# Supplementary information

## Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to arctic environments

Pablo Librado[1*], Clio Der Sarkissian[1*], Luca Ermini[1], Mikkel Schubert[1], Hákon Jónsson[1], Anders Albrechtsen[2], Matteo Fumagalli[3], Melinda A. Yang[4], Cristina Gamba[1], Andaine Seguin-Orlando[1,5], Cecilie D. Mortensen[5], Bent Petersen[6], Cindi A. Hoover[7], Belen Lorente-Galdos[8], Artem Nedoluzhko[9], Eugenia Boulygina[9], Svetlana Tsygankova[9], Markus Neuditschko[10], Vidhya Jagannathan[11], Catherine Thèves[12], Ahmed H. Alfarhan[13], Saleh A. Alquraishi[13], Khaled A.S. Al-Rasheid[13], Thomas Sicheritz-Ponten[6], Ruslan Popov[14], Semyon Grigoriev[15], Anatoly N. Alekseev[15], Edward M. Rubin[7], Molly McCue[16], Stefan Rieder[10], Tosso Leeb[11], Alexei Tikhonov[17], Eric Crubézy[12], Montgomery Slatkin[4], Tomas Marques-Bonet[8], Rasmus Nielsen[18], Eske Willerslev[1], Juha Kantanen[19,20], Egor Prokhortchouk[9], Ludovic Orlando[1,12].

[1]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350K Copenhagen, Denmark;
[2]The Bioinformatics Centre, Department of Biology, University of Copenhagen, Copenhagen, Denmark;
[3]UCL Genetics Institute, Department of Genetics, Evolution and Environment; University College London; London, WC1E 6BT; UK
[4]Department of Integrative Biology, VLSB 3060, University of California, Berkeley, CA 94720-3140, USA;
[5]National High-Throughput DNA Sequencing Centre, University of Copenhagen, Øster Farimagsgade 2D, 1353K Copenhagen, Denmark;
[6]Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, 2800 Lyngby, Denmark;
[7]DOE Joint Genome Institute, Walnut Creek, California 94598, USA;
[8]ICREA (Universitat Pompeu Fabra/Consejo Superior de Investigaciones Cientificas), 08003 Barcelona, Spain; Centro Nacional de Análisis Genómico, 08028 Barcelona, Spain;
[9]National Research Centre Kurchatov Institute, 1, Akademika Kurchatova, Moscow, 123182, Russian Federation;
[10]Agroscope, Swiss National Stud Farm, 1580 Avenches, Switzerland;
[11]Institute of Genetics, University of Bern, 3001 Bern, Switzerland;
[12]Université de Toulouse, University Paul Sabatier (UPS), Laboratoire AMIS, CNRS UMR 5288, 37 allées Jules Guesde, 31000 Toulouse, France;
[13]Zoology Department, College of Science, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia;
[14]Yakutian Research Institute of Agriculture, 677002 Yakutsk, Sakha, Russia
[15]North-Eastern Federal University, Yakutsk, Russian Federation;
[16]College of Veterinary Medicine, University of Minnesota, 1365 Gortner Avenue, St Paul, MN 55108, USA;
[17]Zoological Institute of Russian Academy of Sciences, Universitetskaya nab. 1, 199034 Saint-Petersburg, Russia;
[18]Center for Theoretical Evolutionary Genomics, University of California, Berkeley, Berkeley, California, USA;
[19]Biotechnology and Food Research, MTT Agrifood Research Finland, Jokioinen, Finland;
[20]Department of Biology, University of Eastern Finland, Kuopio, Finland;
*Contributed equally.

## AUTHOR CONTRIBUTIONS

- LO initially conceived and headed the project.
- CT, ANA, AT, EC, and EP provided ancient horse samples.
- JK and EP provided access to modern Yakutian horse samples.
- SR and TL provided access to the Franches-Montagnes horse genomes.
- MM provided access to the Morgan, American Quarter Horse and Standardbred horse genomes.
- ASO, KM, BP, CH, CT, AHA, SAA, KASAR, TSP, SG, AA, AN, EB, EMR, MM, SR, TL, AT, EC, MS, TMB, EW, JH, EP, LO provided reagents and materials.
- LO performed ancient DNA extractions and constructed ancient DNA libraries, with input from CAH, CD and LE.
- JK and RP coordinated sampling and DNA extraction from modern Yakutian horses (Yak1-Yak9).
- AN, EB and EP extracted DNA from modern Yakutian horses (Horse1-Horse3), and performed sequencing at the Kurchatov Institute, Russia.
- ST constructed DNA libraries from modern Yakutian horses (Horse1-Horse3) at the Kurchatov Institute, Russia.
- CD and LE constructed DNA libraries from modern Yakutian horses (Yak1-Yak9).
- ASO and KM sequenced DNA libraries at the Danish National High-Throughput DNA Sequencing Center, Copenhagen, Denmark.
- MN and VJ performed modern DNA extractions and constructed modern DNA libraries for Franches-Montagnes horses sequenced at the Institute of Genetics of the University of Bern.
- MSc performed read mapping, variant calling, and functional characterization, and generated genome alignments; performed DNA damage pattern analyses, BAM rescaling and trimming; did phylogenomics reconstructions based on exome data, with input from BP.
- CD and LE examined DNA damage patterns.
- CD and LE performed metagenomic analyses.
- LO performed phylogenetic reconstruction of mitochondrial genomes with input from LE.
- CD and LE performed phylogenetic reconstructions based on Y-chromosome data.
- HJ performed genotype-based Principal Component Analyses.
- LO performed genotype likelihood-based Principal Component Analyses, NGSAdmix analyses.
- LO computed $F_{ST}$ values, and HJ detected outlier regions.
- MSc performed the analyses of segmental duplications, with input from BL and TMB.
- MSc and HJ estimated genome-wide heterozygosity and inbreeding coverage.
- AA estimated sequencing error rates and performed admixture tests based on the D-statistics.
- CD analysed the results of the admixture tests based on the D-statistics.
- LE did Pairwise-Sequential Markovian Coalescent Inference analyses.
- HJ performed TreeMix and f3-statistics analyses.

- MAY and MSl designed and performed projection analyses, with input from CD and LO.
- PL investigated the relative contribution of *cis*-regulatory elements and protein-coding regions to adaptation, with input from LO.
- LO performed enrichment analyses.
- CD, PL and LO examined outputs enrichment analyses.
- PL, CD, LE, MSc, MAY and LO wrote the supplementary information, with input from CG.
- LO wrote the manuscript, with input from all coauthors.

## TABLE OF CONTENT

## FIGURE LIST

## Section 1: Sample Information

### *1.1 Environmental conditions and inhabitants of Yakutia*

Yakutia, or the Sakha Republic (Russia), is located at the most North-Eastern part of the Russian territory and shows very low population densities. Climatic conditions are extreme and represent the coldest region in the northern hemisphere, with winter temperature records sometimes below -70°C and almost half of the territory consisting of permafrozen soils. The population of Yakutia mainly consists of two human groups, the Russians and the Yakuts. The latter migrated in the region probably between the 13th-15th century (1, 2), leaving their native range in the south, supposedly in the Altai-Sayan and/or Baykal regions (3, 4). Present-day Yakuts constitute a semi-nomadic group specialized in animal husbandry, especially of horse and cattle (5).

Historically, Yakutian horses have represented the main branch of the economy of the Yakuts (6), providing meat, milk, transportation, as well as primary products for clothes (mainly hairs, tendons, and skins). Today, Yakutian horses are mostly exploited as sources of meat and milk, with most males being slaughtered at the end of their first autumn, but still represent an essential part of the Yakutian culture, as shown by the horse symbol featured on the national Yakutian flag.

### *1.2 Yakutian horses*

In this study, we investigate the evolutionary history of Yakutian horses, the most northerly distributed domesticated horse breed in the planet (above the 70 North latitude line), which survive extremely cold conditions with minimal human attention. These horses live in the open air all year round, grazing on vegetation lying under deep snow cover for 7-8 months (6).

They show characteristic morphological traits likely reflecting their adaptation to the extreme climatic conditions of Yakutia, such as a long back, a coarse head, a straight neck, a deep and wide chest (7). They are massively built, with average height at withers not higher than 1.40 m and carry a rump generally higher than withers. Yakutian horses have short limbs and are extremely hairy with small hoof, ears and collar. The mane and tail are very thick and long, and the body hair is thick and long and can reach 10 cm in winter (7).

Yakutian horses also exhibit characteristic metabolic features. For example, during the extremely short period of vegetation growth (late May – late September), they can accumulate important fat reserves and regulate their own metabolic needs according to the seasonal conditions. They have lower winter metabolic levels than outbred horses (8), and show seasonality patterns in aspartate aminotransferase and alanine transaminase activities, which likely reflects the increased participation of carbohydrate metabolism in spring for supporting metabolism recovery, higher energy expenditure and fetal growth (9).

### *1.3 Sampling*

Hair samples from nine Yakutian horses (Yak1 to Yak9) living in rural localities of Yakutia, were collected in 2001 by Pr. Juha Kantanen in the Eveno-Bytantaj District. Three additional hair samples (Horse1 to Horse3) from Yakutian horses were collected by Dr. Alexei Tikhonov in February 2012. These samples originate from three different regions (Kazachie village, in the lower course of the

Yana river; the Srednekolymsk district, in the middle course of the Kolyma river; and, the Betenkes village by the Adycha river, close to Verkhoyansk) and are considered to represent the descendants of the most ancient Yakutian breeds **(Table S1.1)**.

We also sampled a series of nine ancient horse bones and teeth excavated in Yakutia, corresponding to specimens that lived prior to (~5,000-5,500 years Before Present, yBP) and after (~300 yBP) the arrival of the Yakuts in the area **(Table S1.2)**. Samples pre-dating the Yakut settlement include two samples: sample "Yukagir" excavated at the "Yuka" site, on the southern coast of the Dmitry Laptev Strait (between Bolshoi Lyakhowski Island and mainland, Novosoborskie islands, Yakutia, Russia) in the Oyagossky Yar area, and radiocarbon-dated to 4,630 ± 35 uncalibrated yBP (GrA-540209), and sample "Batagai" excavated from the Batagai site, Verkhoyansk District, and radiocarbon-dated to 4,450 ± 35 uncalibrated yBP (Gr 50842). These dates correspond to 3,517-3,351 Before Christ (BC) and 2,939-3,337 BC, respectively, using the IntCal13 calibration in OxCal 4.2 online (http://www.c14.arch.ox.ac.uk/oxcal.html). Six of the samples post-dating the arrival of Yakuts (CGG101392, CGG101393, CGG101394, CGG101395, CGG101396, CGG101397) were presented in (10) and were dated ~200-300 years yBP. Sample ODJ6 was excavated from the Odjuluun site during the MAFSO 2006 expedition (French Archaeological Mission in Oriental Siberia) and was dated to ~250 yBP.

## *1.4 Supplementary Tables for Section 1*

### Table S1.1. Sample Information for the modern Yakutian horses analysed in this study.

| Horse ID | Gender | Tissue type | Geographical coordinates | Region of Origin |
|---|---|---|---|---|
| Yak1 | Male | Hair | 68°11N, 131°41E | Kustur, Sakha Republic, Russia |
| Yak2 | Female | Hair | 68°11N, 131°41E | Kustur, Sakha Republic, Russia |
| Yak3 | Female | Hair | 68°11N, 131°41E | Kustur, Sakha Republic, Russia |
| Yak4 | Male | Hair | 68°11N, 131°41E | Kustur, Sakha Republic, Russia |
| Yak5 | Female | Hair | 68°11N, 131°41E | Kustur, Sakha Republic, Russia |
| Yak6 | Female | Hair | 67°47N, 130°24E | Batagay-Alyta, Sakha Republic, Russia |
| Yak7 | Female | Hair | 67°47N, 130°24E | Batagay-Alyta, Sakha Republic, Russia |
| Yak8 | Male | Hair | 67°47N, 130°24E | Batagay-Alyta, Sakha Republic, Russia |
| Yak9 | Male | Hair | 67°47N, 130°24E | Batagay-Alyta, Sakha Republic, Russia |
| Horse1 | n/a | Hair | n/a | Kazachie village, Sakha Republic, Russia |
| Horse2 | n/a | Hair | n/a | Srednekolymsk, Sakha Republic, Russia |
| Horse3 | n/a | Hair | n/a | Betenkes, Sakha Republic, Russia |

### Table S1.2. Sample Information for the ancient Yakutian horses analysed in this study.

| Horse ID | Tissue type | Age | Collection date | Site name and coordinates | Region of origin | Reference |
|---|---|---|---|---|---|---|
| Yukagir | Bone | 4,630 ± 35 uncalibrated yBP | August 2012 | Yuka: 72°42N, 142°50E | Verkhoyansk | This study |
| Batagai | Bone | 4,450 ± 35 uncalibrated yBP | February 2012 | Batagai: 67°34N, 134°46E | Verkhoyansk | This study |
| CGG101392 | Tooth | 19th century AD | MAFSO 2012 | OursSire2: 66°51N, 131°45.21E | Verkhoyansk | Der Sarkissian et al. 2014 (10) |
| CGG101393 | Bone | 18th /19th century AD | MAFSO 2011 | Bakhtakh: 67°09N, 134°31.01E | Verkhoyansk | Der Sarkissian et al. 2014 (10) |
| CGG101394 | Tooth | 19th century AD | MAFSO 2012 | Yakutia: near 66°53N, 131°51E | Verkhoyansk | Der Sarkissian et al. 2014 (10) |
| CGG101395 | Bone | 18th century AD | MAFSO 2012 | Tysarastaak2: 66°54N, 131°50E | Verkhoyansk | Der Sarkissian et al. 2014 (10) |
| CGG101396 | Tooth | 19th century AD | MAFSO 2010 | Targana1: 66°59 N, 132°59E | Verkhoyansk | Der Sarkissian et al. 2014 (10) |
| CGG101397 | Bone | 19th century AD | MAFSO 2012 | Tumeski: 66°56N, 131°55E | Verkhoyansk | Der Sarkissian et al. 2014 (10) |
| ODJ6 | Bone | First half of 18th century AD | MAFSO 2006 | Odjuluun: 61°52N, 132°24E | Central Yakutia | This study |

“yBP”: years Before Present; “MAFSO”: Mission Archéologique Française en Sibérie Orientale led by Prof. Eric Crubézy.

# 2 Section 2: Genome sequencing

## *2.1 DNA extraction and library preparation for modern Yakutian horses*

Genomic DNA from the nine modern Yakutian horse samples, Yak1 to Yak9, was extracted at the Institute of Veterinary Medicine and Animal Sciences (Estonian University of Life Sciences, Estonia). DNA extraction was performed on hair roots using Puregene® Genomic DNA Purification Kit (Gentra Systems. USA), following the manufacturer's protocol. Genomic DNA from three modern Yakutian horse samples, Horse1 to Horse3, was extracted at the Kurchatov Institute (Moscow, Russia) from hair root and shaft, using the methods described by Gilbert and colleagues (11), with the modifications from (12). More specifically, hair shafts were sliced into small pieces and disgested at 55°C for 24 hours in 5ml of digestion buffer (10mM Tris, 10mM NaCl, 5mM CaCl, 2.5mM EDTA, 1% SDS, 10mb/ml DTT and 0.5mb/ml proteinase K). The digestion supernatant was recovered by spinning at 2,500g for 2 minutes and concentred to 50-100 µL using 30KDa centricons. The final DNA extract was obtained following a terminal purification of MinElute columns (Qiagen).

Indexed Illumina DNA libraries were constructed from the Yak1-Yak9 DNA extracts at the Centre for GeoGenetics (University of Copenhagen, Denmark) in laboratory facilities dedicated to the molecular analysis of fresh DNA. Overall, we followed the procedure for constructing libraries by blunt-end ligation described in the Supplementary Information section 3.1.b.3 of (13) with slight modifications. Briefly, 1 µg of DNA extract was sheared with the Bioruptor (Diagenode) using 4 cycles of 15 seconds on high energy, and 90 seconds off. Sheared DNA extracts were then purified on MinElute columns (Qiagen) and eluted in 22 µL of EB. The whole eluted sheared DNA was used for library construction using the NEBNext Quick DNA Library Prep Master Mix Set for 454 (New England BioLabs, ref: #E60700), with the following modifications. We worked with 25 µL reaction volumes at each step and used a final concentration of 500 nM of Illumina multiplex blunt-end adapters. We used two different adapters that we prepared according to Meyer and Kircher (14). After the end-repair reaction, we used a MinElute PCR purification kit (Qiagen) with an elution in 16 µL EB buffer following a 10 min-incubation at 37°C. Ligation was then performed for 20 minutes at 20°C following a purification step in MinElute columns (Qiagen) with elution in 22 µL of buffer EB (10 min at 37°C). A final fill-in reaction step was performed at 37°C for 20 minutes adding 4 µL of Bst DNA polymerase enzyme mix to each library, consisting of 2.5 µL of adapter-fill in reaction buffer and 1.5 µL of Bst DNA polymerase. The Bst polymerase was then inactivated by incubation at 80°C for 20 min. DNA library were obtained in a final volume of 25 µL.

