Data in Brief 53 (2024) 110246

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

The first complete mitochondrial genome data of the Afghan pika *Ochotona rufescens* (Lagomorpha, Ochotonidae), near the type locality

Anwesha Saha[a,b], Mateusz Baca[b], Danijela Popović[b], Zeinolabedin Mohammadi[c], Urban Olsson[d,e], Emily Roycroft[f], Łucja Fostowicz-Frelik[a,*]

[a] *Institute of Paleobiology, Polish Academy of Sciences, Twarda 51/55, 00–818 Warsaw, Poland*
[b] *Centre of New Technologies (CeNT), University of Warsaw, S. Banacha 2c, 02–097 Warsaw, Poland*
[c] *Department of Biology, Faculty of Sciences, Golestan University, Gorgan, Iran*
[d] *Systematics and Biodiversity, Department of Biology and Environmental Sciences, University of Gothenburg, Box 463, SE-405 30 Gothenburg, Sweden*
[e] *Gothenburg Global Biodiversity Centre, Box 461, SE-405 30 Gothenburg, Sweden*
[f] *Division of Ecology and Evolution, Research School of Biology, The Australian National University, Canberra, Acton, ACT 2601, Australia*

ARTICLE INFO

Article history:
Received 9 October 2023
Revised 3 February 2024
Accepted 19 February 2024
Available online 28 February 2024

Dataset link: ON859137 (Original data)
Dataset link: ON859136 (Original data)

Keywords:
Mitogenome
Museomics
De novo assembly
Lagomorpha
Ochotona
Asia

ABSTRACT

The Afghan pika *Ochotona rufescens* (Gray, 1842) is widely distributed across the mountains of Afghanistan, Iran, Pakistan, and southwestern Turkmenistan, most often at elevations between 2,000 and 3,000 m. Here we present, for the first time, the complete mitochondrial genomes of two specimens of the nominotypical subspecies *Ochotona rufescens rufescens*, de novo assembled from Illumina short reads of fragmented probe-enriched DNA. The lengths of the circular mitogenomes are 16,408 bp and 16,407 bp, respectively. Both mitogenomes contain 13 protein-coding genes (PCGs), two ribosomal RNAs (16S rRNA and 12S rRNA), 22 transfer RNA genes, and a control region. The gene NAD6 and the tRNA (Gln), tRNA (Ala), tRNA (Asn), tRNA (Cys), tRNA (Tyr), tRNA (Ser), tRNA (Glu), and tRNA (Pro) are encoded on the

* Corresponding author.
E-mail address: lfost@twarda.pan.pl (Ł. Fostowicz-Frelik).
Social media: @mimee0991 (A. Saha), @IPaleoPAN (Ł. Fostowicz-Frelik)

https://doi.org/10.1016/j.dib.2024.110246

light strand while the rest are encoded on the heavy strand. The overall nucleotide composition was ~30% for A, 25% for T, 15% for G, and 29% for C. The mitogenome data are available in the GenBank under the accession numbers ON859136 and ON859137.

Specifications Table

Subject	Biological Sciences/Omics/Genomics
Specific subject area	Mitogenomics
Data format	Raw and analysed
Type of data	Tables (.docx), mitogenome map in figure format (.jpg), phylogenetic tree in figure format (.jpg) S1: List of taxa and corresponding accession numbers used to design baits for target enrichment (.xls) S2: Quality of mitogenomic DNA samples in tables (.docx) S3: Sequencing QC data (.zip) S4: Alignment file for phylogenetic analysis in PHYLIP format (.phy) S5: Gene partition file for phylogenetic analysis in NEXUS format (.nex) S6: Phylogenetic tree in NEXUS format (.nex) S7: Additional mitogenome table and figure (.docx) S8: Additional phylogenetic tree based on cytochrome b sequences (.pdf) S9: NCBI accession numbers and taxonomic references (.xls)
Data collection	Genomic DNA was extracted following the modified phenol-chloroform extraction protocol [1]; double-indexed and double-stranded library preparation; mitogenome enrichment with myBaits Custom 20–40 K (Daicel Arbor Biosciences, USA); sequencing: Illumina NextSeq 550 platform; mitogenome assembled de novo in NOVOPlasty v.4.3.1 and annotated in MITOS2 webserver. The circular mitogenome maps were drawn using OGDRAW. Phylogenetic relationships were inferred using IQ-TREE
Data source location	Two individuals of *Ochotona rufescens rufescens*, collected near the original type locality (Lat.: 34° 34′ 59.9874″N; Long.: 68° 57′ 0″E) in Paghman, Kabul Province, Afghanistan, were preserved under the voucher numbers FMNH 102840 (male) and FMNH 102841 (female) in the Field Museum of Natural History, Chicago, IL, USA
Data accessibility	The mitogenome data are available in the GenBank under the accession number ON859136 (https://www.ncbi.nlm.nih.gov/nuccore/ON859136) and ON859137 (https://www.ncbi.nlm.nih.gov/nuccore/ON859137). Raw sequence data are available in BioProject PRJNA1020817, BioSample: SAMN37529232 (specimen FMNH102840), https://www.ncbi.nlm.nih.gov/biosample/37529232; SAMN37529233 (sample FMNH102841), https://www.ncbi.nlm.nih.gov/biosample/37529233, and Mendeley data repository (https://data.mendeley.com/datasets/ggmjwhymmf/3) [2]

