacu* is most closely related to Bolivian *D. tajacu* within the family Tayassuidae. The annotated mitogenome of Peruvian *D. tajacu* provides valuable genomic data for evolutionary research and will serve in conservation and management strategies for the species.

Specifications Table

Subject	Biological Sciences
Specific subject area	Omics: Genomics
Type of data	Table: Gene annotations, base composition Figures: collared peccary, circular mitogenome map, phylogenetic tree Fasta: Mitogenome sequence Fastq: DNA sequence reads Data Format: Raw and analyzed
Data collection	Dna extraction and sequencing: Total genomic DNA was extracted of blood sample using the E.Z.N.A. Tissue DNA Kit. The library was constructed, and then sequenced on the Illumina Novaseq 6000 platform, with paired-end read lengths of 150 bp. Assembly and annotation: Read quality was assessed and filtered using FastQC and Trimmomatic, respectively. Reads were mapped to reference mitogenome available in NCBI. The assembly was performed with SPAdes, and the annotation with Geseq and MITOS2. Transfer RNAs were predicted using tRNAScan-SE. Circular mitogenome map was drawn using CGVIEW. Phylogenetic analysis: IQ-tree was used to construct the Maximum Likelihood phylogenetic tree.
Data source location	Location: Parque Natural de Pucallpa City: Pucallpa, Ucayali Country: Peru Latitude and Longitude: 8°23′40.4′′ S, 74°33′21.6′′ W Sample repository: Genomic DNA was deposited in the germplasm collection of Universidad Nacional de Barranca (contact person: Renzo Valdez-Núñez, rvaldez@unab.edu.pe) under the voucher number N° UNAB-DNA-003.
Data accessibility	Repository name: NCBI BioProject Data identification number: PRJNA1135145 Direct URL to data: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1135145 Repository name: NCBI BioSample Data identification number: SAMN42463660 Direct URL to data: https://www.ncbi.nlm.nih.gov/biosample/?term=SAMN42463660 Repository name: NCBI SRA Data identification number: SRR29795200 Direct URL to data: https://www.ncbi.nlm.nih.gov/sra/?term=SRR29795200 Repository name: NCBI Genbank Data identification number: PQ009497 Direct URL to data: https://www.ncbi.nlm.nih.gov/nuccore/PQ009497 Repository name: Zenodo data Data identification number: 10.5281/zenodo.13382317 Direct URL to data: https://zenodo.org/records/13382318

1. Value of the Data

- The data represents the complete Peruvian mitogenome of *Dicotyles tajacu*, which serves as a food, economic and cultural support for Amazonian inhabitants.
- The mitochondrial genome may be used by researchers to conduct comparative analysis seeking to clarify the taxonomic status among specimens distributed in the Americas [1,2].
- The provided data is a valuable resource for phylogenetic analysis, population genomics, and evolutionary research on Tayassuidae with usefulness in management and conservation strategies.

2. Background

D. tajacu is a critical resource for Amazonian settlers (Fig. 1), who utilize it as a source of bush meat, sold in local market and for skins export in international trade [3,4]. Ecologically, *D. tajacu* plays a significant role as prey for large carnivores, and as seed dispersers, contributing to ecosystems regeneration and maintenance. However, this species faces multiple threats, including forest fragmentation, over-hunting, introduction of exotic species, and disease transmission, all of which increase its population decline [5]. Despite its importance, studies on Peruvian collared peccary have been limited, these focusing mainly on demography, breeding and management [6,7], with no research to date on genetic diversity or genomic characterisation. Currently, three mitogenomes have been deposited into GenBank, one of unknown origin (NC_012103.1), and the others from Bolivia (JN632682.1) and French Guiana (JN632683.1). Therefore, this work seeks to sequence and characterize the mitogenome of *D. tajacu* from Peru as an initial step towards the development of genomic tools for evolutionary research, conservation efforts, and management strategies.

3. Data Description

The complete mitochondrial genome of *D. tajacu* from Ucayali has a total length of 16,836 bp (GenBank accession PQ009497.1), with a nucleotide base composition of A= 34.2 %, C= 26.8 %, T= 25.5 %, G= 13.5 %, resulting in AT and GC contents of 59.7 % and 40.3 %, respectively.