DNA libraries were then PCR-amplified in a final volume of 25 µL containing: 12.5 µL of DNA library, 1 µL Illumina inPE1.0 primer (25 µM, 5′-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T), 1 µL Multiplex Index Primer (25 µM, 5′-CAA GCA GAA GAC GGC ATA CGA GAT NNN NNN GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG, where the N stretch corresponds to a 6 nucleotides index tag), 5 units of AmpliTaq Gold DNA polymerase (Life Technologies), 1x Gold Buffer, 25 mM $MgCl_2$, 1 mg/mL BSA, and 0.25 µM of each dNTP. PCR cycling conditions consisted of an initial denaturation and enzyme activation for 10 minutes at 92°C, followed by 9 cycles of 30 second denaturation at 92°C, 30 seconds annealing at 60°C and 30 second elongation at 72°C, and terminated by a final 7 minute elongation step at 72°C. The resultant amplified DNA fragments were then purified on MinElute columns (Qiagen), and eluted in 25 µL of EB buffer. A blank library was built by replacing DNA samples with EB buffer to monitor contamination and a PCR blank was also included. All blanks were negative after library amplification.

For samples Horse1-Horse3, DNA libraries were prepared at the Kurchatov Institute (Moscow, Russia) using a NEBNext Quick DNA Library Prep Master Mix set for 454 (New England Biolabs), with adapter primers on an Illumina Sequencing Platform and following the manufacturer's instructions. DNA libraries were then PCR-amplified in a final volume of 50 µL containing: 15 µL of DNA library, 5 µL NEBNext®universal primer (10 µM 5′-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC T-3′), 5 µL NEBNext® Multiplex Index Primer (10 µM, 5′-CAA GCA GAA GAC GGC ATA CGA GAT NNN NNN GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T-3′, where the N stretch corresponds to a 6 nucleotides index tag), 2x NEBNext® Q5® Hot Start HiFi PCR Master Mix (NEB) cycling conditions consisted of 3 minutes at 98°C, followed by 7 cycles of 30 seconds denaturation at 98°C and 75 seconds annealing/extension at 65°C, and terminated by a final 5 minute elongation step at 65°C. The resultant amplified DNA fragments were then purified with 45mkl AMPure XP Beads (Beckman Coulter), and eluted in 25 µL H2O. A blank library was built by replacing DNA samples with EB buffer to monitor contamination and a PCR blank was also included. All blanks were negative after library amplification. The purity and amount of DNA libraries were evaluated using a 2100 Bioanalyser (Agilent, USA) and HS Qubit (Invitrogen, USA).

### *2.2 DNA extraction and library preparation for ancient horses from Yakutia*

We performed DNA extraction and library preparation from ancient bones and teeth at the ancient DNA facilities of the Centre for GeoGenetics, University of Copenhagen, Denmark, following strict procedures for limiting contamination by modern DNA. Importantly, those facilities are located in separate buildings (i.e., at about a five minute walking distance) from the post-PCR facilities, where DNA of the modern samples was handled.

For bones, the outer surface was first abraded to a depth of 1-2 mm and bone powder was generated drilling within the sample with a Dremel grinding tool, at low speed. For teeth, powder was obtained by drilling. Digestion of the samples was carried out in 5 mL of 0.5 M EDTA (pH 8.5), 0.1% N-lauryl-sarcosyl, and 1 mg/mL proteinase K (Invitrogen) at 37 °C for 24 hours under rotation. The supernatant was stored and a second digestion was performed on the pellets' leftover from the first digestion. The supernatant from this second digestion was then used for DNA extraction using the silica-based method described in (13) and (10), and the resulting DNA extract was further build into DNA libraries. For samples CGG101392 to CGG101397, the remaining fraction of the first digestion (hereafter referred to as 'simple digestion') was also fully extracted, and prepared into DNA libraries (see below).

For Batagai, a total of 4 indexed TruSeq Illumina DNA libraries were built using the procedure already described in (15) and (16). Briefly, End Repair reaction was performed on 16.5 µL of DNA extract using End-It™ DNA End-Repair Kit (Epicentre) with a first incubation at 4°C for 2 minutes followed by a second incubation at 37°C for 45 minutes. End-repaired DNA templates were then purified with MinElute columns (Qiagen), using 10 volumes of PN buffer (from QIAQuick, Qiagen) and 17.5 µL of Elution Buffer (EB, Qiagen). The whole purified volume was then used for Klenow exo-polyA-tailing (37°C for 30 minutes) and adapter ligation with standard indexed TruSeq adapters. Each DNA library underwent a final purification with Ampure XP beads using 1.8:1 volume ratio between beads and DNA. Purified DNA was eluted in 20 µL of EB solution. Two libraries were then PCR amplified with AmpliTaq Gold DNA polymerase (Life Technologies), and the other two with Accuprime Pfx (Life Technologies). We amplified 7 µL of DNA library in a final PCR volume of 25 µL using 300 nM Illumina PCR primers (PCR primer 1.0: 5'-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GA; PCR primer 2.0: 5'-CAA GCA GAA GAC GGC ATA CGA GAT), a concentration of 1mg/mL BSA, 1µM dXTP (Invitrogen) and 5U of Taq Gold or Accuprime Pfx. For Taq Gold, PCR cycling conditions consisted of a first DNA denaturation at 95 °C for 10 minutes, followed by 15 cycles of denaturation (95°C, 30 seconds), annealing (60°C, 30 seconds) and elongation (72°C, 60 seconds). A final elongation was performed at 72°C for 7 minutes before amplified DNA was purified into 25 µL EB with using Qiagen Minelute purification kit. For Accuprime Pfx, PCR cycling conditions consisted of a first DNA denaturation at 95 °C for 2 minutes, followed by 15 cycles of denaturation (95°C, 15 seconds), annealing (68°C, 30 seconds) and elongation (68°C, 40 seconds). A final elongation was performed at 68°C for 3 minutes before amplified DNA was purified into 25 µL EB with using Qiagen Minelute purification kit.

For sample Batagai, we also prepared one indexed Illumina DNA library based on blunt-end ligation, following the procedure described in (10) and in **section S2.1**, without prior DNA shearing and using PN buffer (from QIAQuick, Qiagen) instead of PB for MinElute (Qiagen) purification steps. The same type of DNA libraries was prepared for samples CGG101392 to CGG101397, ODJ6, and Yukagir, for which no other library types (e.g. TruSeq) were prepared. For samples CGG101392 to CGG101397, one library was built per extract fraction (i.e., following simple or double digestion) as described in (10). DNA libraries were amplified using the same conditions as in **section S2.1**, except that two rounds of amplification were performed as described in (17), with the number of PCR cycles varying from 10 to 12 and 500nM of adapters. We used the whole 25 µL of DNA libraries in the first amplification round, except for sample Batagai (7 µL).

DNA contamination from the laboratory and reagents was monitored through mock DNA extractions and mock library constructions carried out at the same time as for ancient samples. All blanks were negative after final amplification of the libraries.

Our implementation of DNA libraries from fossil remains is detailed in **Table S2.1**, including the amount of fossil material processed, the type/number of DNA libraries constructed, the type of DNA polymerase used, and the number of library amplification PCR cycles performed.

### *2.3 DNA Shotgun sequencing*

Sequencing of the genomes from the modern Yakutian horses Yak1 to Yak9 and from all ancient Yakutian horses was performed at the Danish National High-Throughput DNA Sequencing Centre, Copenhagen, Denmark, where a total number of 9,684,698,721 sequencing reads were produced. DNA libraries were first inspected and quantified on an Agilent 2100 Bioanalyzer High Sensitivity DNA chip. Modern DNA libraries were mixed in different pools and paired-end sequenced (98 bp reads, PE) on Illumina HiSeq 2000/2500 platforms (**Table S2.2**).

The DNA libraries constructed on the three modern Yakutian horses Horse1, Horse2 and Horse3 were sequenced at the Kurtchatov Institute, Moscow, Russia, on the Illumina GAIIx platform, where a total number of 155,582,799 sequencing reads were generated (**Table S2.2**).

Ancient DNA libraries were sequenced in pools that did not include any of the DNA libraries built on modern samples. Additionally, DNA libraries from Batagai and CGG101397 were sequenced separately on Illumina MiSeq and HiSeq2000/2500 platforms, using both paired-end (PE) and single-end sequencing technologies (SE; **Table S2.3**). Library pools included DNA libraries that differed in their barcode sequence by a minimum edit distance of 2 bp. Casava (version 1.8.s, Illumina) was used for basecalling, applying 100% match to the expected index of each library. Reads failing the matching filter were removed before downstream analyses.

## *2.4 Comparative horse dataset*

We compared the complete genomes of Yakutian horses, that we characterized here for the first time, to a panel comprising 32 present-day and ancient horse genomes, representing eleven horse breeds/populations (**Table S2.4**).

Twenty of these were characterized in (18) (European Nucleotide Archive Project number PRJEB10098) and belong to the following domesticated horse breeds: Franches-Montagnes (N=12), Morgan (N=1), American Quarter Horse (later also referred to as "Quarter", N=4) and Standardbred (N=3). Domesticated horse genomes were generated from Illumina TruSeq v2 libraries sequenced using one lane (0.5 lane for the Std_M1009 individual) of 100-bp pair-ended (PE) Illumina HiSeq (v3 chemistry) at the Institute of Genetics of the University of Bern, Switzerland and at the University of Minnesota, USA.

Our comparative panel also included previously published data from seven present-day domesticated horses belonging to the six domesticated breeds (Thoroughbred, Arabian, Standardbred, Norwegian Fjord, Icelandic horses (13, 16), and Mongolian horses (19), as well as from three Przewalski's horses (13, 19). Sequence reads for the Mongolian and Przewalski's horses published in (19) were retrieved from the Sequence Read Archive (SRA). For Mng_D2628 we used runs with SRA accession numbers SRR1167052-3, SRR1167891, SRR1167892 (mate 1 reads only), and SRR1167093. For Mng_D2629, we used runs with accession numbers SRR1167108-10, and SRR1167893. For Prz_D2630 we used runs with accession numbers SRR1167030, SRR1167031 (mate 1 reads only), SRR1167045, SRR1167257, and SRR1167890. For Prz_D2631 we used runs with accession numbers SRR1167258, and SRR1167048-50.

The published data of two Late Pleistocene horses, pre-dating domestication, were also included in the comparative datasets. These horses, labelled CGG10022 and CGG10023, were excavated in the Taymyr peninsula, Russia, and dated to 42,692 ± 891 (UBA-16478) and 16,099 ± 192 calibrated years before present (yBP; UBA-16479), respectively (13, 16). This is equivalent to 39,851-41,633 BC and 13,957-14,341 BC.

## *2.5 Sequence read processing and alignment against reference genomes*

The sequencing reads obtained from present-day and ancient Yakutian horses, as well as from all the horses in our comparative panel, were processed using the PALEOMIX pipeline (20) and following the procedure described in (16). Mapping results are reported in **Tables S2.5-2.6.** The sequencing data are available from the European Nucleotide Archive Project number PRJEB10854.

The three modern Yakutian horses Horse1, Horse2 and Horse3 were not sequenced at sufficient depth (i.e. their average depth-of-coverage was 0.34X, 0.21X and 0.13X, respectively) to enable their complete nuclear genome characterization. For those samples, we thus restricted our analyses to their mitochondrial genome (sequenced at 69.63, 47.66 and 43.94X; see **section S5.1**).

Preliminary screening of libraries constructed from ancient Yakutian horse specimens showed that the Batagai and CGG101397 samples were characterized by endogenous contents compatible with the cost-effective sequencing of their complete genome (their final endogenous content estimates are provided in **Table S2.6** and respectively correspond to 38.22% and 45.39% of the reads sequenced and passing quality filters). These specimens were therefore whole-genome sequenced at relatively high coverage (18.29-20.25X). Two other ancient horses, Yukagir and ODJ6, showed lower endogenous content (0.93% and 9.46%; **Table S2.6**) and were not further sequenced. For these samples, we therefore restricted our analyses to their complete mitochondrial genomes (sequenced at 911.47 and 30.60X; see **section S5.1**). Likewise, specimens CGG101392 to CGG101396 were excluded from further horse genomic analyses, as they showed low endogenous content (**Table S2.6**). As their DNA libraries consisted of a majority of reads of exogenous origin (probably from micro-organisms), these were included in metagenomic analyses (see **section S3**).

Read length distributions were computed for all samples based on collapsed and non-collapsed PE reads. For non-collapsed PE reads, we only considered pairs of reads in which both mates were mapped to the same chromosome, to different strands, and where the mate mapped to the positive strand was located 5' to the mate mapped to the negative strand. Each collapsed read or pair of mates was counted once. To exclude extreme outliers (potentially) resulting from mis-alignments, we truncated the distribution at the 99.9th quantile. The distributions are shown in **Figures S2.1-2.2**.

A ca. 10-bp periodicity was observed for modern specimens Yak1 to Yak9, but was more pronounced for ancient specimens CGG101397 and Batagai. This pattern has been proposed by (15) as a nucleosome protection footprint, possibly resulting from apoptotic degradation of DNA in hair shafts (21) and *post-mortem* fragmentation in bones (15). This pattern is in line with previous observations for a series of other ancient samples, including the Palaeo-Saqqaq Eskimo (22), Mesolithic hunter-gatherers (23), Siberian and North mammoths (24) and Late Pleistocene horses (16).

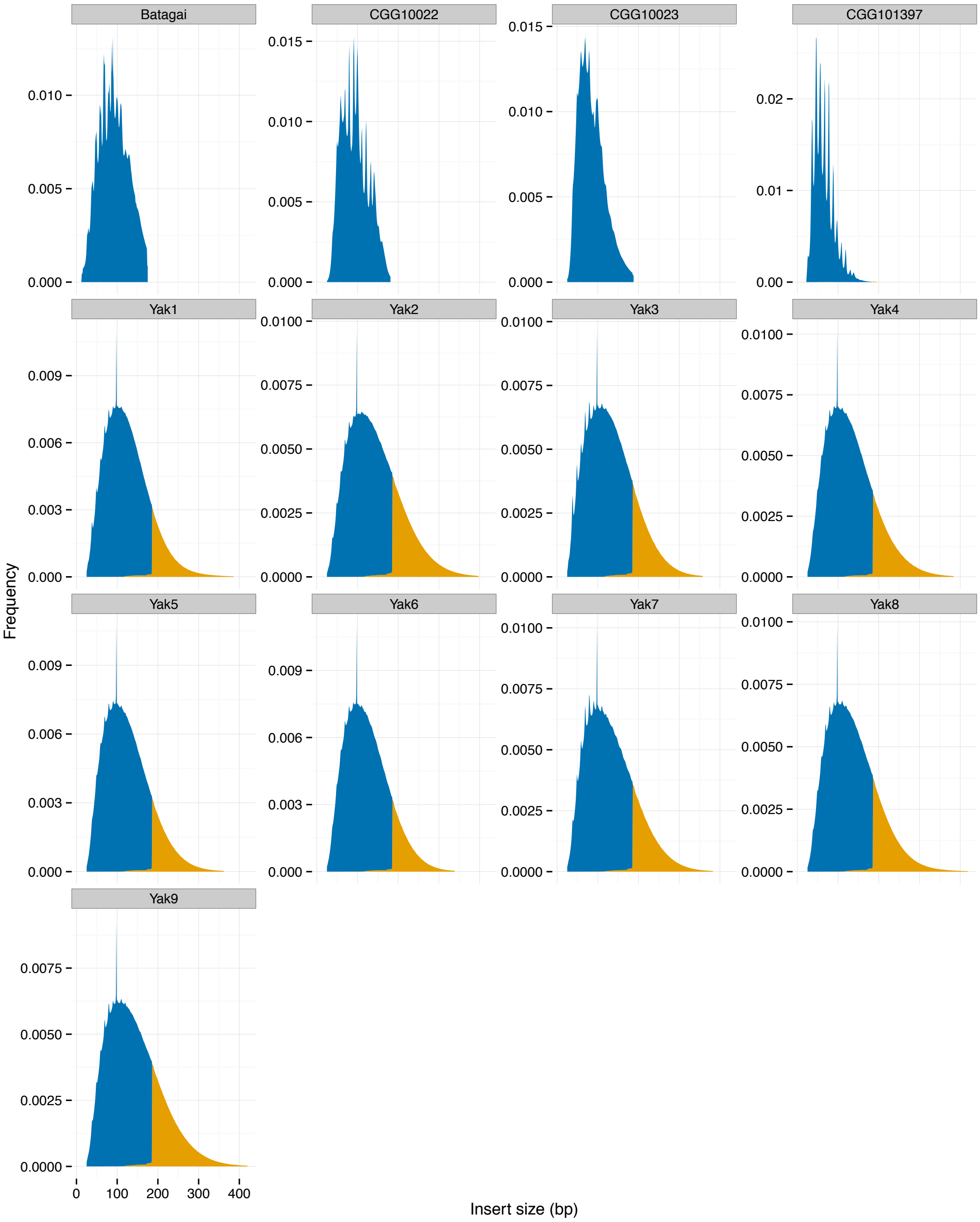


**Figure S2.1. Size distribution of horse library inserts for the present-day and ancient Yakutian specimens.**

Collapsed reads are indicated in blue. The physical distance covered between retained mate pairs is shown in yellow). The distribution of insert sizes observed for two Late Pleistocene horse genomes previously characterized, CGG10022 and CGG10023 (16), are shown for comparison.