1. Value of the Data

- This data will be useful for designing appropriate conservation policies for the species.
- This is new genomic information for research on hybridization in *Ochotona*.
- The mitogenome will be useful in population genomics studies of *Ochotona rufescens* to monitor habitat-driven cryptic diversity among the species.
- The mitogenome from the topotypical population will facilitate taxonomic evaluation of the *Ochotona rufescens* complex.

Table 1
Raw mitogenome sequence data of two specimens of *Ochotona rufescens*.

	FMNH 102840	FMNH 102841
Total reads obtained	2,924,244	1,116,748
Total mapped reads	5347	35,744
Mapped paired reads	1148	12,169
Mean Coverage (x)	28.5737	217.6494
Mean Mapping Quality	59.89	59.86

2. Background

Until now, *Ochotona* genomics research has primarily revolved around North American, Russian, and Chinese species, lacking any substantial information from the Indian subcontinent and central Asia. Previous multi-locus phylogenies indicated the climate and habitat-driven diversification among the Afghan pika *Ochotona rufescens* populations and separated the species into O. *rufescens* and O. *vizier* [3]; the latter species recognized for the Zagros region population [3].

3. Data Description

The Afghan pika *Ochotona rufescens* (Gray, 1842) is widely distributed across the mountains of Afghanistan, Iran, Pakistan, and southwestern Turkmenistan at elevations up to 3000 m. [4]. This diurnal species is often seen in gorges, valleys, and slopes with talus of arid montane habitats [4]. It consumes various green plants and agricultural crops and spends cold winter and summer drought in the chambered burrows and natural rock crevices [3]. *O. rufescens* is considered Least Concern (LC) on the IUCN Red List of Threatened Species [5].

Here we report the first complete mitogenomes of two specimens of *Ochotona rufescens rufescens*, collected near the type locality, which are 16,408 bp and 16,407 bp in length (GenBank No. ON859136 and ON859137), respectively. The raw mitogenome sequencing data are summarized in Table 1. Each mitogenome contains two ribosomal RNAs (16S rRNA and 12S rRNA), 13 protein-coding genes (PCGs), 22 transfer RNA (tRNA) genes, and a noncoding control region (D-loop). The arrangement of these 37 genes encoded on either the heavy (H) or the light (L) strand is presented in Fig. 1 for the specimen FMNH 102840 and in Figure ii (S7.docx) for the specimen FMNH 102841. The total length of PCGs is 11,395 bp, transcribing 3798 amino acids, which accounts for ∼69% of the entire mitogenome. The gene NAD6 and the tRNA (Gln), tRNA (Ala), tRNA (Asn), tRNA (Cys), tRNA (Tyr), tRNA (Ser), tRNA (Glu), and tRNA (Pro) are encoded on the light strand while rest are encoded on the heavy strand (Fig. 1; Table 2) (see also Table ii and Figure ii in S7.docx). The overall nucleotide composition is estimated as ∼A: 30%, T: 25%, G: 15%, and C: 29%. Precise nucleotide composition for each individual is given in Table 3.