Fig. 1. Image of female *D. tajacu* adult.

Table 1
Base composition and AT/GC skewness of mitogenome of *D. tajacu*.

Sequence	size (bp)	A%	G%	T%	C%	A + T %	G + C %	AT skew	GC skew
Mitogenome	16,836	34.2	13.5	25.5	26.8	59.7	40.3	0.146	−0.330
Protein-coding protein	11,350	33.6	12.0	25.9	28.5	59.5	40.5	0.129	-0.407
tRNAs	1432	36.2	15.1	26.7	22.0	62.9	37.1	0.151	-0.186
rRNAs	2527	38.0	17.0	22.7	22.3	60.7	39.3	0.252	−0.135
Control region	1403	30.4	16.6	27.2	25.8	57.6	42.4	0.056	−0.217

Table 2
Gene organization of *D. tajacu* mitogenome. *Positive and negative values correspond to intergenic or overlapping regions between genes, respectively.

Gene	Type	Start	Stop	Strand	Length (bp)	Start	Stop	Codon	Anticodon	Intergenic* Nucleotide
trnF	tRNA	1	70	H	70			TTC	GAA	0
s-rRNA	rRNA	70	1021	H	952					−1
trnV	tRNA	1021	1088	H	68			GTA	TAC	−1
l-rRNA	rRNA	1089	2663	H	1575					0
trnL2	tRNA	2663	2737	H	75			TTA	TAA	−1
ND1	Coding	2739	3693	H	955	ATG	T–			1
trnI	tRNA	3694	3763	H	70			ATC	GAT	0
trnQ	tRNA	3761	3833	L	73			CAA	TTG	−3
trnM	tRNA	3834	3903	H	70			ATG	CAT	0
ND2	Coding	3904	4945	H	1042	ATA	T–			0
trnW	tRNA	4946	5013	H	68			TGA	TCA	0
trnA	tRNA	5019	5086	L	68			GCA	TGC	6
trnN	tRNA	5093	5166	L	74			AAC	GTT	6
Ori-L		5167	5198	H	32					0
trnC	tRNA	5199	5265	L	67			TGC	GCA	32
trnY	tRNA	5266	5331	L	66			TAC	GTA	0
COX1	Coding	5333	6877	H	1545	ATG	TAA			1
trnS2	tRNA	6881	6949	L	69			TCA	TGA	3
trnD	tRNA	6956	7024	H	69			GAC	GTC	6
COX2	Coding	7025	7720	H	696	ATG	AGA			0
trnK	tRNA	7713	7779	H	67			AAA	TTT	−8
ATP8	Coding	7781	7984	H	204	ATG	TAA			1
ATP6	Coding	7942	8622	H	681	ATG	TAA			−43
COX3	Coding	8622	9405	H	784	ATG	T–			−1
trnG	tRNA	9406	9473	H	68			GGA	TCC	0
ND3	Coding	9475	9821	H	347	ATA	TA-			1
trnR	tRNA	9822	9889	H	68			CGA	TCG	0
ND4L	Coding	9890	10,186	H	297	ATG	TAA			0
ND4	Coding	10,180	11,557	H	1375	ATG	T–			−7
trnH	tRNA	11,558	11,625	H	68			CAC	GTG	0
trnS1	tRNA	11,626	11,684	H	59			AGC	GCT	0
trnL1	tRNA	11,685	11,754	H	70			CTA	TAG	0
ND5	Coding	11,755	13,575	H	1821	ATA	TAA			0
ND6	Coding	13,559	14,086	L	528	ATG	TAA			−17
trnE	tRNA	14,087	14,155	L	69			GAA	TTC	0
CytB	Coding	14,160	15,299	H	1140	ATG	AGA			4
trnT	tRNA	15,300	15,369	H	70			ACA	TGT	0
trnP	tRNA	15,369	15,433	L	65			CCA	TGG	−1
Ori-H		16,001	16,203	L	203					0
D-loop		15,434	16,836	H	1393					0