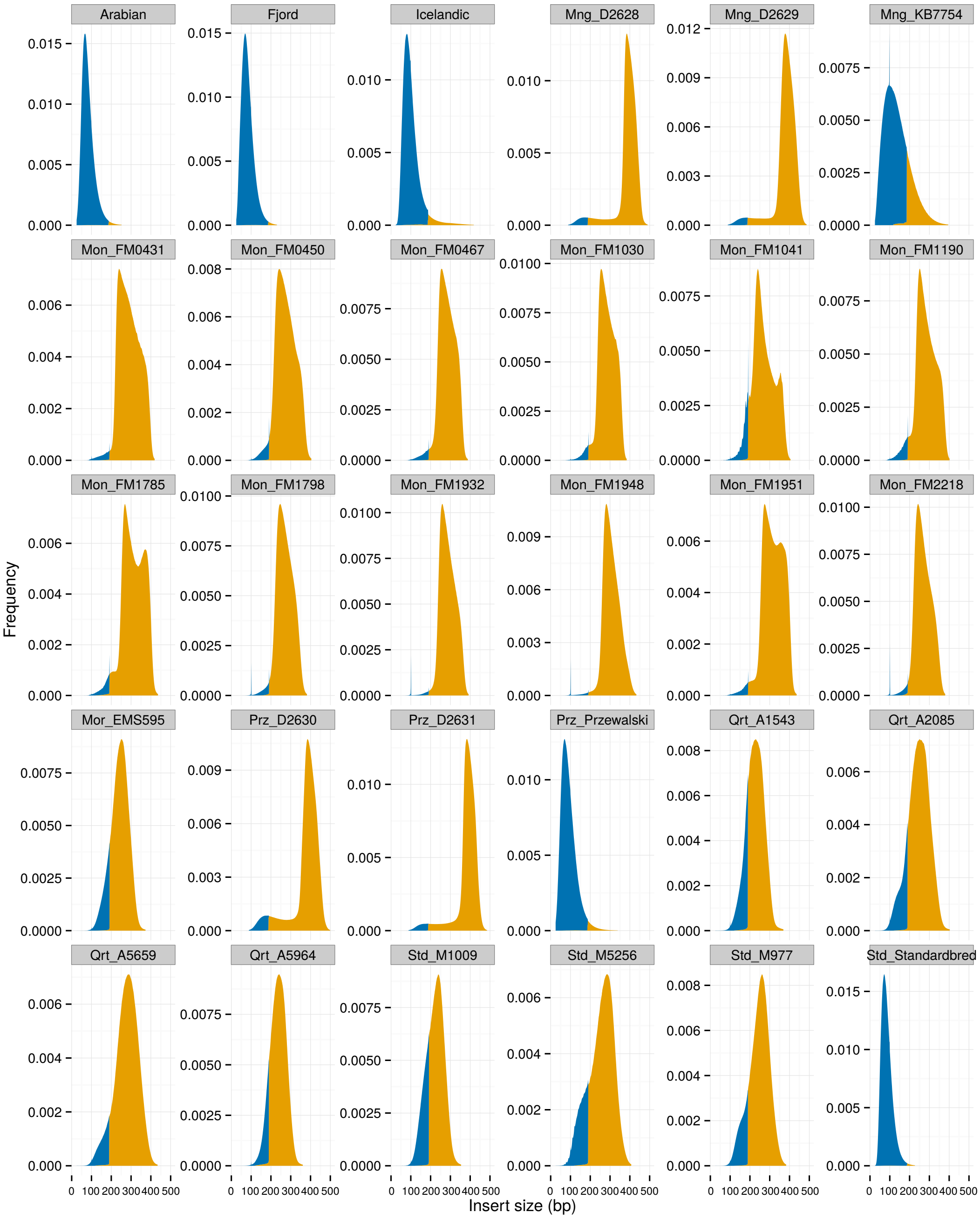


**Figure S2.2. Size distribution of horse library inserts for the comparative dataset of modern horses.** The Arabian, Fjord, Icelandic, St_Standardbred, Thoroughbred (Twilight) and Prz_Przewalski genomes were characterized in (13); Mng_D2628, Mng_D2629, Prz_D2630 and Prz_D2631 were characterized in (19); the others were characterized in (18). See **Figure S2.1** for additional captions.

## *2.6 Characterizing and correcting the sequencing patterns of post-mortem damage of ancient samples*

We assessed the presence of sequence patterns typical of *post-mortem* damage in each library generated from the ancient samples Batagai, CGG101397, Yukagir and ODJ6 in order to validate the authenticity of the data (25). This was achieved using the program mapDamage v2.02 (26) with default parameters.

Batagai showed expected nucleotide mis-incorporation signatures of *post-mortem* DNA damage, with increasing C→T substitution rates towards 5' read termini, and increasing complementary G→A substitution levels towards 3' read ends (**Figure S2.3A**). In addition, we observed an excess of purines at the genomic position preceding sequencing starts, in line with depurination driving *post-mortem* DNA fragmentation (**Figure S2.4A**). Similar patterns were observed in the sequence data underlying the genome of samples CGG101397, Yukagir and ODJ6. For CGG101397 and ODJ6 C→T and G→A mis-incorporation rates were more limited, most likely due to the younger age of the specimens (**Figures S2.3-2.6**). For sample ODJ6, we observed abnormally elevated error rates for all classes of mis-incorporations at the 7th and the 17th positions from the 3' end (**Figure S2.5A**). As these libraries were sequenced as part of the same sequencing run, this indicates a bias in the base call or in the sequencing at cycles 7 and 17. This, however, does not impact the accuracy of the complete mitochondrial genome sequence for ODJ6, as we estimated that these errors involve only 0.03% of the bases mapped, these errors being randomly distributed along the mitochondrial genome sequence obtained at 30.6X coverage. We also noticed elevated mis-incorporation rates in the last ~7 nucleotides sequenced for four of the DNA libraries of sample Batagai (**Figure S2.3A**). These are likely due to palindromic artifacts, previously observed in TruSeq DNA library data (27), which introduce, during library formation, copies of the sequence starts towards sequence ends, as long as native templates show some level of sequence complementarity.

In order to limit the impact of such palindromic artifacts and other types of nucleotide mis-incorporations in the downstream analyses of the Batagai and CGG101397 complete genomes, we trimmed low-quality regions from both read termini as described in **Table S2.7**. We also used a statistical model of DNA damage to rescale the quality scores of likely damaged positions. More specifically, approximate Bayesian estimation of damage parameters was first computed in mapDamage v2.02 (26) with default parameters for each library of both ancient samples. Resulting posterior distribution estimates of cytosine deamination rates at both double stranded regions ($\delta_d$) and single stranded overhangs ($\delta_s$), and the probabilities of reads not terminating in overhangs ($\lambda$) are shown in **Figure S2.7**. We then rerun mapDamage v2.02 for rescaling quality scores of likely damaged positions of each library using the per-library options, as shown in (**Table S2.7**), in addition to the parameter --seq-length=15, which was used to fit the model based on the first 15 bases at both 5' and 3' ends for all libraries. All subsequent analyses involving the nuclear genome sequences of the ancient horses Batagai and CGG101397 rely on the sequences post-trimming and rescaling.

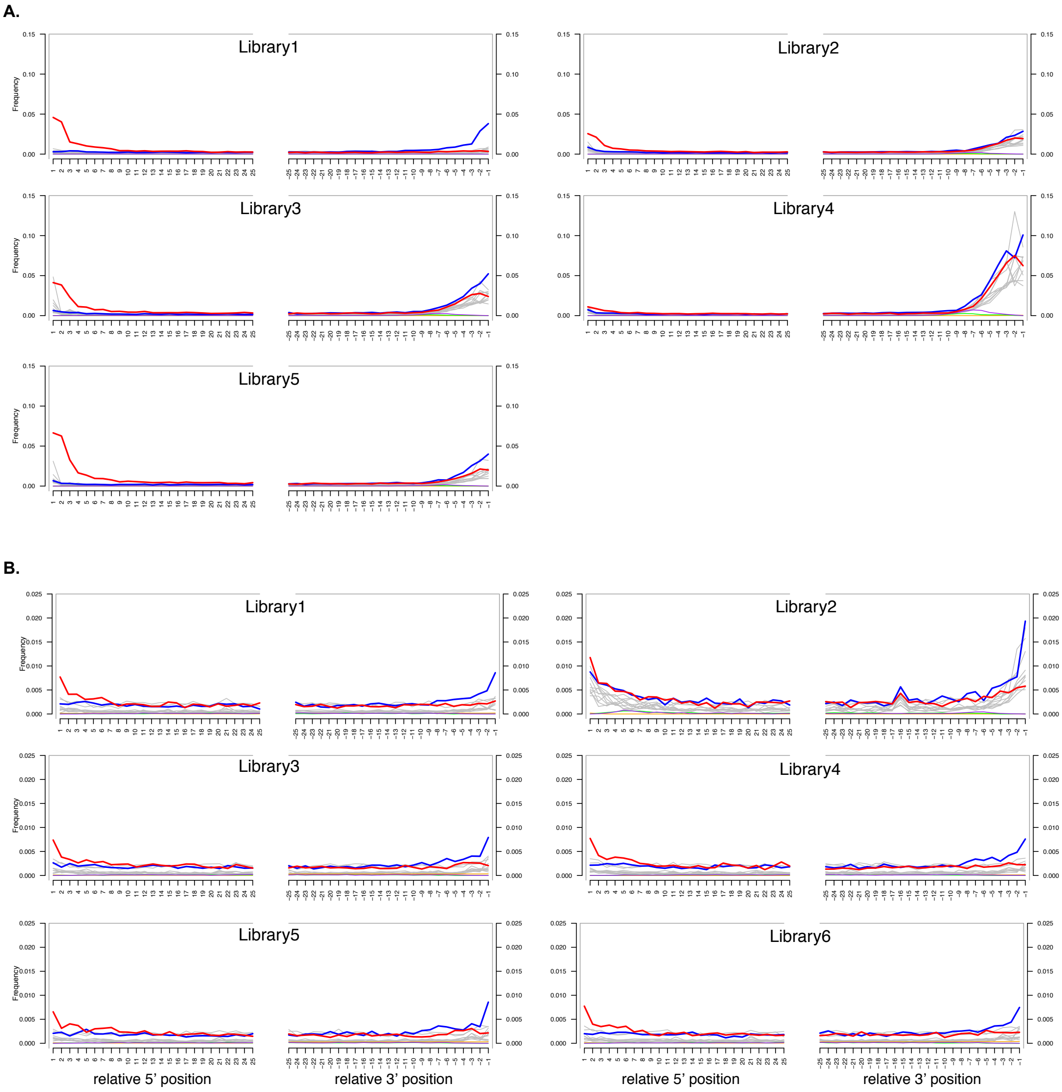
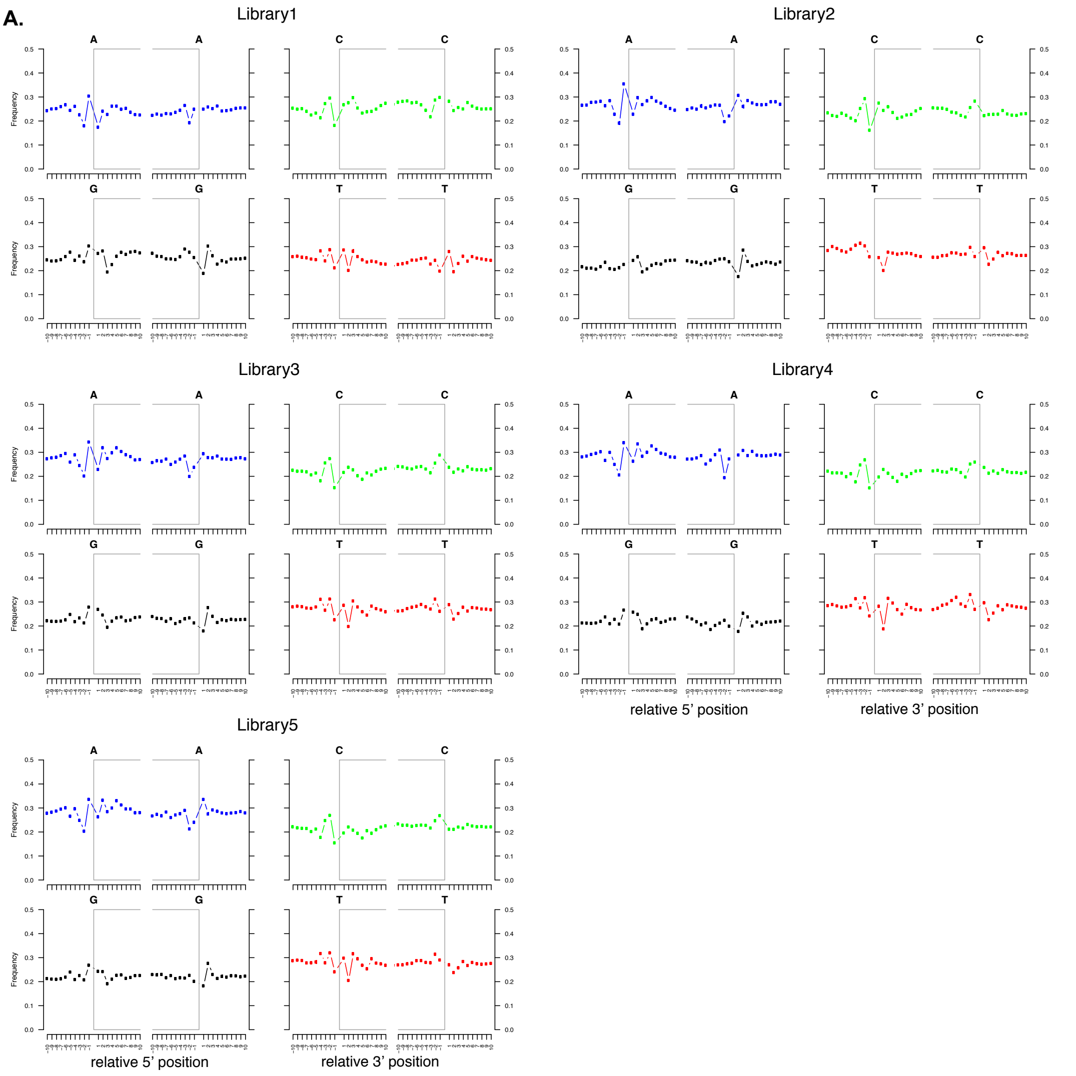


**Figure S2.3. Nucleotide mis-incorporation patterns at 5'- and 3'- read termini for the ancient samples Batagai and CGG101397.**

Nucleotide mis-incorporation patterns along the first and last 25 read positions obtained for the Batagai (A) and the CGG101397 (B) before trimming and rescaling. A. For Batagai, Library 1 is a blunt-ended library (New England Biolabs) amplified with AmpliTaq Gold DNA polymerase (Life Technologies), Library2, Library3 and Library5 are TruSeq libraries (Illumina) amplified with AmpliTaq Gold DNA polymerase, Library4 is a TruSeq library amplified with Accuprime Pfx (Life Technologies). B. For CGG101397, all libraries are blunt-ended libraries amplified with AmpliTaq Gold DNA polymerase. Mis-incorporation frequencies are shown for the first and the 25 nucleotides of the reads aligned to the horse reference nuclear genome EquCab2.0. The x-axis provides read positions relative to read starts (positive numbers) and/or read ends (negative numbers).