The phylogenetic position of *Ochotona rufescens* in the *Conothoa* subgenus is robustly supported (Fig. 2; bootstrap value=100; node with "*"), in agreement with a comprehensive phylogeny revising the *Conothoa* taxonomy [6]. The description of this novel mitogenome is an important addition to ochotonid taxonomic and evolutionary research.

4. Experimental Design, Materials and Methods

4.1. Biological sample

Two individuals of *O. r. rufescens* were sampled from the mammalian collection of the Field Museum of Natural History, Chicago, IL, USA (voucher FMNH 102840 - male and FMNH 102841 – female; https://collections-zoology.fieldmuseum.org/). These specimens were originally collected

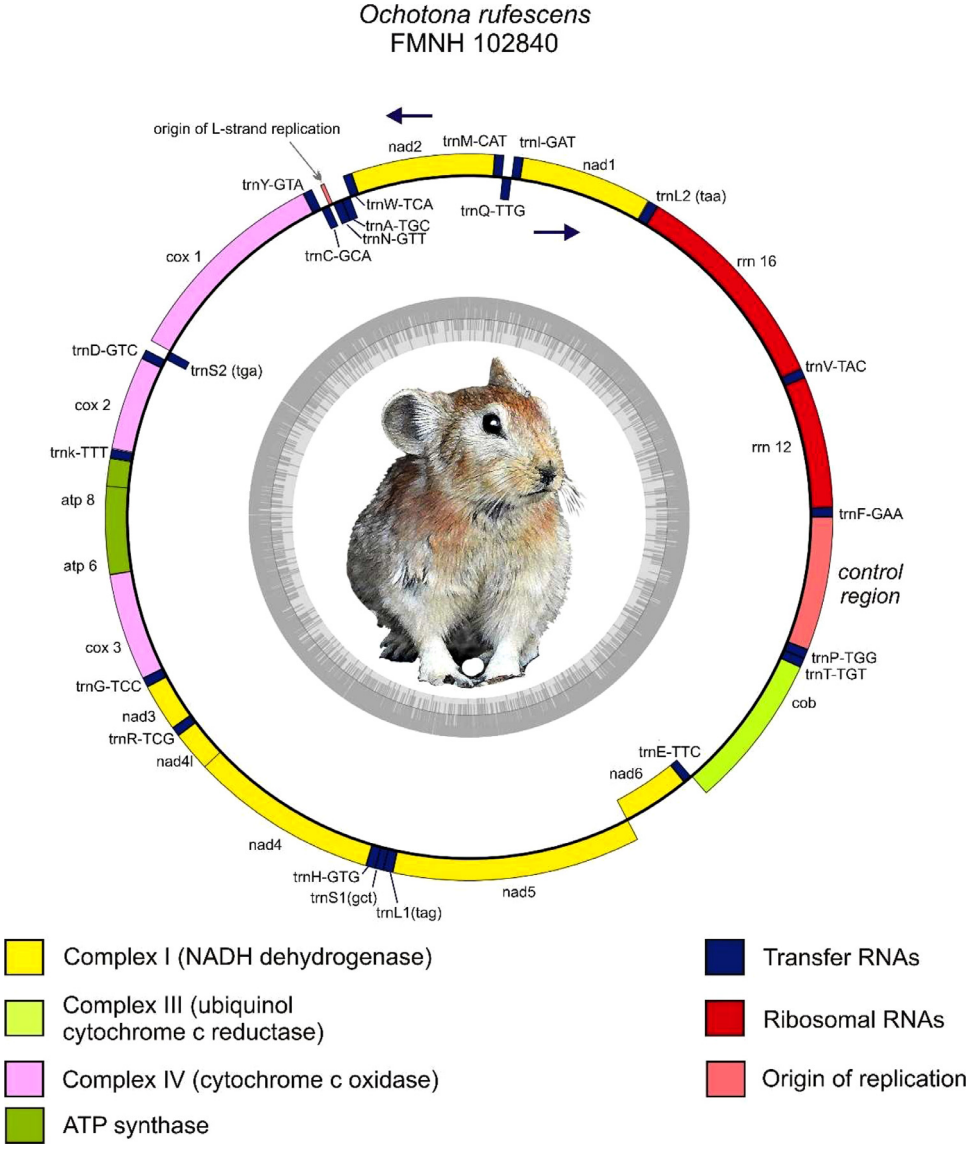

Fig. 1. Circular map of *Ochotona r. rufescens* (FMNH 102840) mitochondrial genome. Genes encoded on the heavy (H) and light (L) strands represent the outer and inner circle, respectively. The direction of the arrows symbolizes the transcription direction (both clockwise and anti-clockwise) of the genes. The inner ring (gray) indicates the GC content of the genome. Circular mitogenome map of individual FMNH 102841 is given in supplementary file S7. The image of *O. rufescens* by Mohammad Amin Ghaffari (https://www.inaturalist.org/photos/213736083; CC BY 4.0), modified.