The sequence displays a low positive AT-skew (0.146) and negative GC-skew (−0.330) (Table 1). The mitogenome assembly had an average sequencing depth of 7148.36 X (Supplementary Figure 1). Mitogenome annotation revealed the presence of 37 genes, including 13 protein-coding genes, 22 tRNA genes, two rRNA genes and one control region (Fig. 2). Nine of these genes (one protein-coding gene, NADH dehydrogenase subunit *ND6*, and eight tRNA genes – *trnQ* (TTG), *trnA* (TGC), *trnN* (GTT), *trnC* (GCA), *trnY* (GTA), *trnS2* (TGA), *trnE* (TTC), and *trnP* (TGG) were located on the complementary strand (Table 2). The mitogenome presented 10 gene spacer regions, rang-

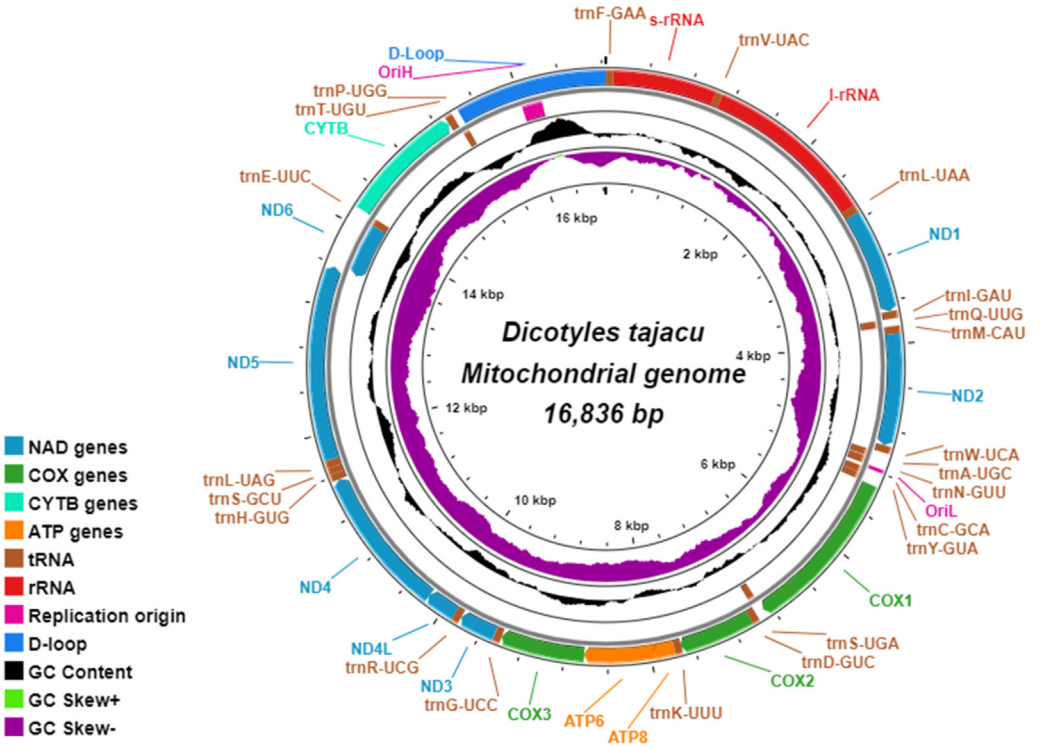

Fig. 2. The complete mitogenome map of *D. tajacu*. In the outermost circle is showed the position of genes: outer genes present transcription on heavy strand (H-strand) in a clockwise direction, and inner genes are transcribed on light strand (L-strand) in a counterclockwise sense. The innermost and middle circles display GC skew and GC content, respectively. The legend is presented in bottom left corner where genes are color-labelled according to their different functional categories.