A.
Library1
Library2
Library3
Library4
Library5
A
C
G
T
Frequency
0.0
0.1
0.2
0.3
0.4
0.5
relative 5' position
relative 3' position

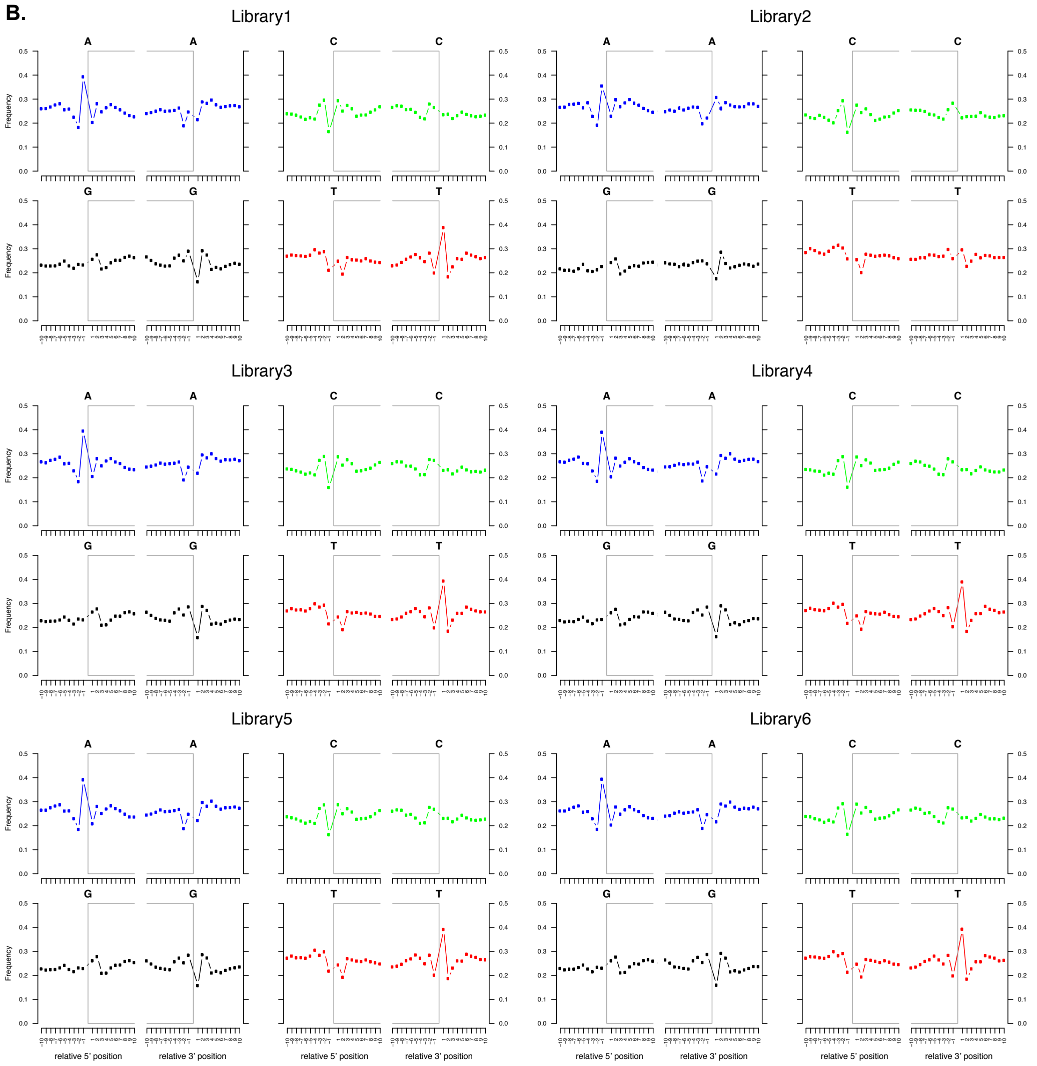


**Figure S2.4. DNA fragmentation patterns at 5'- and 3'- read termini for ancient the samples Batagai and CGG101397.**

DNA fragmentation patterns obtained for the Batagai (A) and the CGG101397 (B) specimens before trimming and rescaling. DNA fragmentation is shown for reads aligned to the horse reference nuclear genome EquCab2.0: within 10 bp preceding read starts (positions -1 to -10 on the left panels of each base composition profile) and within the 10 bp following read ends (post-adapter and/or quality trimming; positions 1 to 10 on the right panels of each base composition profile). See **Figure S2.3** for additional captions.

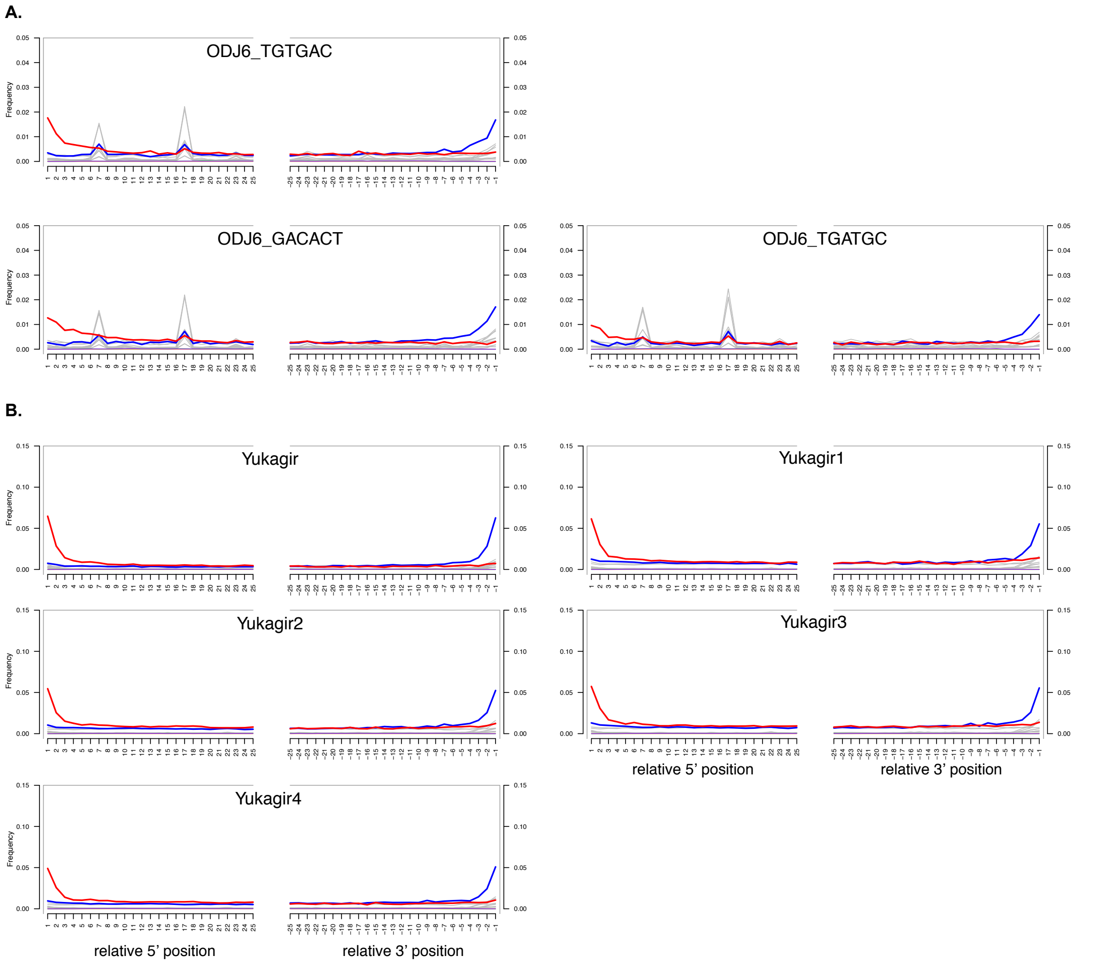
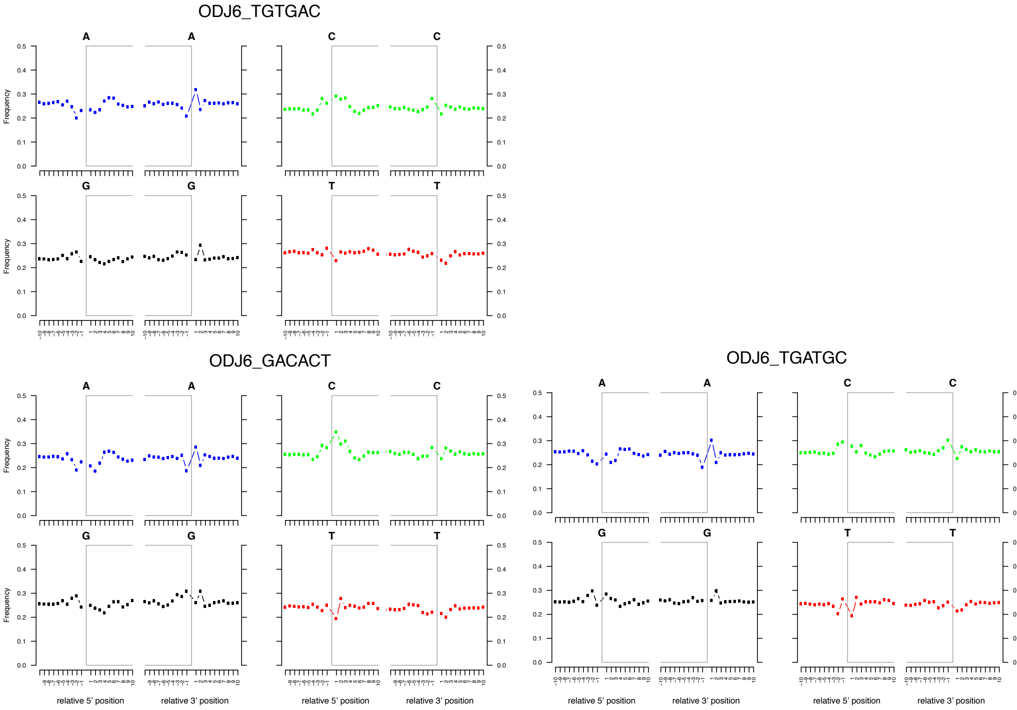


**Figure S2.5. Nucleotide mis-incorporation patterns at 5'- and 3'- read termini for the ancient samples ODJ6 and Yukagir.**
Nucleotide mis-incorporation patterns along the first and last 25 read positions obtained for the ODJ6 (A) and the Yukagir (B) specimens. All libraries are blunt-ended libraries (New England Biolabs) amplified with AmpliTaq Gold DNA polymerase (Life Technologies). See **Figure S2.3** for additional captions.

A.



ODJ6_TGTGAC

ODJ6_GACACT

ODJ6_TGATGC

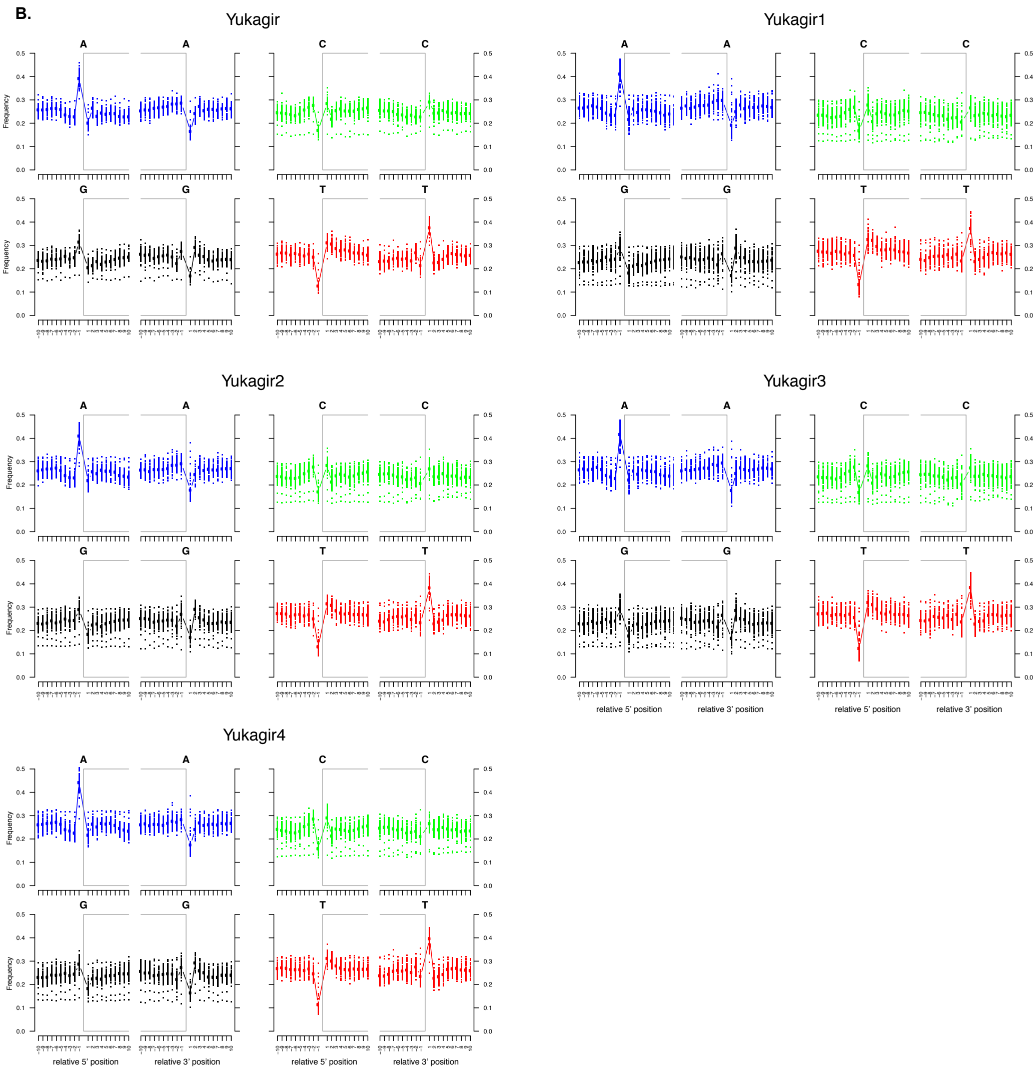


**Figure S2.6. DNA fragmentation patterns at 5'- and 3'- read termini for the ancient samples ODJ6 and Yukagir.**

DNA fragmentation patterns obtained for the ODJ6 (A) and the Yukagir (B) specimens. See **Figures S2.3 and S2.4** for additional captions.