during a Street Expedition to Afghanistan (SETA) on July 13, 1965, near the type locality (latitude: 34° 34′ 59.987″N; longitude: 68° 57′ 0″E) in Paghman, Kabul Province, Afghanistan. Dried skin tissue was obtained by ZM and UO from the museum and transferred to AS for genomic analyses.

Total genomic DNA was extracted from the tissue with a modified phenol-chloroform extraction method [1], suitable for skin and toepads sampled from poorly preserved old museum

Table 2
Mitogenome feature of *Ochotona r. rufescens* (GenBank accession number ON859136). Protein coding genes (PCGs) are represented in bold letters and the genes encoded on the light (L) strand are italicized. The mitogenome feature table of the specimen FMNH 102841 is in the supplementary file (S7.docx).

Gene	*Position*		*Size (bp)*	*Amino acid length*	*Strand*
	Start	*end*			
tRNA (Phe)	1	69	69		H
12S rRNA	70	1030	961		H
tRNA(Val)	1031	1098	68		H
16S rRNA	1100	2678	1579		H
tRNA (Leu)	2679	2753	75		H
NAD1	2756	3711	**956**	**319**	**H**
tRNA (Ile)	3712	3778	67		H
tRNA (Gln)	*3777*	*3848*	*72*		*L*
tRNA (Met)	3849	3917	69		H
NAD2	3918	4961	**1044**	**348**	**H**
tRNA (Trp)	4962	5027	66		H
tRNA (Ala)	*5030*	*5097*	*68*		*L*
tRNA (Asn)	*5098*	*5170*	*73*		*L*
rep_origin	5171	5202	32		
tRNA (Cys)	*5203*	*5269*	*67*		*L*
tRNA (Tyr)	*5270*	*5337*	*68*		*L*
COX1	5339	6880	**1542**	**514**	**H**
tRNA (Ser)	*6883*	*6951*	*69*		*L*
tRNA (Asp)	6955	7023	69		H
COX2	7024	7707	**684**	**228**	**H**
tRNA (Lys)	7711	7777	67		H
ATP8	7779	7983	**205**	**68**	**H**
ATP6	7940	8620	**681**	**227**	**H**
COX3	8620	9403	**784**	**261**	**H**
tRNA (Gly)	9404	9472	69		H
NAD3	9473	9819	**347**	**116**	**H**
tRNA (Arg)	9820	9887	68		H
NAD4-L	9888	10,184	**297**	**99**	**H**
NAD4	10,178	11,555	**1378**	**459**	**H**
tRNA (His)	11,556	11,624	69		H
tRNA (Ser)	11,625	11,683	59		H
tRNA (Leu)	11,684	11,753	70		H
NAD5	11,754	13,565	**1812**	**604**	**H**
NAD6	*13,561*	*14,085*	***525***	***175***	***L***
tRNA (Glu)	*14,086*	*14,154*	*69*		*L*
COB	14,158	15,297	**1140**	**380**	**H**
tRNA (Thr)	15,297	15,364	68		H
tRNA (Pro)	*15,365*	*15,433*	*69*		*L*
Control region	15,434	16,408	975		

Table 3
Nucleotide composition (in %) of two mitogenomes of *Ochotona r. rufescens*.

