

The complete mitochondrial genome of the extinct Pleistocene horse (*Equus cf. lenensis*) from Kotelny Island (New Siberian Islands, Russia) and its phylogenetic assessment

Artem V. Nedoluzhko^{a,*} , Fedor S. Sharko^{b,c,*}, Eugenia S. Boulygina^{b,*}, Svetlana V. Tsygankova^b, Natalia V. Slobodova^b, Natalia M. Gruzdeva^b, Sergey M. Rastorguev^b, Natalia N. Spasskaya^d and Evgeny N. Maschenko^{e,f}

^aFaculty of Biosciences and Aquaculture, Nord University, Bodø, Norway; ^bNational Research Centre Kurchatov Institute, Moscow, Russia; ^cResearch Center of Biotechnology of the Russian Academy of Sciences, Institute of Bioengineering, Moscow, Russia; ^dZoological Museum of Moscow Lomonosow State University, Moscow, Russia; ^eBorissiak Paleontological Institute, Russian Academy of Sciences, Moscow, Russia; ^fAcademy of the Sciences Sakha (Yakutia) Republic, Yakutsk, Russia

ABSTRACT

The complete mitochondrial genome from the Pleistocene stallion horse (*Equus cf. lenensis*) which complete skull was found in 1901 on Kotelny Island (New Siberian Archipelago, Sakha Republic, Russia) is published in this paper. The mitochondrial DNA (mtDNA) is 16,584 base pairs (bp) in length and contained 13 protein-coding genes, 2 rRNA genes, 22 tRNA genes. The overall base composition of the genome in descending order was 32.3% – A, 28.5% – C, 13.4% – G, 25.8% – T without a significant AT bias of 58.2%.

ARTICLE HISTORY

Received 5 October 2019
Accepted 27 October 2019

KEYWORDS

Mitochondrial genome;
Pleistocene horse; Equidae;
ancient DNA; evolution

Lenskaya or Lena horse (*Equus cf. lenensis* Russanov, 1968) is extinct Equidae species, which roamed in the Northeastern Siberia during the Late Pleistocene. Presumably, this species lived up to the Middle Holocene in the Arctic zone of Yakutia (northern of 70° north latitude); the latest remains of a frozen mummy is dated by 5450–5310 cal BP (Boeskorov et al. 2014; 2018).



The investigated Lena horse specimen was collected in 1901 on the south coast of the Kotelny Island (New Siberian Archipelago, Sakha Republic, Russia) by Russian Polar Expedition of 1900–1902 that was led by the Russian Arctic explorer Eduard von Toll (1858–1902). The first description of the Late Pleistocene mammals collection from the Kotelny Island was conducted by M. V. Pavlova (Pavlova 1906). The specimen (part of the cranium in a good safe, that was belonged to a stallion by 5–8 years age old) is stored in the Mammal's Laboratory of the Borissiak Paleontological Institute, Russian Academy of Sciences (collection number PIN 301/1). The absolute age (AMS) of the sample established by collagen from the lower jaw bone is 21,105 ± 55 cal BP (IGAN/CIR UG6966). Morphological characters of the PIN 301/1 specimen allow to define it as *Equus cf. lenensis*.

Lena horse DNA was extracted from the bone powder at the ancient DNA facilities of the National Research Center “Kurchatov Institute” (Moscow, Russia), following the methodology described previously (Orlando et al. 2013).

DNA-library was prepared using an Ovation® Ultralow Library System V2 (NuGEN, Redwood City, CA). Amplified DNA library was quantified using a high-sensitivity chip on a 2100 Bioanalyser instrument (Agilent Technologies, Palo Alto, CA). To sequence, the DNA-library of the *E. cf. lenensis* specimen the S2 flowcell of Illumina Novaseq6000 genome analyzer (Illumina, San Diego, CA) was used with paired-end reads with 150 base pairs (bp) length.