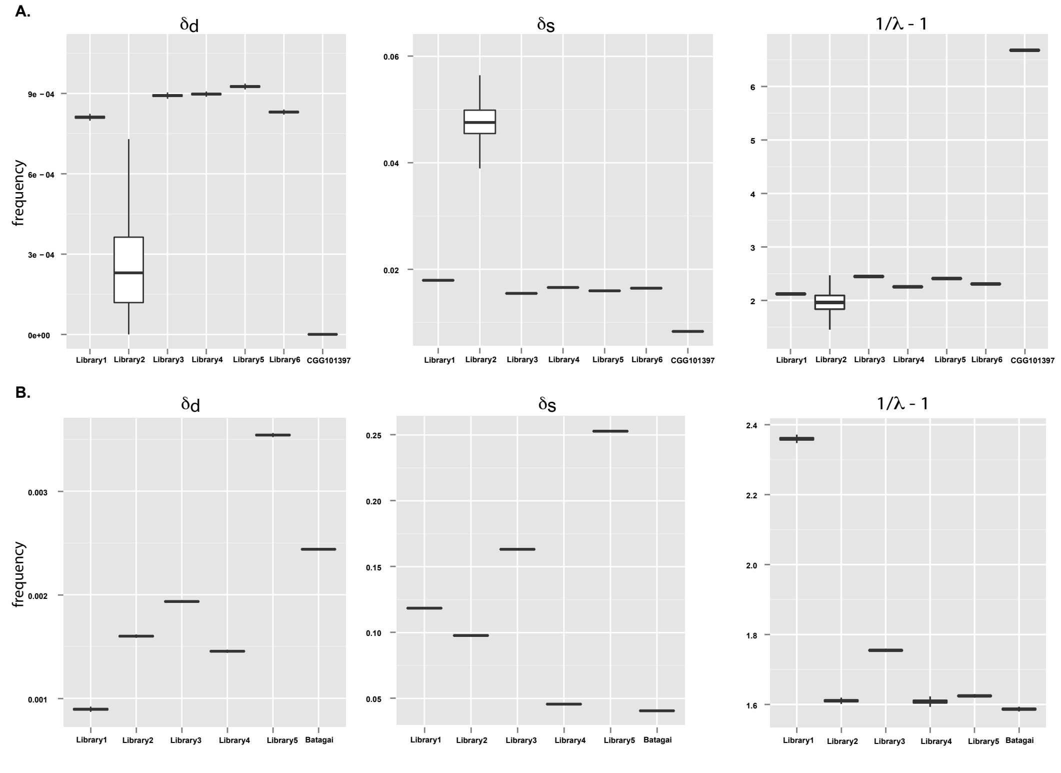


**Figure S2.7. Posterior distributions estimated for three DNA damage parameters ($\delta_d$, $\delta_s$ and 1/λ - 1) for ancient samples Batagai and CGG101397.**
"$\delta_d$", cytosine deamination rates at double stranded regions; "$\delta_s$", at single stranded overhangs ($\delta_S$); "λ", probability of reads not terminating in overhangs. Parameter distributions are shown on a per-library and a per-sample basis. A. CGG101397. B. Batagai. For clarity, λ is converted in a size estimate of average overhang length (in bp) using the following the formula: (1/λ - 1), following (13) and (16).

## *2.7 Sample-wise error rates*

The sequencing error rate was estimated using a methodology similar to the one described in (28), and implemented in the Supplementary section S4.4 of (13). The method is based on the idea that genomes from the same species should have the same expected number of derived alleles when compared to an outgroup sequence. We used the high-quality genome of EquCab2.0 (Thoroughbred horse Twilight; (29)) as the prototype horse genome for comparison, and the genome of *Equus africanus somaliensis* (30) as the outgroup defining ancestral alleles. More specifically, let $A_i$ and $a_i$ be the true and observed number of ancestral alleles for a given sample i. We define $D_i$ and $d_i$ as the true and observed number of derived alleles present in the same genome. Under the assumption that the genomes are phylogenetically equidistant from the outgroup, and given an error rate $\varepsilon_i$ the expected number of derived allele is:

$$E\,[d_i] = d_i\,(1 - \varepsilon_i) + A_i\varepsilon_i \approx d_i(1 - \varepsilon_i) + a_i\varepsilon_i$$

Under the assumption that the reference high-quality genome j contains no error such as that the estimates of $A_j$ and $D_j$ are equal to $a_j$ and dj, respectively, an estimate of the overall error rate can then be obtained as

$$\varepsilon_i = (d_i - d_j) / (a_j - d_j)$$

Here, the high-quality genome j is represented by the Illumina sequencing data generated by Orlando and colleagues (13) for the individual Twilight, which underpins the EquCab2.0 reference genome. Error rates for each substitution category are estimated on a maximum likelihood framework summing over the true state of the allele in the sequenced genome, as described in (13). We estimated the sequencing error rates for each sequenced genome with and without stringent quality filtering (minimal mapping quality, mapQ, of 30 and minimal base quality, baseQ, of 20). For both the high-quality horse and outgroup genomes, we considered Phred-mapping score mapQ>35 and Phred-base quality score baseQ>25 in order to match the expectations of high-quality data.

For unfiltered data, we observed very low average sequencing error rate for the genomes of modern Yakutian horses, spanning 0.14%-0.18% errors per base, for samples Yak5 and Yak2, repectively (**Figure S2.8)**. The per-base error rate of the genomes newly characterized in our comparative panel (18) was on average equal to 0.25% and ranged from 0.17% (sample Mon_FM0450) to 0.44% (sample Mon_FM1948). In absence of quality filtering, the genome of the sample CGG101397 showed an overall error rate similar to that of modern individuals (0.09%, **Figure S2.8**). The per-base error rate of sample Batagai was found to be larger, and equal to 0.23%, reminiscent of the quality estimated within our comparative genome panel (**Figure S2.8**). Importantly, error rates were particularly increased for G→A and C→T base substitutions within the most ancient sample Batagai (0.14% each). Those substitution types are known to correspond to nucleotide mis-incorporations at cytosine residues that have been deaminated into uracil residues post-mortem (25) (see also **section S2.6**).

We also calculated sequencing error rates after quality filtering (baseQ≥20 and mapQ≥30) and we estimated overall average error rate per base to range from 0.04% (Yak6) to 0.07% (Batagai). We found remarkably low sequencing error rates (<0.01% to 0.09%) across all substitution types (**Figure S2.9**). The highest error rates were again observed for G→A and C→T base substitutions for the ancient genome of sample Batagai (0.05% each) and the modern genomes of the Yak2 and Yak3 horses (0.09% each for both), most likely due to significant cytosine deamination levels in the hair tissues analysed for these samples (**Figure S2.9**).

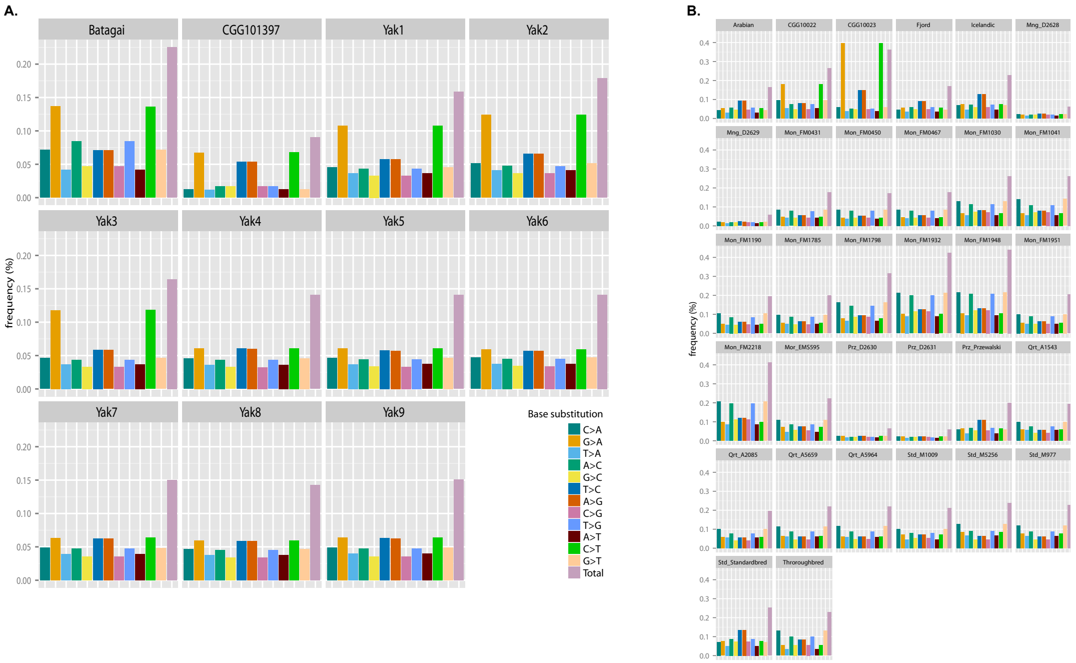


**Figure S2.8. Estimated error rates for ancient and present-day horse genomes without quality filtering.**

A. Ancient and present-day horse genomes characterized in this study. B. Ancient and present-day horse genomes of the comparative panel. Error rates are provided for each substitution type and across all substitution types (“Total”). “High-quality genome” = EquCab2.0; outgroup genome = *Equus asinus somaliensis*, both filtered to keep bases with mapping scores >35 and base-quality scores >25.

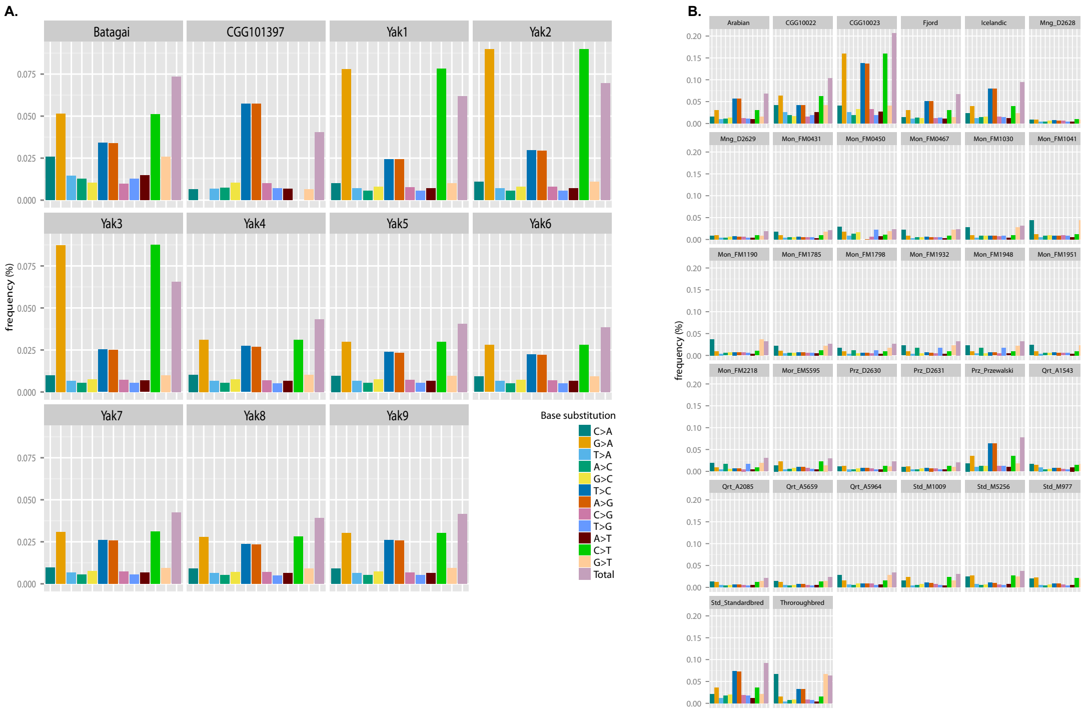


**Figure S2.9. Estimated error rates for ancient and present-day horse genomes with quality filtering (mapQ≥30, baseQ≥20).**

A. Ancient and present-day horse genomes characterized in this study. B. Ancient and present-day horse genomes of the comparative panel. See **Figure S2.8** for additional captions.

## 2.8 *Supplementary Tables for Section 2*

### Table S2.1. Ancient DNA extraction and library information.

| Horse ID | Tissue type | Powder used (mg) | Library | Library type | Polymerase | PCR cycles |
|---|---|---|---|---|---|---|
| Yukagir | Bone | 348 | Yukagir | BE | TG | 12+10 |
| | | | Yukagir1 | BE | TG | 12+10 |
| | | | Yukagir2 | BE | TG | 12+10 |
| | | | Yukagir3 | BE | TG | 12+10 |
| | | | Yukagir4 | BE | TG | 12+10 |
| Batagai | Bone | 225 | Library1 | BE | TG | 12+10 |
| | | | Library2 | TS | TG | 15 |
| | | | Library3 | TS | TG | 15 |
| | | | Library4 | TS | AP | 15 |
| | | | Library5 | TS | TG | 15 |
| CGG101392 | Tooth | 605 | CGG101392 | BE | TG | 12+10 |
| | | | CGG101392R | BE | TG | 12+10 |
| CGG101393 | Bone | 673 | CGG101393AR | BE | TG | 12+10 |
| | | 263 | CGG101393B | BE | TG | 12+10 |
| | | | CGG101393BR | BE | TG | 12+10 |
| CGG101394 | Tooth | 482 | CGG101394 | BE | TG | 12+10 |
| | | | CGG101394R | BE | TG | 12+10 |
| CGG101395 | Bone | 326 | CGG101395 | BE | TG | 12+10 |
| | | | B_CGG101395R | BE | TG | 12+10 |
| CGG101396 | Tooth | 361 | CGG101396 | BE | TG | 12+10 |
| | | | CGG101396R | BE | TG | 12+10 |
| CGG101397 | Bone | 727 | Library1 (TuRE$) | BE | TG | 12+10 |
| | | | Library2 (Tu$) | BE | TG | 12+10 |
| | | | Library3 | BE | TG | 12+10 |
| | | | Library4 | BE | TG | 12+10 |
| | | | Library5 | BE | TG | 12+10 |
| | | | Library6 | BE | TG | 12+10 |
| ODJ6 | Bone | 241 | TGTGAC | BE | TG | 10+10 |
| | | | TGATGC | BE | TG | 10+10 |
| | | | GACACT | BE | TG | 10+10 |

“BE”: Blunt-Ended library (New England Biolabs); “TS”: TruSeq library (Illumina); “TG”: AmpliTaq Gold DNA polymerase (Life Technologies); “AG”: Accuprime Pfx (Life Technologies). When two amplification rounds were performed, the number of PCR cycles used in the first and second rounds are indicated prior and following the + sign, respectively. $: labels from (10).

## Table S2.2. Sequencing information for the present-day Yakutian horses.

| Horse ID | Library | Sequencing Run | Number of raw reads | Number of retained reads | Number of collapsed pairs |
|---|---|---|---|---|---|
| Yak1 | Yak1_1 | 98PE | 515,474,672 | 286,081,689 | 225,881,271 |
| | Yak1_2 | 98PE | 63,444,452 | 41,497,687 | 21,671,067 |
| | Total | 98PE | 578,919,124 | 327,579,376 | 247,552,338 |
| Yak2 | Yak2_1 | 98PE | 891,452,624 | 541,301,679 | 343,163,472 |
| | Yak2_2 | 98PE | 49,195,198 | 31,384,438 | 17,622,213 |
| | Total | 98PE | 940,647,822 | 572,686,117 | 360,785,685 |
| Yak3 | Yak3_1 | 98PE | 516,321,776 | 299,198,159 | 213,417,404 |
| | Yak3_2 | 98PE | 56,254,278 | 35,359,754 | 20,654,528 |
| | Total | 98PE | 572,576,054 | 334,557,913 | 234,071,932 |
| Yak4 | Yak4_1 | 98PE | 430,886,056 | 249,367,905 | 178,365,934 |
| | Yak4_2 | 98PE | 58,496,056 | 37,195,922 | 21,057,224 |
| | Total | 98PE | 489,382,112 | 286,563,827 | 199,423,158 |
| Yak5 | Yak5_1 | 98PE | 530,254,434 | 299,574,265 | 226,771,443 |
| | Yak5_2 | 98PE | 49,402,642 | 31,394,424 | 17,788,042 |
| | Total | 98PE | 579,657,076 | 330,968,689 | 244,559,485 |
| Yak6 | Yak6_1 | 98PE | 587,021,086 | 331,444,133 | 251,436,596 |
| | Yak6_2 | 98PE | 43,036,440 | 26,425,349 | 16,437,552 |
| | Total | 98PE | 630,057,526 | 357,869,482 | 267,874,148 |
| Yak7 | Yak7_1 | 98PE | 1,153,069,686 | 676,648,042 | 467,572,888 |
| | Yak7_2 | 98PE | 22,598,580 | 15,139,546 | 7,352,490 |
| | Total | 98PE | 1,175,668,266 | 691,787,588 | 474,925,378 |
| Yak8 | Yak8_1 | 98PE | 440,059,126 | 258,885,881 | 178,074,119 |
| | Yak8_2 | 98PE | 38,375,780 | 26,238,927 | 11,975,400 |
| | Total | 98PE | 478,434,906 | 285,124,808 | 190,049,519 |
| Yak9 | Yak9_1 | 98PE | 458,586,934 | 279,105,845 | 175,926,495 |
| | Yak9_2 | 98PE | 30,194,094 | 20,356,673 | 9,705,442 |
| | Total | 98PE | 488,781,028 | 299,462,518 | 185,631,937 |
| Horse1 | Library1 | 50SE/98SE | 33,495,082 | 14,752,467 | n/a |
| | Library2 | 50SE/98SE | 19,533,897 | 7,518,876 | n/a |
| | Library3 | 50SE/98SE | 70,896,686 | 66,503,514 | n/a |
| | Total | 50SE | 123,925,665 | 88,774,857 | n/a |
| Horse2 | Library4 | 50SE | 21,980,216 | 20,782,070 | n/a |
| Horse3 | Library4 | 98SE | 9,676,918 | 9,379,056 | n/a |

“Number of retained reads”: number of sequencing reads retained after adapter removal; “Number of collapsed pairs”: number of overlapping read pairs collapsed into a single sequence; “98PE”: 98 bp pair-ended sequencing reads (PE); “50SE/98SE”: 50/98 bp single-ended reads.