Specimen	A	T	C	G
FMNH 102840	30.15	24.52	29.52	15.81
FMNH 102841	29.74	25.16	29.48	15.62

specimens. The extracted DNA was subjected to spectrophotometric quantification. Since the DNA was degraded, sonication was not performed prior to library preparation. An input of ~400 ng of extracted DNA was used as the template for constructing a double-indexed, double-stranded library based on the protocol of Meyer and Kircher [7]. The library of one individual (FMNH 102840) was subjected to shotgun sequencing to check the endogenous DNA content before proceeding to target enrichment. The pre-enrichment library (~150 ng) was further enriched for the mitogenome following the manufacturer's protocol, using a total of 28,756 unique

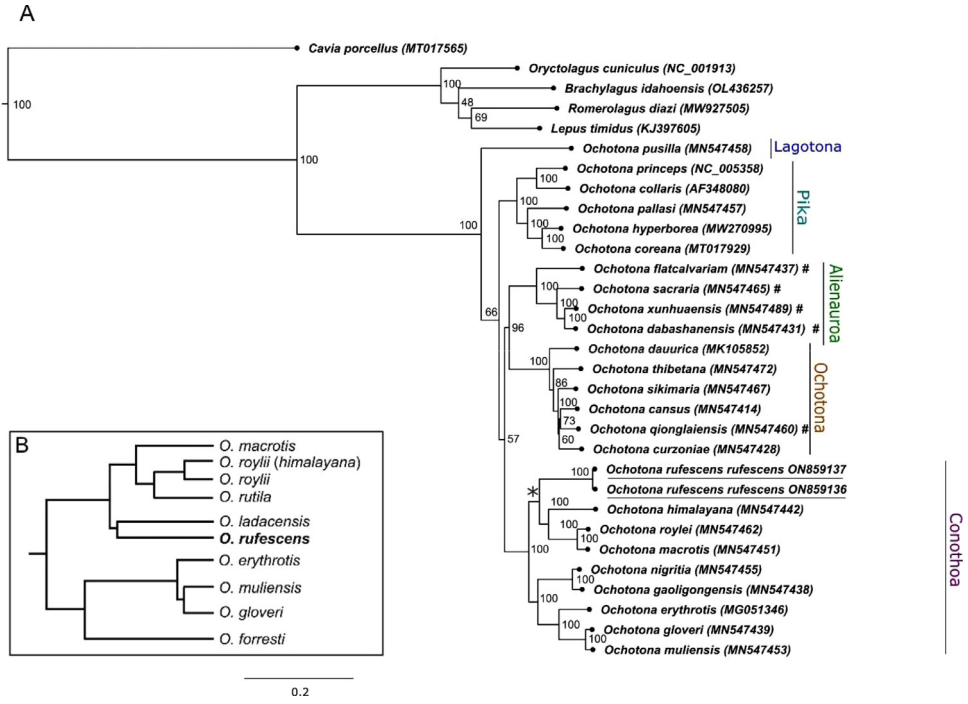

Fig. 2. (A) Phylogeny of *Ochotona rufescens* is constructed with 28 lagomorph species (four leporid and 24 ochotonid) mitochondrial genomes using the maximum-likelihood (ML) method. Vertical lines on the right-hand side of the tree indicate five ochotonid subgroups [22]. Numbers adjacent to each node indicate the corresponding ML bootstrap support. The GenBank accession numbers for the underlined taxa are generated in this study. The node with an asterisk (*) signifies our studied taxon's phylogenetic position within Ochotonidae. The '#' indicates the species names following those in Wang et al. [22], not the direct GenBank entries (see Supplementary Material). (B) Revised taxonomy of the *Conothoa* subgenus according to Lissovsky et al. [6]. Note that our topology of the *rufescens/macrotis* clade differs from the phylogeny in [6]; see S8 file for further information.

custom-designed RNA probes with 3 × tilling, which were based on mitogenomes of 71 species, representing all five extant Euarchontoglires orders, obtained from GenBank (Supplementary file S1.xls; list of species, taxonomic classification, and corresponding accession numbers). These probes were included in myBaits Custom 20–40 K kit (Cat# 300248.V5 201119-92; Daicel Arbor Biosciences, USA). Enriched libraries were pooled and sequenced on Illumina NextSeq 550 using 2 × 75 bp mode.