307,010,901 Illumina paired-end reads were generated for DNA library for PIN 301/1 Pleistocene horse specimen. Obtained Illumina reads were processed through PALEOMIX pipeline (Schubert, Ermini, et al. 2014), mapping was done against the *E. caballus* reference sequence (assembly EquCab3.0) using Bowtie 2 under the “very-sensitive” and “rescale” options. We used mapDamage2 (Jonsson et al. 2013), that implemented in PALEOMIX, to model postmortem DNA damage from nucleotide misincorporation patterns (Figure 1(A)).

5,732,968 reads were filtered by PALEOMIX then were used for *de novo* assembling of mitochondrial DNA sequence by a Norgal software package (Al-Nakeeb et al. 2017). The resulting consensus sequence was annotated using the MITOS (Bernt et al. 2013). The obtained annotation was then used to define partitions in the subsequent phylogenetic analysis.

CONTACT Artem V. Nedoluzhko  artem.nedoluzhko@nord.no  Faculty of Biosciences and Aquaculture, Nord Universitetet, Bodo, 8049 Norway

*These authors equally contributed to this work.

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

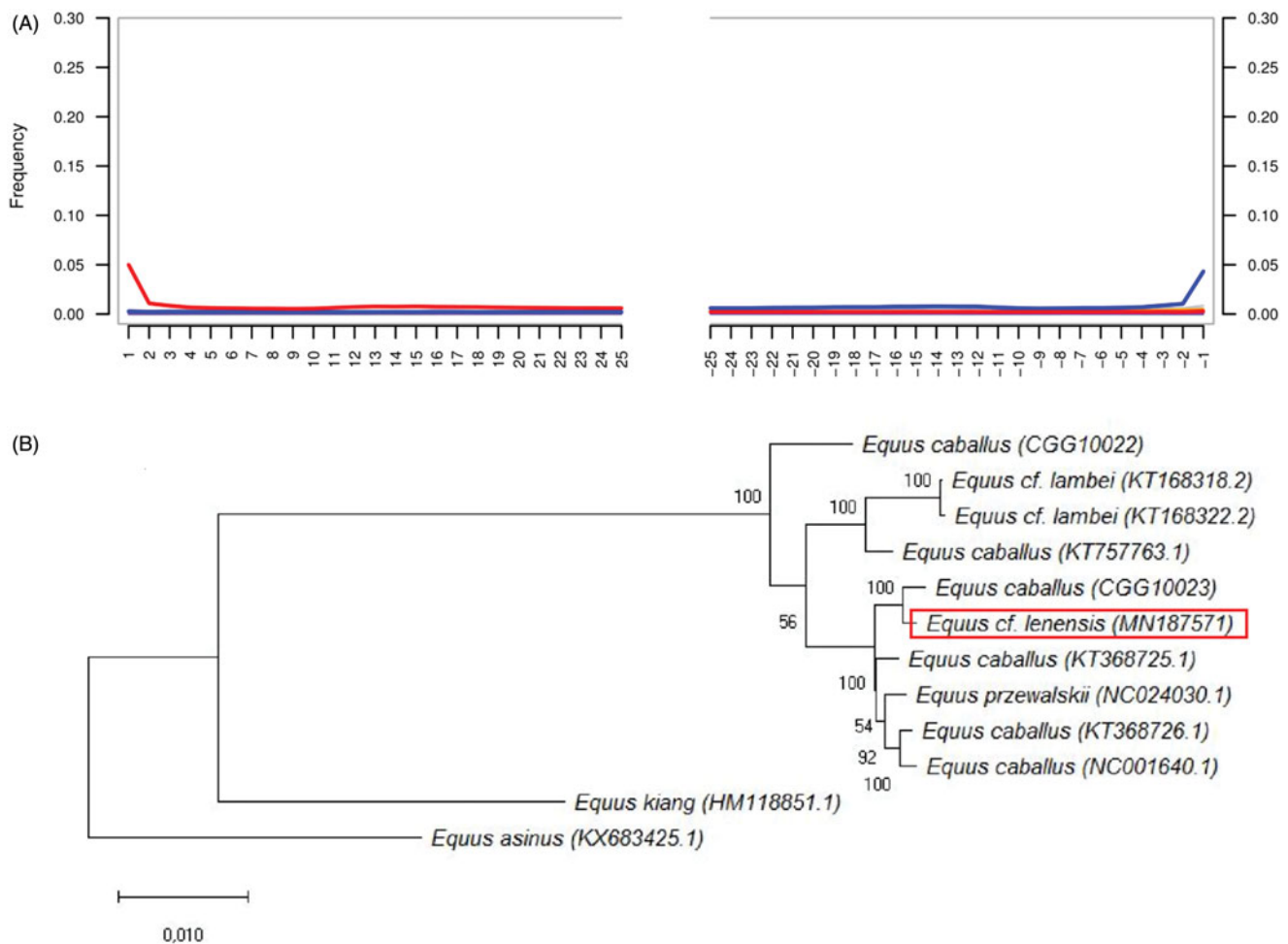


Figure 1. (A) Postmortem DNA damage patterns in DNA reads that were implemented for de novo assembling of *Equus cf. lenensis* mitochondrial DNA sequence. (B) Maximum likelihood phylogenetic tree reconstruction of the Equidae species, including the extinct Pleistocene *Equus cf. lenensis* horse based on their CDS.

As result, the mitogenome of Pleistocene horse (*E. cf. lenensis*) consists of 16,584 bp (GenBank accession number: MN187571) and includes 13 protein-coding genes (PCGs), 2 rRNA genes and 22 tRNA genes.