ing from 1 to 32 bp, with the longest region between *trnN* and *trnC*. Likewise, 10 gene overlap regions were identified, ranging from 1 to 43 bp, with the longest overlap region between the *ATP8* and *ATP6* genes. All 13 protein-coding genes were represented by the initiator codon ATN, specifically 10 with ATG (*ND1, ND4L, ND4, ND6, COX1, COX2, COX3, CytB, ATP6, ATP8*), and three with ATA (*ND2, ND3, ND5*). Termination codons varied with *COX1, ATP6, ATP8, ND4L, ND5*, and *ND6* ending with TAA; *COX2* and *CytB* ending with AGA; *ND1, ND2, ND3, ND4, COX3* ending with a single T which are completed by the addition of 3′ A residues to the mRNA by post-transcriptional polyadenylation [8]. The 22 tRNA genes ranged in length from 59 bp to 75 bp, and all exhibiting the typical cloverleaf secondary structure, except for *trnS1* which lacks a dihydrouridine arm. The *s-rRNA* (12S) and *l-rRNA* (16S) genes have a length of 952 bp and 1575 bp, respectively, and separated by *trnV* gene. The phylogenetic analysis based on protein-coding genes yielded a strong bootstrap support across all nodes, ranging from 94 % a 100 % (Fig. 3). The families Suidae, Tayassuidae, and Camelidae were confirmed as monophyletic, as well as Suidae and Tayassuidae are sister taxa and clustered together to form the suborder Suina. Within the family Tayassuidae, Peruvian *D. tajacu* was closely related to Bolivian *D. tajacu* with respect to the others analyzed.

4. Experimental Design, Materials and Methods

4.1. Sample collection

The blood sample was collected from an adult female *D. tajacu* from the Parque Natural de Pucallpa, Ucayali, Peru (8°23′40.4'' S, 74°33′21.6'' W). The blood sample was stored at 4°C and transported to the Laboratory of Molecular Analysis and Genomics of Universidad Nacional de Barranca.

4.2. DNA extraction and sequencing

Total genomic DNA was extracted using the E.Z.N.A. Tissue DNA Kit (Omega Bio-tek Inc., Georgia, USA) following the manufacturer's instruction. The quantity and integrity of genomic DNA was assessed by Qubit 4.0 Fluorometer and 1.5 % agarose gel electrophoresis, respectively. Genomic DNA was deposited and stored at −80 °C in the germplasm collection of Universidad Nacional de Barranca (contact person: Renzo Valdez-Núñez, rvaldez@unab.edu.pe) under the voucher number N° UNAB-DNA-003, and only 30 µL was used for sequencing. The library was constructed, and then sequenced on the Illumina Novaseq 6000 platform, with paired-end read lengths of 150 bp.

4.3. Mitogenome assembly and annotation

Read quality was assessed using FastQC [9]. Adapters and low-quality reads were trimmed using Trimmomatic 0.39 [10]. Reads were mapped to the *D. tajacu* reference mitogenome (NC_012103.1) available in NCBI. The assembling was performed with SPAdes 3.15.5 [11]. The base composition and GC/AT skew were estimated by using Bioedit [12]. Mitogenome annotation was performed with Geseq [13] and MITOS2 [14] based on the vertebrate mitochondrial code. Transfer RNAs (tRNA) were predicted using the on-line software tRNAScan-SE 2.0 (https://lowelab.ucsc.edu/tRNAscan-SE/, accessed on 17 July 2024) [15]. The circular architecture of the mitogenome was designed and visualized using CGVIEW [16].