**Table S2.3. Sequencing information for the ancient horses from Yakutia.**

| Horse ID | Library | Sequencing Run | Number of PE reads | Number of SE reads | Number of retained reads (PE and SE) | Number of collapsed pairs |
|---|---|---|---|---|---|---|
| Yukagir | Yukagir | 94SE | n/a | 48,101,535 | 47,813,678 | n/a |
| | Yukagir1 | 94SE | n/a | 4,866,139 | 4,711,842 | n/a |
| | Yukagir2 | 94SE | n/a | 7,560,056 | 7,255,501 | n/a |
| | Yukagir3 | 94SE | n/a | 4,394,290 | 4,319,878 | n/a |
| | Yukagir4 | 94SE | n/a | 6,237,908 | 6,113,260 | n/a |
| | Total | | n/a | 71,159,928 | 70,214,159 | n/a |
| Batagai | Library1 | 94SE | n/a | 60,936,512 | 60,624,116 | n/a |
| | Library2 | 94SE | n/a | 262,457,296 | 218,190,094 | n/a |
| | Library3 | 51SE, 94SE, 98PE | 615,757,748 | 300,147,502 | 562,609,645 | 216,372,041 |
| | Library4 | 51SE, 94SE, 98PE | 874,262,382 | 31,595,163 | 397,541,599 | 254,737,837 |
| | Library5 | 94SE, 98PE | 44,142,598 | 278,586,185 | 232,218,000 | 15,226,633 |
| | Total | | 1,534,162,728 | 933,722,658 | 1,471,183,454 | 486,336,511 |
| CGG101392 | CGG101392 | 94SE | n/a | 12,545,745 | 12,393,476 | n/a |
| | CGG101392R | 94SE | n/a | 16,379,222 | 10,871,116 | n/a |
| | Total | | n/a | 28,924,967 | 23,264,592 | n/a |
| CGG101393 | CGG101393AR | 94SE | n/a | 18,302,900 | 11,862,758 | n/a |
| | CGG101393B | 94SE | n/a | 10,401,702 | 9,491,647 | n/a |
| | CGG101393BR | 94SE | n/a | 9,849,670 | 8,969,882 | n/a |
| | Total | | n/a | 38,554,272 | 30,324,287 | n/a |
| CGG101394 | CGG101394 | 94SE | n/a | 8,645,215 | 8,531,157 | n/a |
| | CGG101394R | 94SE | n/a | 7,267,126 | 7,033,095 | n/a |
| | Total | | n/a | 15,912,341 | 15,564,252 | n/a |
| CGG101395 | CGG101395 | 94SE | n/a | 11,693,072 | 11,259,928 | n/a |
| | B_CGG101395R | 94SE | n/a | 12,187,372 | 12,023,328 | n/a |
| | Total | | n/a | 23,880,444 | 23,283,256 | n/a |
| CGG101396 | CGG101396 | 94SE | n/a | 12,009,410 | 11,709,424 | n/a |
| | CGG101396R | 94SE | n/a | 29,418,042 | 12,384,464 | n/a |
| | Total | | n/a | 41,427,452 | 24,093,888 | n/a |
| CGG101397 | Library1 | 51SE, 98PE | 322,207,922 | 1,998,389 | 162,988,335 | 158,261,243 |
| | Library2 | 51SE | n/a | 642,694 | 565,437 | n/a |
| | Library3 | 51SE, 98PE | 614,836,976 | 2,030,628 | 309,052,534 | 304,207,048 |
| | Library4 | 51SE, 98PE | 806,664,104 | 1,629,180 | 404,510,458 | 399,065,920 |
| | Library5 | 51SE, 98PE | 761,335,186 | 1,729,611 | 382,214,058 | 376,946,693 |
| | Library6 | 51SE, 98PE | 875,205,176 | 1,157,255 | 436,691,907 | 430,865,890 |
| | Total | | 3,380,249,364 | 9,187,757 | 1,696,022,729 | 1,669,346,794 |
| ODJ6 | TGTGAC | 94SE | n/a | 17,507,548 | 17,669,585 | n/a |
| | TGATGC | 94SE | n/a | 20,127,349 | 20,216,200 | n/a |
| | GACACT | 94SE | n/a | 92,964,045 | 93,269,115 | n/a |
| | Total | | n/a | 130,598,942 | 131,154,900 | n/a |

“Number of retained reads”: number of sequencing reads retained after adapter removal; “Number of collapsed pairs”: number of overlapping read pairs collapsed into a single sequence; “PE”: pair-ended; “SE”: single-ended.

**Table S2.4. Genome characteristics for the ancient and present-day horses of the comparative panel.**

| Horse ID | Breed | Gender | Reference | # Mapped | Cov. (X) |
|---|---|---|---|---|---|
| Mon_FM1798 | Franches-Montagnes | M | Der Sarkissian et al. 2015 (18) | 42,505,483 | 22.62 |
| Mon_FM1932 | Franches-Montagnes | M | Der Sarkissian et al. 2015 (18) | 20,527,420 | 10.82 |
| Mon_FM1948 | Franches-Montagnes | M | Der Sarkissian et al. 2015 (18) | 16,600,211 | 8.72 |
| Mon_FM2218 | Franches-Montagnes | M | Der Sarkissian et al. 2015 (18) | 22,771,087 | 12.05 |
| Mon_FM1190 | Franches-Montagnes | M | Der Sarkissian et al. 2015 (18) | 30,749,884 | 16.37 |
| Mon_FM1041 | Franches-Montagnes | M | Der Sarkissian et al. 2015 (18) | 14,695,522 | 7.96 |
| Mon_FM1951 | Franches-Montagnes | M | Der Sarkissian et al. 2015 (18) | 36,627,376 | 19.24 |
| Mon_FM0467 | Franches-Montagnes | M | Der Sarkissian et al. 2015 (18) | 28,836,871 | 15.29 |
| Mon_FM_431 | Franches-Montagnes | M | Der Sarkissian et al. 2015 (18) | 21,988,334 | 11.62 |
| Mon_FM1785 | Franches-Montagnes | M | Der Sarkissian et al. 2015 (18) | 28,230,241 | 14.89 |
| Mon_FM1030 | Franches-Montagnes | M | Der Sarkissian et al. 2015 (18) | 32,187,975 | 17.02 |
| Mon_FM0450 | Franches-Montagnes | M | Der Sarkissian et al. 2015 (18) | 29,840,675 | 16.15 |
| Mor_EMS595 | Morgan | F | Der Sarkissian et al. 2015 (18) | 25,216,660 | 14.02 |
| Qrt_A5964 | American Quarter Horse | F | Der Sarkissian et al. 2015 (18) | 25,565,430 | 13.97 |
| Qrt_A5659 | American Quarter Horse | F | Der Sarkissian et al. 2015 (18) | 25,576,377 | 13.67 |
| Qrt_A1543 | American Quarter Horse | F | Der Sarkissian et al. 2015 (18) | 22,265,239 | 12.55 |
| Qrt_A2085 | American Quarter Horse | M | Der Sarkissian et al. 2015 (18) | 22,456,409 | 12.32 |
| Std_M977 | Standardbred | F | Der Sarkissian et al. 2015 (18) | 22,685,006 | 12.35 |
| Std_M5256 | Standardbred | M | Der Sarkissian et al. 2015 (18) | 20,975,678 | 11.44 |
| Std_M1009 | Standardbred | M | Der Sarkissian et al. 2015 (18) | 13,252,954 | 7.68 |
| Std_Standardbred | Standardbred | M | Orlando et al. 2013; Schubert et al. 2014 (13, 16) | 27,145,773 | 12.31 |
| Mng_D2629 | Mongolian | F | Do et al. 2014 (19) | 44,899,823 | 24.09 |
| Mng_D2628 | Mongolian | M | Do et al. 2014 (19) | 43,833,220 | 23.54 |
| Fjord | Norwegian Fjord | F | Orlando et al. 2013; Schubert et al. 2014 (13, 16) | 17,734,587 | 7.73 |
| Icelandic | Icelandic | M | Orlando et al. 2013; Schubert et al. 2014 (13, 16) | 16,814,503 | 8.51 |
| Arabian | Arabian | F | Orlando et al. 2013; Schubert et al. 2014 (13, 16) | 24,648,073 | 10.82 |
| Thoroughbred | Thoroughbred | F | Wade et al. 2009 (29) | 74,144,230 | 41.14 |
| Prz_D2630 | Przewalski | M | Do et al. 2014 (19) | 31,465,668 | 16.91 |
| Prz_D2631 | Przewalski | F | Do et al. 2014 (19) | 47,061,560 | 25.25 |
| Prz_Przewalski | Przewalski | M | Orlando et al. 2013; Schubert et al. 2014 (13, 16) | 20,310,866 | 9.65 |
| CGG10022 | Late Pleistocene | F | Orlando et al. 2013; Schubert et al. 2014 (13, 16) | 53,947,568 | 25.60 |
| CGG10023 | Late Pleistocene | M | Schubert et al. 2014 (16) | 18,032,105 | 7.72 |

"F": Female; "M": Male; "#Mapped": number of sequencing reads mapping uniquely and with high confidence (mapping quality mapQ≥ 25). "Cov.": genome coverage.

**Table S2.5. Mapping results for present-day Yakutian horses.**

| | | | | Mapped reads[a] | | Coverage (X) | |
|---|---|---|---|---|---|---|---|
| **Horse ID** | **Library** | **%end.** | **%clon.** | **Nuclear genome** | **Mitochondrial genome** | **Nuclear genome** | **Mitochondrial genome** |
| Yak1 | Yak1_1 | 75.22 | 2.82 | 215,185,844 | 168,417 | 9.19 | 1,044.85 |
| | Yak1_2 | 75.09 | 7.59 | 31,158,714 | 21,245 | 1.32 | 132.83 |
| | **Total** | **75.20** | **3.45** | **246,344,558** | **189,662** | **10.52** | **1,177.67** |
| Yak2 | Yak2_1 | 74.45 | 6.47 | 403,013,210 | 231,724 | 17.37 | 1,470.14 |
| | Yak2_2 | 75.19 | 7.50 | 23,598,996 | 13,693 | 1.01 | 85.76 |
| | **Total** | **74.49** | **6.53** | **426,612,206** | **245,417** | **18.38** | **1,555.91** |
| Yak3 | Yak3_1 | 74.79 | 3.22 | 223,770,293 | 155,116 | 9.54 | 965.51 |
| | Yak3_2 | 73.74 | 7.24 | 26,074,838 | 17,190 | 1.11 | 106.46 |
| | **Total** | **74.68** | **3.65** | **249,845,131** | **172,306** | **10.65** | **1,071.97** |
| Yak4 | Yak4_1 | 74.27 | 3.47 | 185,197,680 | 154,190 | 7.91 | 976.88 |
| | Yak4_2 | 73.26 | 7.76 | 27,249,122 | 22,023 | 1.16 | 139.33 |
| | **Total** | **74.14** | **4.04** | **212,446,802** | **176,213** | **9.07** | **1,116.20** |
| Yak5 | Yak5_1 | 75.18 | 2.82 | 225,218,752 | 217,060 | 9.53 | 1,372.03 |
| | Yak5_2 | 74.60 | 7.25 | 23,421,400 | 21,955 | 0.99 | 138.18 |
| | **Total** | **75.12** | **3.26** | **248,640,152** | **239,015** | **10.52** | **1,510.21** |
| Yak6 | Yak6_1 | 76.38 | 3.02 | 253,152,852 | 220,636 | 10.79 | 1,404.27 |
| | Yak6_2 | 75.24 | 7.01 | 19,881,239 | 17,349 | 0.85 | 110.31 |
| | **Total** | **76.29** | **3.32** | **273,034,091** | **237,985** | **11.64** | **1,514.58** |
| Yak7 | Yak7_1 | 73.59 | 5.11 | 497,929,694 | 496,523 | 21.16 | 3,158.80 |
| | Yak7_2 | 74.53 | 8.46 | 11,282,826 | 10,973 | 0.47 | 68.58 |
| | **Total** | **73.61** | **5.19** | **509,212,520** | **507,496** | **21.63** | **3,227.38** |
| Yak8 | Yak8_1 | 77.41 | 2.48 | 200,394,370 | 172,475 | 8.64 | 1,106.54 |
| | Yak8_2 | 75.99 | 8.46 | 19,938,048 | 15,680 | 0.84 | 98.74 |
| | **Total** | **77.28** | **3.05** | **220,332,418** | **188,155** | **9.48** | **1,205.28** |
| Yak9 | Yak9_1 | 77.06 | 3.17 | 215,068,521 | 196,393 | 9.21 | 1,254.99 |
| | Yak9_2 | 75.45 | 8.20 | 15,359,392 | 13,557 | 0.65 | 85.12 |
| | **Total** | **76.95** | **3.52** | **230,427,913** | **209,950** | **9.85** | **1,340.11** |
| Horse1 | Library1 | 35.40 | 2.70 | 5,222,589 | 6,919 | 0.11 | 25.69 |
| | Library2 | 46.19 | 1.36 | 3,472,669 | 3,815 | 0.09 | 16.05 |
| | Library3 | 10.83 | 79.28 | 7,204,071 | 9,762 | 0.13 | 27.89 |
| | **Total** | **17.91** | **63.58** | **15,899,329** | **20,496** | **0.34** | **69.63** |
| Horse2 | **Library4** | **50.53** | **3.43** | **10,500,497** | **16,170** | **0.21** | **47.66** |
| Horse3 | **Library4** | **38.87** | **9.13** | **3,645,791** | **8002** | **0.13** | **43.94** |

[a] sequencing reads mapping uniquely and with high confidence (mapping quality MQ≥ 25); "%end.": percentage of endogenous DNA calculated as 100 * Number of Mapped Reads (uniquely and with high confidence) / Number of Retained Reads (after trimming for adapters and low-quality ends); "%clon.": percentage of reads that were PCR duplicates; %end. and %clon. were calculated considering reads mapping to the nuclear or the mitochondrial horse genomes separately.