4.2. Complete mitogenome generation

Raw reads were demultiplexed using bcl2fastq (Illumina), adapters and low-quality bases were trimmed, and overlapping reads were collapsed using AdapterRemoval v.2 [8]. The mitogenome was de novo assembled in NOVOPlasty v.4.3.1 [9] with default parameters to reproduce the candidate mitogenomes. The sequence assembled by NOVOPlasty was used as a reference for mapping using BWA-MEM [10]. Duplicates and reads with low mapping quality (mapQ < 30) were removed using samtools [11]. Number of unique mapped reads, mean coverage and mapping quality of the assembled genome was estimated by Qualimap 2 [12], followed by a manual checking in Tablet software [13]. The mitogenomes were annotated on the MITOS2 webserver ([14]; http://mitos2.bioinf.uni-leipzig.de/index.py) using RefSeq 89 Metazoa reference method. Further, the annotations for individual PCG's start and stop codons were manually con-

firmed by nucleotide BLAST analysis. Circular mitogenome maps were drawn using OGDRAW [15].

To reconstruct the mitogenomic phylogeny of available ochotonid species we used an alignment of 24 *Ochotona* species and five outgroup taxa (*Cavia porcellus* MT017565, a rodent, and the leporid lagomorphs *Romerolagus diazi* MW927505, *Lepus timidus* KJ 397605, *Brachylagus idahoensis* OL436257, and *Oryctolagus cuniculus* NC_001913), aligned with MAFFT [16]. Ambiguously aligned sites were subsequently removed with GBLOCKS v 0.91b [17]. A Maximum-Likelihood (ML) based phylogeny (Fig. 2) was inferred from the resulting alignment (16,333 bp) on IQ-TREE web server ([18]; http://iqtree.cibiv.univie.ac.at/), simultaneously using the model selection (auto) algorithm [19], partition model with varying evolutionary rate [20] and ultrafast bootstrap [21] method for 1000 iterations.

Limitations

Not applicable.

Ethics Statement

The authors have read and follow the ethical requirements for publication in Data in Brief. The current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

Data Availability

ON859137 (Original data) (GenBank)
ON859136 (Original data) (GenBank)

CRediT Author Statement

Anwesha Saha: Methodology, Software, Investigation, Formal analysis, Writing – original draft; **Mateusz Baca:** Methodology, Software, Writing – review & editing, Project administration; **Danijela Popović:** Methodology, Data curation; **Zeinolabedin Mohammadi:** Resources, Validation, Writing – review & editing; **Urban Olsson:** Resources, Writing – review & editing; **Emily Roycroft:** Methodology, Validation, Writing – review & editing; **Łucja Fostowicz-Frelik:** Conceptualization, Visualization, Supervision, Funding acquisition, Writing – review & editing.

Acknowledgements

We are grateful to Lawrence R. Heaney (Field Museum of Natural History, Chicago, Illinois, USA) for granting access to the tissue samples of *Ochotona r. rufescens*.

This work was supported by the National Science Centre (Cracow, Poland) grant No. 2015/18/E/NZ8/00637.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary Materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2024.110246.

References

[1] E. Roycroft, A.J. MacDonald, C. Moritz, A. Moussalli, R.P. Miguez, K.C. Rowe, Museum genomics reveals the rapid decline and extinction of Australian rodents since European settlement, Proc. Natl. Acad. Sci. U. S. A. 118 (2021) e2021390118, doi:10.1073/pnas.2021390118.

[2] A. Saha, M Anwesha, D. Baca, Z. Popovic, U Mohammadi, Olsson, E. Roycroft, L Fostowicz-Frelik, The first complete mitochondrial genome data of the Afghan pika *Ochotona rufescens rufescens* (Lagomorpha), near the type locality, Mendeley Data V3 (2023), doi:10.17632/ggmjwhymmf.3.

[3] Z. Mohammadi, F. Ghorbani, M. Aliabadian, A.A. Lissovsky, F. Yazdani Moghaddam, U. Olsson, Multilocus phylogeny reveals habitat-driven cryptic diversity in *Ochotona rufescens* (Ochotonidae), Zool. Scr. 51 (2022) 617–628, doi:10.1111/zsc.12564.

[4] T.E. Lacher Jr, W.J. Murphy, J. Rogan, A.T. Smith, N.S. Upham, Evolution, phylogeny, ecology, and conservation of the clade glires: lagomorpha and rodentia, in: D.E. Wilson, R.A. Mittermeier, T.E. Lacher (Eds.), Handbook of the Mammals of the World Volume 6, Lynx Edicions in Association with Conservation International and IUCN, 2016, pp. 15–60.