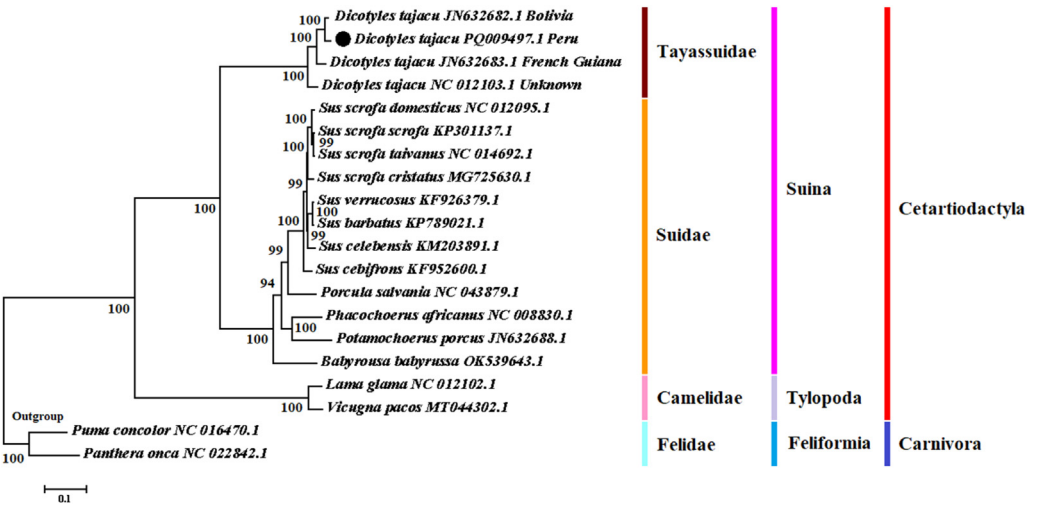

Fig. 3. Phylogenetic relationships of *D. tajacu* and 19 other mammalian mitogenome sequences. Bootstrap support values of the Maximum likelihood tree are indicated on each branch nodes. The black circle marks out the examined species.

4.4. Phylogenetic analysis

To confirm the phylogenetic position of *D. tajacu*, its mitochondrial genome was compared and aligned using MAFFT 7.487 [17]. The analysis included 12 mitogenomes from Suidae, three from Tayassuidae (related to *D. tajacu*), and two from Camelidae. *Puma concolor* (NC_016470.1) and *Panthera onca* (NC_022842.1) from the family Felidae were used as outgroups. All mitogenomes were retrieved from GeneBank. The Maximum Likelihood phylogenetic tree was inferred based on concatenated protein-coding genes using IQ-TREE 2.3.2 [18], with the optimal nucleotide substitution model GTR+I+G selected by JModeltest 2.1.10 [19], 1000 ultrafast bootstrap and 1000 SH-aLRT replicates.

Limitations

Not applicable.

Ethics Statement

The sample collection from collared peccary specimen was conducted in accordance with the Peruvian National Law No. 30407: "Animal Protection and Welfare" and Supreme Decree N° 019-2021-MINAM of Ministry of Environment of Peruvian Government for access to genetic resources

CRediT Author Statement

Julio César Chávez-Galarza: Conceptualization, Methodology, Formal analysis, Investigation, Resources, Data curation, Writing-Original draft, Writing – Review and Edition, Visualization, Supervision, Administration project, Funding acquisition; **Victor Manuel Arévalo-Rojas:** Conceptualization, Investigation, Writing – Original draft, Administration project; **Gladys Luz Garay Livia:** Conceptualization, Investigation, Writing – Review and Edition; **Rubén Dario Ferro-Mauricio:** Methodology, Formal analysis, Investigation, Data curation, Visualization; **Daniel Vecco:** Conceptualization, Investigation, Resources, Writing – Review and Edition, Supervision; **Agustín Cerna-Mendoza:** Conceptualization, Resources, Writing – Review and Edition, Supervision; **Aldi Alida Guerra-Teixeira:** Resources, Writing – Review and Edition, Supervision; **Dora Henriques:** Conceptualization, Formal analysis, Writing – Original draft; **Fredy Fabian Domínguez:** Methodology, Investigation, Writing, Review and Edition; **Wilder Macedo-Córdova:** Investigation, Writing, Review and Edition; **Miguel Alexis Llanto López:** Methodology, Investigation, Writing – Review and Edition.

Data Availability

Mitochondrial genome of Dicotyles tajacu (Original data) (Zenodo).

Acknowledgments

We would like to thank the staff of Parque Natural de Ucayali for their support in the blood collection of collared peccary specimens.

This work was supported by the research project "Revalorizando nuestros recursos zoogenéticos: Identificación de genes evolutivos y de importancia económica en *Tayassu* spp (Suborden: Suina) mediante genómica comparativa" with grant number R.C.O. N° 629-2022-UNAB.

Declaration of Competing Interest

No potential conflict of interest was reported by the authors.

Supplementary Materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2024.111142.