## Table S2.6. Mapping results for the ancient horses from Yakutia.

| | | | | Mapped reads[a] | Coverage (X) | | | |
|---|---|---|---|---|---|---|---|---|
| **Horse ID** | **Library** | **%end.** | **%clon.** | **Nuclear genome** | **Mitochondrial genome** | **Nuclear genome[a]** | **X-chromosome** | **Mitochondrial genome** |
| Yukagir | Yukagir | 0.72 | 89.21 | 344,200 | 32,871 | 9.08x10$^{-3}$ | - | 184.28 |
| | Yukagir1 | 1.34 | 81.33 | 63,295 | 32,427 | 1.87x10$^{-3}$ | - | 181.30 |
| | Yukagir2 | 1.26 | 84.03 | 91,508 | 32,626 | 2.61x10$^{-3}$ | - | 182.73 |
| | Yukagir3 | 1.48 | 80.33 | 63,745 | 32,424 | 1.86x10$^{-3}$ | - | 181.04 |
| | Yukagir4 | 1.51 | 82.95 | 92,345 | 32,490 | 2.68x10$^{-3}$ | - | 182.11 |
| | **Total** | **0.93** | **86.81** | **655,093** | **162,838** | **1.81x10$^{-2}$** | - | **911.47** |
| Batagai [b] | Library1 | 31.34 | 7.77 | 19,001,067 | 2,591 | 0.54 | - | 11.69 |
| | Library2 | 50.16 | 3.93 | 109,449,139 | 13,985 | 3.38 | - | 72.66 |
| | Library3 | 35.47 | 3.20 | 199,575,502 | 33,838 | 6.69 | - | 187.28 |
| | Library4 | 30.48 | 4.27 | 121,150,831 | 21,837 | 4.28 | - | 119.13 |
| | Library5 | 48.70 | 6.32 | 113,095,345 | 15,991 | 3.4 | - | 81.66 |
| | **Total** | **38.22** | **4.37** | **562,271,884** | **88,242** | **18.29** | **8.87** | **472.42** |
| CGG101392 | CGG101392 | 0.19 | 1.24 | 23,783 | 45 | 5.89x10$^{-4}$ | - | 0.16 |
| | CGG101392R | 0.36 | 22.95 | 39,409 | 81 | 1.00x10$^{-3}$ | - | 0.34 |
| | **Total** | **0.27** | **16.00** | **63,192** | **126** | **1.59x10$^{-3}$** | - | **0.50** |
| CGG101393 | CGG101393AR | 2.42 | 4.46 | 286,689 | 279 | 8.77x10$^{-3}$ | - | 1.31 |
| | CGG101393B | 0.56 | 6.87 | 52,735 | 58 | 1.32x10$^{-3}$ | - | 0.19 |
| | CGG101393BR | 3.21 | 12.02 | 288,377 | 167 | 6.24x10$^{-3}$ | - | 0.57 |
| | **Total** | **2.07** | **8.28** | **627,801** | **504** | **1.63x10$^{-2}$** | - | **2.06** |
| CGG101394 | CGG101394 | 1.69 | 1.25 | 143,836 | 37 | 3.95x10$^{-3}$ | - | 0.13 |
| | CGG101394R | 11.24 | 1.96 | 790,281 | 749 | 2.14x10$^{-2}$ | - | 2.99 |
| | **Total** | **6.00** | **1.85** | **934,117** | **786** | **2.54x10$^{-2}$** | - | **3.12** |
| CGG101395 | CGG101395 | 0.82 | 6.93 | 92,883 | 24 | 2.52x10$^{-3}$ | - | 0.08 |
| | B_CGG101395R | 2.66 | 4.21 | 319,312 | 170 | 7.74x10$^{-3}$ | - | 0.67 |
| | **Total** | **1.77** | **4.83** | **412,195** | **194** | **1.03x10$^{-2}$** | - | **0.75** |
| CGG101396 | CGG101396 | 0.42 | 1.91 | 48,828 | 15 | 1.40x10$^{-3}$ | - | 0.06 |
| | CGG101396R | 1.68 | 48.73 | 207,982 | 200 | 5.90x10$^{-3}$ | - | 0.84 |
| | **Total** | **1.07** | **43.61** | **256,810** | **215** | **7.30x10$^{-3}$** | - | **0.90** |
| CGG101397[b] | Library1 | 50.35 | 6.53 | 82,062,890 | 23,82 | 2.15 | - | 88.50 |
| | Library2 | 9.95 | 0.49 | 56,265 | 15 | 9.5x10$^{-4}$ | - | 0.04 |
| | Library3 | 46.33 | 14.89 | 143,194,831 | 41,660 | 3.72 | - | 154.62 |
| | Library4 | 44.17 | 18.29 | 178,676,610 | 50,828 | 4.65 | - | 189.27 |
| | Library5 | 45.60 | 17.21 | 174,275,030 | 49,215 | 4.7 | - | 188.32 |
| | Library6 | 43.86 | 18.25 | 191,542,625 | 52,707 | 5.03 | - | 198.89 |
| | **Total** | **45.39** | **16.29** | **769,808,251** | **218,247** | **20.25** | **8.93** | **819.64** |
| ODJ6 | TGTGAC | 7.79 | 12.11 | 1,363,036 | 536 | 4.26 x10$^{-2}$ | - | 2.66 |
| | TGATGC | 3.51 | 4.68 | 705,771 | 603 | 2.18 x10$^{-2}$ | - | 2.96 |
| | GACACT | 10.34 | 9.40 | 9,610,375 | 5,157 | 0.30 | - | 24.98 |
| | **Total** | **8.94** | **9.46** | **11,679,182** | **6,296** | **0.37** | - | **30.60** |

[a] sequencing reads mapping uniquely and with high confidence (mapping quality MQ≥ 25); [b] statistics relative to the mapping to the nuclear genomes are given post-trimming/rescaling (see **section S2.6**); "%end.": percentage of endogenous DNA calculated as 100 * Number of Mapped Reads (uniquely and with high confidence) / Number of Retained Reads (after trimming for adapters and low-quality ends); "%clon.": percentage of reads that were PCR duplicates; %end. and %clon. were calculated considering reads mapping to the nuclear or the mitochondrial horse genomes separately.

**Table S2.7. Number of bases trimmed at read ends and mapDamage v2.02 parameters used for rescaling bases in the BAM file alignments to the horse nuclear genome sequence for the Batagai, the CGG101397 and two previously published Late Pleistocene genomes.**

| | | | #bases trimmed | |
|---|---|---|---|---|
| **Horse ID** | **Library name** | **mapDamage2 rescaling option** | **5'-end** | **3'-end** |
| Batagai | Library1 | No rescaling | 0 | 0 |
| | Library2 | --diff-hangs | 0 | 0 |
| | Library3 | --forward | 2 | 2 |
| | Library4 | --forward | 6 | 6 |
| | Library5 | --forward | 2 | 2 |
| CGG101397 | Library1 | --forward | 7 | 7 |
| | Library2 | No rescaling | 3 | 3 |
| | Library3 | No rescaling | 9 | 9 |
| | Library4 | No rescaling | 2 | 2 |
| | Library5 | No rescaling | 8 | 8 |
| | Library6 | No rescaling | 1 | 1 |
| CGG10022[a] | ACTGCC_DI | --diff-hangs | 0 | 0 |
| | ACTTGA | --diff-hangs | 0 | 0 |
| | CGTAGT | --diff-hangs | 0 | 0 |
| | CTTGTA_AP | No rescaling | 0 | 0 |
| | CTTGTA_APem | No rescaling | 0 | 0 |
| | CTTGTA_TG | --diff-hangs | 0 | 0 |
| | GCAACG_TI | Defaults | 0 | 0 |
| | TGACCA_AP | No rescaling | 0 | 0 |
| | TGACCA_TG | No rescaling | 0 | 0 |
| | TGCAGG_SI | Defaults | 0 | 0 |
| CGG10023 [a] | 13-idx1 | --diff-hangs | 0 | 0 |
| | 13-idx10 | --diff-hangs | 0 | 0 |
| | 13-idx11 | --diff-hangs | 0 | 0 |
| | 13-idx12 | --diff-hangs | 0 | 0 |
| | 13-idx2 | --diff-hangs | 0 | 0 |
| | 13-idx24 | Defaults | 0 | 0 |
| | 13-idx25 | Defaults | 0 | 0 |
| | 13-idx26 | Defaults | 0 | 0 |
| | 13-idx3 | --diff-hangs | 0 | 0 |
| | 13-idx4 | --diff-hangs | 0 | 0 |
| | 13-idx9 | --diff-hangs | 0 | 0 |

[a] Schubert et al. 2014 (16); “--diff-hangs”: damage models are allowed to be different for 5’- and 3’-overhangs. “--forward”: damage models are estimated using only the 5'-end of the sequencing reads.

# 3 Section 3: Microbial profiling of ancient horses DNA

We performed metagenomic analyses on single end DNA sequences of each library of the four ancient specimens for which we generated shotgun-sequencing data (samples Yukagir, Batagai, CGG101397, and ODJ6). We followed the methodology implemented in the PALEOMIX pipeline (20), and applied in (10). Briefly, single-end sequencing reads were mapped to the markers of the MetaPhlAn database using the Bowtie2 v2.1.0 (31), with default parameters (sensitive global alignment strategy). PCR duplicates were then removed from each DNA library. The resultant unique high-quality hits were then profiled for their microbial content using MetaPhlAn version 1.7.7 (32). For comparison, we also generated microbial profiles on a subset of shotgun sequences of the two previously published Late Pleistocene horses (CGG10022 and CGG10023; (16)) and we also incorporated microbial profiles obtained from other six ancient horse remains (18th-19th century AD) analysed in (10) **(Table S3.1)**. All the ancient samples analysed were buried in similar conditions, namely in the Yakutian permafrost. For sample CGG101397, the profile obtained from Library2 was excluded from the analyses, as no microbial taxon could be identified by MetaPhlAn from the limited number of reads generated for this library (and mapping to the MetaPhlAn database), leaving five libraries analysed here for CGG101397.

We compared taxon abundances across samples using a suite of analyses described in (20) and (10). In order to reduce stochastic effects of low-abundance taxa, we excluded those with a relative abundance <1% across all samples. The Shannon index was used as a measure of microbial diversity at the genus level and was calculated using the function 'diversity' of the R package 'vegan' (http://cran.r-project.org/package=vegan). Genus-level diversity was also characterized using Principal Coordinate Analyses (PCoA) of Bray-Curtis distances among microbial profiles and the R function 'pcoa'. The structure among microbial profiles at the genus level was established using the R package 'pvclust', Manhattan distances and an average linkage clustering method (33). This method allowed estimating cluster support through Approximately Unbiased p-values and Bootstrap Probabilities from 10,000 bootstrap iterations. In order to identify the ecological origin of the microbial diversity retrieved from ancient horse remains, the microbial profiles were finally compared to those previously published from human (34) and soil samples (35) means of PCoA of Bray-Curtis distances at the genus level.

The relative abundances of microbial classes in the ancient horses samples **(Figure S3.1)**, PCoA **(Figure S3.2)** and hierarchical clustering analyses of genus abundances **(Figure S3.3)** showed a high similarity amongst the microbial profiles characterized from the different libraries obtained from each individual sample. Per-sample clustering of microbial profiles was also supported by hierarchical clustering even when different samples showed high relative abundance similarities in PCoA, e.g., Batagai and Yukagir. We also found that the microbial diversity across all ancient samples was closer to that observed in soil samples than in human samples **(Figure S3.4)**. Altogether, these results support the deposition soil as the main source of the microbial diversity recovered, as well as the low impact, if any, of contamination derived from humans during the handling of the samples post-excavation.

Our analyses identified a clear segregation between the microbial profiles of specimens Batagai and Yukagir on the one hand, and of the rest of the samples on the other hand **(Figures S3.1-3.3)**. This segregation was driven by high relative abundances in the microbial class *Gammaproteobacteria* (82.2-99.5%) in the Batagai and Yukagir specimens, whereas the microbial profiles of the other samples were dominated by

*Actinobacteria* (22.7-100.0%; **Figure S3.1**). In Batagai and Yukagir, high abundances of *Gammaproteobacteria* were caused by high abundances of the environmental *Pseudomonas* genus (73.7-97.5% for *P. fluorescens* and unclassified *Pseudomonas*), which is identified in the other samples at lower abundances (07.5%). The *Pseudomonas* genus was previously found to dominate the microbial sequence data generated from a ~700,000 year-old horse sample preserved from the permafrost of the Yukon Territory, Canada (13).

Among the other ancient horse samples analysed here, CGG101397 appeared differentiated from the other specimens (in accordance with (10)), due to high abundances of *Actinobacteria*, the only class identified in four out of the five libraries profiled for this sample. In CGG101397, the *Actinobacteria* class was characterized by a single genus, *Mycobacterium* (unclassified species); lower but relatively high abundances in *Mycobacterium* were also detected in other samples (4.81-54.4%), together with the *Rhodococcus erythropolis* species (0.0-33.5%).

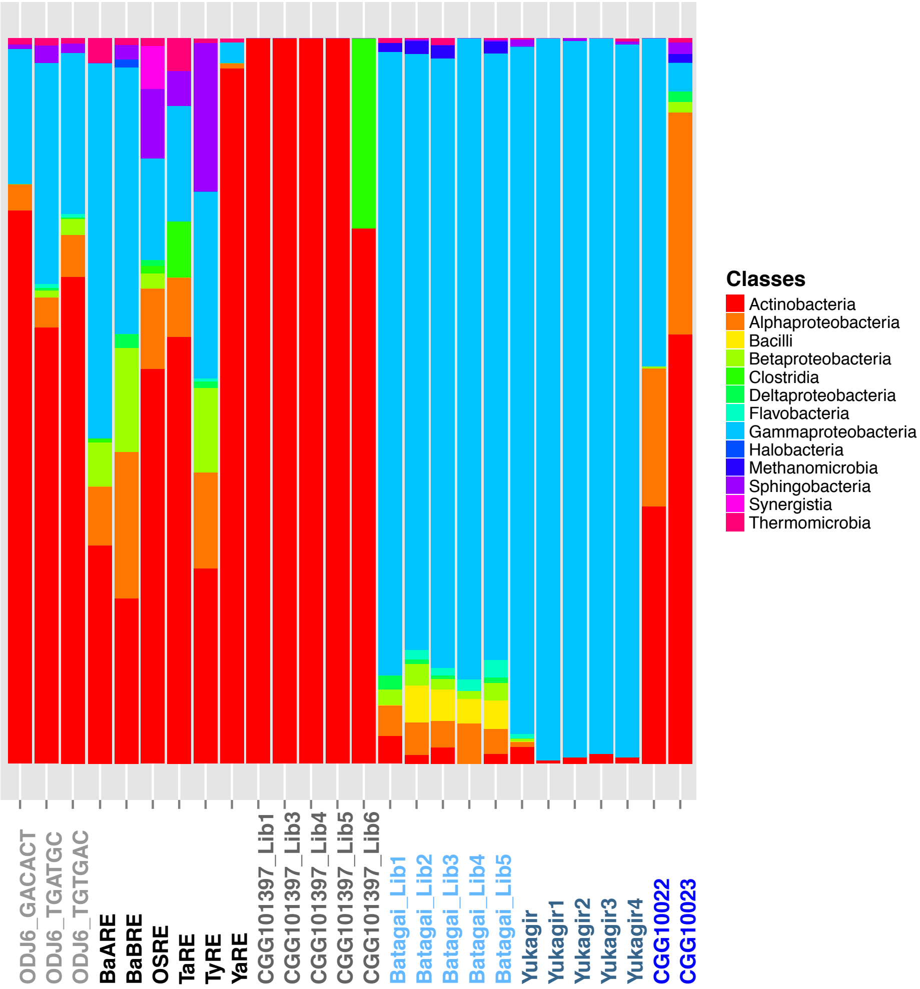


**Figure S3.1. Relative abundance of microbial classes in the DNA extracts of ancient horse remains preserved in the permafrost.**

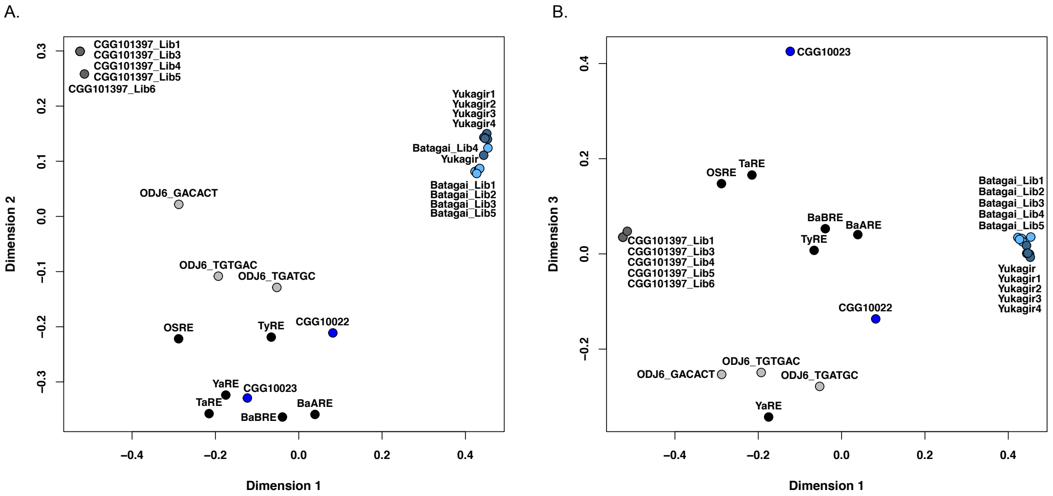


**Figure S3.2. Principal Coordinate Analysis of Bray-Curtis distances between microbial DNA profiles at the genus level in the DNA extracts of ancient horse remains preserved in the permafrost.**

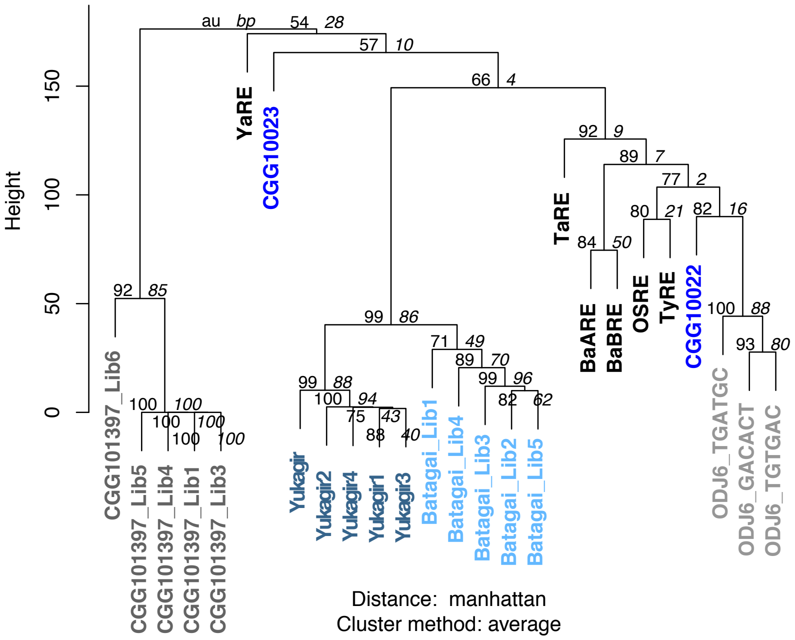


**Figure S3.3. Hierarchical clustering of Manhattan distances between microbial DNA at the genus level in the DNA extracts of ancient horse remains preserved in the permafrost.**

Cluster support is given approximately unbiased p-values (au) and bootstrap values (bp; 10,000 iterations).