Eleven of the 13 PCGs (*NAD6*, *NAD4*, *NAD5*, *NAD4L*, *COB*, *NAD1*, *COX2*, *ATP8*, *ATP6*, *COX3*, *COX1*) used ATG as start codon, another two (*NAD2*, *NAD3*) used ATA. Nine genes (*COX1*, *COX2*, *ATP8*, *ATP6*, *COX3*, *NAD4L*, *NAD4*, *NAD5*, *NAD6*) ended with a TAA stop codon, but for two ones of them (*COX3* and *NAD4*) TAA stop codon is completed by the addition of 3' A residues to the mRNA, another three (*NAD1*, *NAD2*, *NAD3*) used TAG and *COB* gene ended with a AGA stop codon.

The phylogenetic analysis for coding part of mitochondrial genome sequences (CDS) was performed for the *E. cf. lenensis* (MN187571) and other Equidae species:

- Extinct horse *E. cf. lambei* – KT168318.2, and KT168322.2 (Heintzman et al. 2017);
- Middle Pleistocene horse – KT757763.1 (Orlando et al. 2013);
- Three Late Pleistocene horses – CGG10022, and CGG10023 (Schubert, Jonsson et al. 2014), also as Batagai horse – KT368725.1 (Librado et al. 2015);
- Modern *E. caballus* – KT368726.1 (Librado et al. 2015), and NC_001640.1 (Xu and Arnason 1994);

- Przewalski's horse – *E. przewalskii* (NC_024030.1).

The CDS of Tibetan wild ass (*E. kiang*) (HM118851.1), and Qingyang donkey – *E. asinus* (KX683425) were used as an outgroup (Luo et al. 2011; Guo et al. 2017). The phylogenetic relationships were reconstructed using Maximum likelihood analysis was conducted using RAxML. Nodal support was evaluated using 1000 replications of rapid bootstrapping implemented in RAxML (Stamatakis et al. 2008). Maximum likelihood phylogenetic tree reconstruction (Figure 1(B)) has been drawn in MEGA X (Kumar et al. 2018).

This study represents the first mitochondrial DNA analysis of *E. cf. lenensis* phylogenetic position based on the available mitochondrial genomes of Equidae species for comparison. Our data show that Lena horse does not represent a separate mitochondrial lineage within *Equus* genus, also as previously published Middle and Late Pleistocene extinct horses from Siberia (Orlando et al. 2013; Schubert, Jonsson et al. 2014; Librado et al. 2015).