References

[1] J. Gongora, S. Morales, J.E. Bernal, C Moran, Phylogenetic divisions among Collared peccaries (*Pecari tajacu*) detected using mitochondrial and nuclear sequences, Mol. Phylogenet. Evol. 41 (2006) 1–11, doi:10.1016/j.ympev.2006.05.015.
[2] C. Groves, P. Grubb, Ungulate Taxonomy, The Johns Hopkins University Press, Baltimore, USA, 2011.
[3] T. Fang, R.E. Bodmer, P. Puertas, P. Mayor, P.E. Pérez-Peña, R. Acero, D. Haymann, Certificación De Pieles De Pecaríes En La Amazonía peruana: Una estrategia Para La Conservación y Manejo De Fauna en La Amazonía Peruana, Wust Ediciones, Lima, Perú, 2008.
[4] M.P. Quiceno-Mesa, D. Cruz-Antia, N. Van Vliet, J. Neves de Aquino, T. Schor, La invisibilidad de las cadenas comerciales de carne de monte en la triple frontera amazónica entre Colombia, Perú y Brasil, Rev. Colomb. Amazón. 7 (2014) 51–71.
[5] S. Mandujano, R. Reyna-Hurtado, Recent studies on white-lipped peccary and collared peccary in the neotropics, in: S. Gallina-Tessaro (Ed.), Ecology and Conservation of Tropical Ungulates in Latin America, Springer International Publishing, Switzerland, 2019, pp. 415–438.
[6] P. Pérez-Peña, M.S. Riveros, P. Mayor, M.C. Ramos-Rodríguez, et al., Estado poblacional del sajino *(Pecari tajacu)* y huangana *(Tayassu pecari)* en la Amazonía peruana, Folia Amazón. 26 (2017) 103–120, doi:10.24841/fa.v26i2.429.
[7] M.E. Rengifo-Pinedo, A. Flores Mere, D.N. Torres, D.N. Rengifo, El Desarrollo de Una Crianza Sostenible en Pecaríes, Editorial Tercer Sol. Mexico, 2022.
[8] D. Ojala, J. Montoya, G. Attardi, tRNA punctuation model of RNA processing in human mitochondria, Nature 290 (1981) 470–474, doi:10.1038/290470a0.
[9] S. Andrews, FastQC: a quality control tool for high throughput sequence data [Online], 2010, http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
[10] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (2014) 2114–2120, doi:10.1093/bioinformatics/btu170.
[11] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, et al., SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing, J. Comput. Biol. 19 (2012) 455–477, doi:10.1089/cmb.2012.0021.
[12] T.A Hall, BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT, Nucleic Acids Sympos. Ser. 41 (1999) 95–98.
[13] M. Tillich, P. Lehwark, T. Pellizzer, E.S. Ulbricht-Jones, A. Fischer, R. Bock, S. Greiner, GeSeq – versatile and accurate annotation of organelle genomes, Nucleic Acids Res. 45 (2017) W6–W11, doi:10.1093/nar/gkx391.
[14] M. Bernt, A. Donath, F. Jühling, F. Externbrink, C. Florentz, G. Fritzsch, J. Pütz, M. Middendorf, P.F. Stadler, MITOS: improved de novo metazoan mitochondrial genome annotation, Mol. Phylogenet. Evol. 69 (2013) 313–319, doi:10.1016/j.ympev.2012.08.023.
[15] T.M. Lowe, P.P. Chan, tRNAscan-SE on-line: search and contextual analysis of transfer RNA genes, Nucl. Acids Res. 44 (2016) W54-57, doi:10.1093/nar/gkw413.
[16] J.R. Grant, P. Stothard, The CGView Server: a comparative genomics tool for circular genomes, Nucleic Acids Res. 36 (2008) W181–W184, doi:10.1093/nar/gkn179.
[17] K. Katoh, D.M. Standley, MAFFT multiple sequence alignment soft-ware version 7: improvements in performance and usability, Mol. Biol. Evol. 30 (2013) 772–780, doi:10.1093/molbev/mst010.
[18] L-T. Nguyen, H.A. Schmidt, A. von Haeseler, B.Q. Minh, IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies, Mol. Biol. Evol. 32 (2015) 268–274, doi:10.1093/molbev/msu300.
[19] D. Darriba, G. Taboada, R. Doallo, D. Posada, jModelTest 2: more models, new heuristics and parallel computing, Nat. Methods 9 (2012) 772, doi:10.1038/nmeth.2109.