[5] A.T. Smith, C. Johnson[IUCN], Ochotona rufescens, The IUCN Red List of Threatened Species, 2016 T41269A45184750 https://dx.doi.org/10.2305/IUCN.UK.2016-3.RLTS.T41269A45184750.en (Accessed 27 September 2023).

[6] A.A. Lissovsky, S.P. Yatsentyuk, E.V. Obolenskaya, N.P. Koju, D. Ge, Diversification in highlands: phylogeny and taxonomy of pikas of the subgenus *Conothoa* (Lagomorpha, Ochotonidae), Zool. Scr. 51 (2022) 267–287, doi:10.1111/zsc.12538.

[7] M. Meyer, M. Kircher, Illumina sequencing library preparation for highly multiplexed target capture and sequencing, Cold Spring Harb. Protoc. (2010), doi:10.1101/pdb.prot5448.

[8] M. Schubert, S. Lindgreen, L. Orlando, AdapterRemoval v2: rapid adapter trimming, identification, and read merging, BMC Res. Notes 9 (2016) 1–7, doi:10.1186/s13104-016-1900-2.

[9] N. Dierckxsens, P. Mardulyn, G. Smits, NOVOPlasty: de novo assembly of organelle genomes from whole genome data, Nucleic Acids Res. 45 (2017) e18, doi:10.1093/nar/gkw955.

[10] H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform, Bioinformatics 26 (2010) 589–595, doi:10.1093/bioinformatics/btp698.

[11] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 genome project data processing subgroup, the sequence alignment/map format and SAMtools, Bioinformatics 25 (2009) 2078–2079, doi:10.1093/bioinformatics/btp352.

[12] K. Okonechnikov, A. Conesa, F. García-Alcalde, Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data, Bioinformatics 32 (2015) 1–3, doi:10.1093/bioinformatics/btv566.

[13] I. Milne, M. Bayer, L. Cardle, P. Shaw, G. Stephen, F. Wright, D. Marshall, Tablet - next generation sequence assembly visualization, Bioinformatics 26 (2009) 401–402, doi:10.1093/bioinformatics/btp666.

[14] A. Donath, F. Jühling, M. Al-Arab, S.H. Bernhart, F. Reinhardt, P.F. Stadler, M. Middendorf, M. Bernt, Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes, Nucleic Acids Res. 47 (2019) 10543–10552, doi:10.1093/nar/gkz833.

[15] S. Greiner, P. Lehwark, R. Bock, OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes, Nucleic Acids Res. 47 (2019) W59–W64, doi:10.1093/nar/gkz238.

[16] T. Nakamura, K.D. Yamada, K. Tomii, K. Katoh, Parallelization of MAFFT for large-scale multiple sequence alignments, Bioinformatics 34 (2018) 2490–2492, doi:10.1093/bioinformatics/bty121.

[17] G. Talavera, J. Castresana, Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments, Syst. Biol. 56 (2007) 564–577, doi:10.1080/10635150701472164.

[18] J. Trifinopoulos, L.T. Nguyen, A. von Haeseler, B.Q. Minh, W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis, Nucleic Acids Res. 44 (2016) 232–235, doi:10.1093/nar/gkw256.

[19] S. Kalyaanamoorthy, B.Q. Minh, T.K.F. Wong, A. Von Haeseler, L.S. Jermiin, ModelFinder: fast model selection for accurate phylogenetic estimates, Nat. Methods 14 (2017) 587–589, doi:10.1038/nmeth.4285.

[20] O. Chernomor, A. Von Haeseler, B.Q. Minh, Terrace aware data structure for phylogenomic inference from supermatrices, Syst. Biol. 65 (2016) 997–1008, doi:10.1093/sysbio/syw037.

[21] D.T. Hoang, O. Chernomor, A. Von Haeseler, B.Q. Minh, L.S. Vinh, UFBoot2: improving the ultrafast bootstrap approximation, Mol. Biol. Evol. 35 (2018) 518–522, doi:10.1093/molbev/msx281.

[22] X.Y. Wang, D. Liang, W. Jin, M.K. Tang, S.Y. Liu, P. Zhang, Out of Tibet: genomic perspectives on the evolutionary history of extant pikas, Mol. Biol. Evol. 37 (2020) 1577–1592, doi:10.1093/molbev/msaa026.