Acknowledgments

This work was carried out using high-performance computing resources of federal center for collective usage at NRC "Kurchatov Institute", <http://computing.kiae.ru/>. E. Mashchenko was supported by the Russian Academy of Sciences program: "The evolution of the organic world. The

role and influence of planetary processes. Project: “Comparative analysis and synthesis of paleontological and molecular genetic data on the phylogeny of key groups of terrestrial vertebrates”.

Disclosure statement

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

ORCID

Artem V. Nedoluzhko  <http://orcid.org/0000-0001-7040-0892>

References

- Al-Nakeeb K, Petersen TN, Sicheritz-Ponten T. 2017. Norgal: extraction and de novo assembly of mitochondrial DNA from whole-genome sequencing data. *Bmc Bioinformatics*. 18(1):510.
- Bernt M, Donath A, Juhling F, Externbrink F, Florentz C, Fritzsche G, Putz J, Middendorf M, Stadler PF. 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol Phylogenet Evol*. 69(2): 313–319.
- Boeskorov GG, Potapova OR, Mashchenko EN, Protopopov AV, Kuznetsova TV, Agenbroad L, Tikhonov AN. 2014. Preliminary analyses of the frozen mummies of mammoth (*Mammuthus primigenius*), bison (*Bison priscus*) and horse (*Equus* sp.) from the Yana-Indigirka Lowland, Yakutia, Russia. *Integr Zool*. 9(4):471–480.
- Boeskorov GG, Potapova OR, Protopopov AV, Plotnikov VV, Maschenko EN, Shchelchkova MV, Petrova EA, Kowalczyk R, van der Plicht J, Tikhonov AN. 2018. A study of a frozen mummy of a wild horse from the Holocene of Yakutia, East Siberia, Russia. *Mamm Res*. 63(3): 307–314.
- Guo X, Bao PJ, Pei J, Ding XZ, Liang CN, Yan P, Lu DX. 2017. Complete mitochondrial genome of Qingyang donkey (*Equus asinus*). *Conservation Genet Resour*. 9(2):269–271.
- Heintzman PD, Zazula GD, MacPhee R, Scott E, Cahill JA, McHorse BK, Kapp JD, Stiller M, Wooller MJ, Orlando L, et al. 2017. A new genus of horse from Pleistocene North America. *Elife*. 66:e29944.
- Jonsson H, Ginolhac A, Schubert M, Johnson PL, Orlando L. 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*. 29(13):1682–1684.
- Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol*. 35(6):1547–1549.
- Librado P, Der Sarkissian C, Ermini L, Schubert M, Jonsson H, Albrechtsen A, Fumagalli M, Yang MA, Gamba C, Seguin-Orlando A, et al. 2015. Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to subarctic environments. *Proc Natl Acad Sci USA*. 112(50):E6889–6897.
- Luo Y, Chen Y, Liu FY, Jiang CH, Gao YQ. 2011. Mitochondrial genome sequence of the Tibetan wild ass (*Equus kiang*). *Mitochondr DNA*. 22(1–2):6–8.
- Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M, Schubert M, Cappellini E, Petersen B, Moltke I, et al. 2013. Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*. 499(7456):74–78.
- Pavlova MV. 1906. Description of the fossil mammals collected on Russian Arctic Expeditions in 1900–1903. *Mem Imp St Petersburg Acad*. 21(1):1–40.
- Schubert M, Ermini L, Der Sarkissian C, Jonsson H, Ginolhac A, Schaefer R, Martin MD, Fernandez R, Kircher M, McCue M, et al. 2014. Characterization of ancient and modern genomes by SNP detection and phylogenomic and metagenomic analysis using PALEOMIX. *Nat Protoc*. 9(5):1056–1082.
- Schubert M, Jonsson H, Chang D, Der Sarkissian C, Ermini L, Ginolhac A, Albrechtsen A, Dupanloup I, Foucal A, Petersen B, et al. 2014. Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc Natl Acad Sci USA*. 111(52):E5661–5669.
- Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol*. 57(5):758–771.
- Xu X, Arnason U. 1994. The complete mitochondrial DNA sequence of the horse, *Equus caballus*: extensive heteroplasmy of the control region. *Gene*. 148(2):357–362.