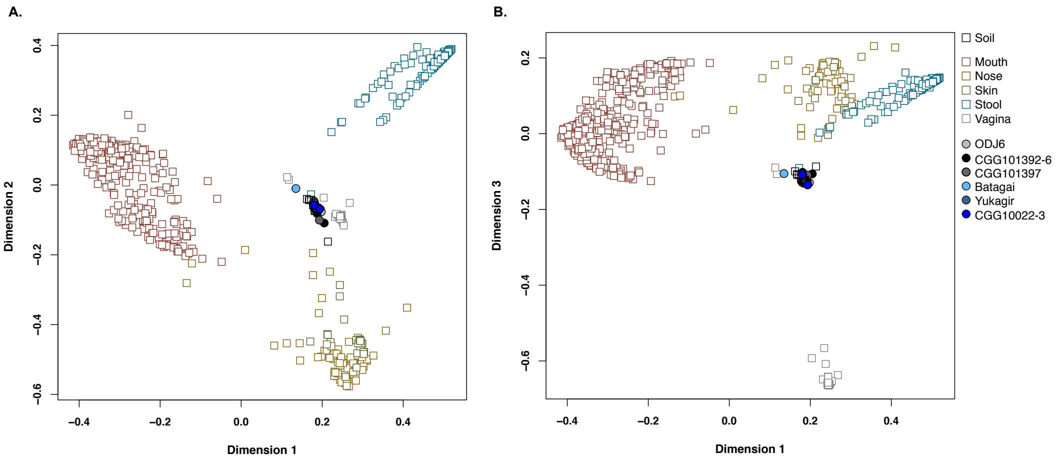


**Figure S3.4. Principal Coordinate Analysis of Bray-Curtis distances between microbial DNA profiles at the genus level in various human associated microbiomes, soils and ancient permafrost horse remains.**

## *3.1 Supplementary Tables for Section 3*

Table S3.1. Information underlying the microbial DNA profiles of ancient horse specimens.

| Horse ID | Reference | Library | #trimmed reads | #mapped reads[a] | # genera | Shannon index[b] |
|---|---|---|---|---|---|---|
| Yukagir | This study | Yukagir | 47,813,678 | 17,208 | 16 | 0.42 |
| | | Yukagir1 | 4,711,842 | 1,790 | 3 | 0.13 |
| | | Yukagir2 | 7,255,501 | 2,977 | 5 | 0.17 |
| | | Yukagir3 | 4,319,878 | 1,598 | 4 | 0.17 |
| | | Yukagir4 | 6,113,260 | 2,327 | 5 | 0.19 |
| Batagai | This study | Batagai_Lib1 | 60,624,116 | 5,797 | 11 | 0.74 |
| | | Batagai_Lib2 | 218,190,094 | 20,824 | 23 | 1.18 |
| | | Batagai_Lib3 | 241,003,960 | 20,576 | 22 | 1.04 |
| | | Batagai_Lib4 | 22,967,850 | 1,909 | 7 | 0.75 |
| | | Batagai_Lib5 | 211,992,359 | 17,465 | 22 | 1.17 |
| CGG101397 | This study | CGG101397_Lib1 | 1,981,276 | 408 | 1 | 0.00 |
| | | CGG101397_Lib2 | 565,437 | 59 | unclassified | n/a |
| | | CGG101397_Lib3 | 2,020,738 | 393 | 1 | 0.00 |
| | | CGG101397_Lib4 | 1,621,611 | 308 | 1 | 0.00 |
| | | CGG101397_Lib5 | 1,722,373 | 316 | 1 | 0.00 |
| | | CGG101397_Lib6 | 1,146,356 | 280 | 2 | 0.57 |
| ODJ6 | This study | ODJ6_GACACT | 9,580,824 | 8,965 | 12 | 1.27 |
| | | ODJ6_TGATGC | 20,127,349 | 23,725 | 23 | 1.56 |
| | | ODJ6_TGTGAC | 84,990,191 | 75,369 | 32 | 1.73 |
| CGG10022 | Schubert et al. 2014 (16) | CGG10022 | 49,964,648 | 3,100 | 12 | 1.99 |
| CGG10023 | Schubert et al. 2014 (16) | CGG10023 | 31,478,854 | 10,952 | 34 | 1.81 |
| BaARE | Der Sarkissian et al. 2014 (10) | CGG101393AR | 11,862,758 | 7,858 | 20 | 2.36 |
| BaBRE | Der Sarkissian et al. 2014 (10) | CGG101393BR | 8,969,882 | 8,410 | 27 | 2.78 |
| OSRE | Der Sarkissian et al. 2014 (10) | CGG101392R | 10,871,116 | 11,421 | 19 | 2.25 |
| TaRE | Der Sarkissian et al. 2014 (10) | CGG101396R | 12,384,464 | 8,407 | 18 | 2.56 |
| TyRE | Der Sarkissian et al. 2014 (10) | B_CGG101395R | 12,023,328 | 9,276 | 18 | 2.17 |
| YaRE | Der Sarkissian et al. 2014 (10) | CGG101394R | 7,033,095 | 12,991 | 15 | 1.50 |

[a] to the MetaPhlAn database. [b] at the genus level.

# 4 Section 4: Genomic variation

## *4.1 SNP variation*

### 4.1.1 Variant calling against the horse reference EquCab2.0

Variants were called against the horse reference genome, as detailed in the Supplementary Information section S8.3 (13) and in (30), using the PALEOMIX pipeline (20).

### 4.1.2 Functional categorization of SNPs variation

Filtered SNPs were categorized according to their effect on genes, transcripts, protein sequences and regulatory regions using the Ensembl Variant Effect Predictor script and annotations from Ensembl v76 (36), as previously described in (30) and (16). Chromosomes X (chrX), which shows variable ploidy across samples, and the pseudo-chromosome Unknown (chrUn), which consists of contigs that could not be scaffolded within autosomal and sex chromosomes, were excluded from the functional characterization. When a variant was assigned to multiple classifications (e.g. variants adjacents to multiple genes/transcripts), we counted this variant only once in each category. This rule was also applied to each classification related to genes. The following categories of variants are listed:

1. Genes
1.1. *Intron*, intronic regions.
1.2. *Non-coding exon*, variant found in a non-coding region of a gene (e.g. non-coding RNA).
1.3. *5' UTR*, variant found in the 5' untranslated region of a gene.
1.4. *3' UTR*, variant found in the 3' untranslated region of a gene.
1.5. *Splice site*, variant found in the splice region: within 1-3 bp of the exon, or within 3-8 bp of the intron.
1.6. *Mature miRNA*, variant found within the sequence of mature miRNA.
1.7. *Coding exons*, coding regions.
1.7.1. *Frameshift*, variants leading to a frameshift the amino acid sequence.
1.7.2. *Synonymous*, variants not changing the encoded amino acid.
1.7.3. *Non-synonymous* variants, leading to a different encoded amino acid.
1.7.4. *Stop gain*, variant leading to gain a stop codon.
1.7.5. *Stop loss*, variant leading to lose a stop codon.

2. Outside genes
2.1. *Intergenic*, variant located between genes but not in 1.1 and 1.2.
2.2. *Upstream*, variant found less than 5 kb upstream of the 5'-termini of a gene.
2.3. *Downstream*, variant found less than 5 kb downstream of the 3'-termini of a gene.

Results for all genomes sequenced in this study are shown in **Table S4.2**.

### 4.1.3 Non-synonymous mutations specific to Yakutian horses.

We collected mutations potentially specific to Yakutian horses, using criteria based on the work from Baye and colleagues (37). The ancient sample Batagai was found to be genetically differentiated from more recent sample CGG101397 and modern Yakutian horses (see **section S5.3**), it was then excluded from the population of Yakutian horses in this specific analysis. Briefly, we collected all sites for which a variant was called in at least 8 out of 10 Yakutian horse genomes and 16 out of the 19 domesticated horses present in our comparative panel. We did not filter SNPs by physical proximity, (unlike (37), who suggested a minimum distance of 0.3 cM in humans) to accommodate downstream analyses for which different marker densities may be required.

We then extracted loci with differences in allele frequency superior to 0.4, corresponding to 429 SNPs, including 170 non-synonymous (in 130 genes) and 259 synonymous mutations (dN/dS = 0.68). The corresponding 170 non-synonymous substitutions are listed in **Table S4.3**.

We carried out functional enrichment for the 130 genes as described in **section S7.3** Using humans as model organisms, no PheWAS or other phenotype was found to be significantly enriched; other results for functionnal enrichment can be found in **Tables S4.4-S4.8**.

### 4.1.4 Analysis of loci associated with important Mendelian traits in horses

Using the genotyping calls produced above (see **section S4.1.1**), we examined SNPs at 49 loci for known Mendelian traits, including traits collected in the Online Mendelian Inheritance in Animals (OMIA) database (38) in (13) and in (39) as described in (16) (**Table S4.9**). We attempted to call SNPs at all loci in both modern and ancient samples, and in cases where coverage was incompatible with our genotyping criteria, we recorded the number of high-quality nucleotides (BaseQ≥35) observed at the given site.

Results are listed in **Tables S4.10-S4.11**. In the majority of cases, Batagai and the other Late Pleistocene horses belonging to the same population cluster (specimens CGG10022 and CGG10023) carry the reference alleles with the exceptions of genes mainly associated with racing performance (*COX4/1*, *ACN9*), and body size (*ZFAT*). For example, heterozygosity is observed within all three specimens for *ACN9*, while homozygosity for the alternative allele is observed in *CKM* in Batagai and heterozygosity in CGG10023. Alleles associated with racing performance are also carried by modern Yakutian horses and sample CGG101397. Additionally, two loci in *PROP1* are associated with dwarfism, and Batagai is homozygote alternate for both loci, while this condition is not observed in Late Pleistocene and modern Yakutian horses. Interestingly, Batagai also shows alternate allele homozygosity for *HMGA2,* which is associated with larger body size. This condition is also shared with four modern Yakutian horses (Yak2, Yak3, Yak5 and Yak7). While CGG10022 and CGG10023 are heterozygous at four loci in *ZFAT*, Batagai is homozygous for the non-alternate allele, as the reference. These SNPs located in *ZFAT*, are associated with wither height in domesticated horses and the heterozygosity condition of one of them (at positions chr9:74,798,143) has been associated with a ~0.5 cm increase in height at the withers (40). Modern Yakutian horses and CGG101397 show both heterozygosity and homozygosity for the alternate allele, except for three horses Yak2, Yak7 and Yak8 carrying the same allelic status as Batagai.

We observed homozygosity for recessive MC1R mutations associated with chestnut coat (Yak6, Yak8, and Yak11), suggesting that these horses were chesnut in color. Specimens Yak3 and Yak7 were homozygous for the 11 bp recessive deletion associated with black coat color, suggesting these horses were black. No allele associated with Tobiano and Sabino spotting, Leopard complex spotting, silver coat color, and champagne dilution were detected across modern and ancient horses from Yakutia. Finally, none of the mutations associated with various coat color phenotypes was observed in the ancient samples Batagai, CGG101397 and some modern Yakutian horses (Yak5 and Yak9).

### *4.2 Screening for segmental duplications*

We identified segmental duplications using mrCaNaVaR v0.31 and mrFAST v2.0.0.5 (41). BED files were manipulated using BEDTools v2.0.17 (42). The analysis was carried out using the Thoroughbred horse (Twilight) and the modern Yakutian horses (Yak1 to Yak9).

We first masked the reference sequence to account for existing duplication in the EquCab2.0 reference genome. To do this, we generated overlapping 36 bp kmers across the entire nuclear genome, using a step size of 5 bp. The genome was then indexed using mrFAST (using a window-size of 12), the 36-mers were mapped back to this genome, and the location of any 36-mer mapping to more than one region was masked in the genome. We additionally masked the regions corresponding to the UCSC repeat-tracts ("Interrupted repeats", "Repeat Masker", and "Simple repeats"). Consequently, a total of 1.2 Gbp of the reference genome was masked, representing ~49% of the EquCab2.0 reference genome. The masked FASTA reference genome was then indexed using mrFAST as above. We next split the reads of the genomes of interest into non-overlapping 36-mers, starting 10 bp from the 5'-end, thereby excluding the more variable (quality-wise) 5'- and 3'-regions of the read. These were mapped to the masked FASTQ using mrFAST as described above. Before investigating the copy-number variation in our samples, we expanded the masked regions in the FASTQ sequences by 36 bp, to account for lower coverage in regions proximate to the masked regions, and indexed the resulting FASTQ sequences using mrCaNaVaR, specifying no gaps.

We then called read-depth and copy-numbers for each window in each specimen with mrCaNaVaR using the alignments generated by mrFAST above. We defined segmental duplications as any region consisting of at least 5 continuous copy-windows ("CW"), covering at least a total of 10 Kb. In this region the copy-number exceeds the mean by at least three standard deviations as calculated using the control regions identified by mrCaNaVaR, while allowing one non-trailing window exceeding the mean by at least 2 standard deviations. The mean and standard deviation were calculated separately for the X chromosome in male specimens. Before we collected regions with segmental duplications, we excluded putative mis-assembled regions where the copy number exceeded 50-fold the ploidy of the given chromosome in any sample. We also excluded any region in which segmental duplications were present in the Illumina sequence data from the Thoroughbred individual underlying the horse reference genome, in order to account for some segmental duplications existing in other domesticated horses and duplications not recorded in the EquCab2.0 reference genome (which was originally assembled based on Sanger sequencing (29)). This resulted in the exclusion of 176 million bp. We then plotted the distribution of copy-numbers called for control regions for each sample (excluding the X chromosome). In order to ensure that results were comparable across samples, we excluded from downstream analyses Yak2 and Yak7, for which the distribution deviated from the remaining samples (**Figure S4.1**), possibly because these genomes were sequenced at a much higher coverage (18.38 and 21.63X versus 9.48-11.64X for the other Yakutian samples). We then called segmental duplications using the same method described above.

We next concatenated the filtered BED files with BEDTools v2.0.17 for the seven Yakutian horse genomes considered and calculated the number of individuals for which a given region was called. We employed the BEDTools command 'genomecov' with arguments '-bg', and determined overlap with annotated sequences from Ensembl v76 (excluding pseudo-genes and introns) using the BEDTools intersect command with arguments '-loj'. For each Ensembl v76 transcript, we recorded the maximum number of samples for which a given segmental duplication overlapped that transcript at any window. Overall, we identified a total number of 260 regions affected by segmental duplications within Yakutian horses.

We compared these results with previous scans for CNVs in the horse (43, 44) and observed that a large proportion of putative segmental duplications from these studies were contained in the set of regions collected for the Thoroughbred (Twilight). For (43), we found an overlap accounting for 2.8 Mbp of the 5.0 Mbp (their Table S4) of regions called in that study, and for (44), we found an overlap accounting for 12.0 Mbp of the 28.5Mbp they analysed ((44); Table S3). These regions may thus represent a combination of common segmental duplications among the domesticated horses, and false positives common to our and their studies.

When examining the functional characterization of the segmental duplications identified in Yakutian horses, we found 178 annotated genomic regions, of which 170 were genes coding for 91 proteins of known function (**Table S4.12**) and 79 uncharacterized proteins (**Table S4.13**). In the remaining regions covered by segmental duplications, we found four small nuclear RNA, one RNA, one small nucleolar RNA, one ribosomal RNA, and one microRNA.

We carried out functional enrichment (GO-terms, pathways, phenotypes, disease) for the genes encompassed by the detected segmental duplications as described in **section 7.3**. No GO-term was found to be significantly enriched; other enrichment results are presented in **Tables S4.13-S4.15**.

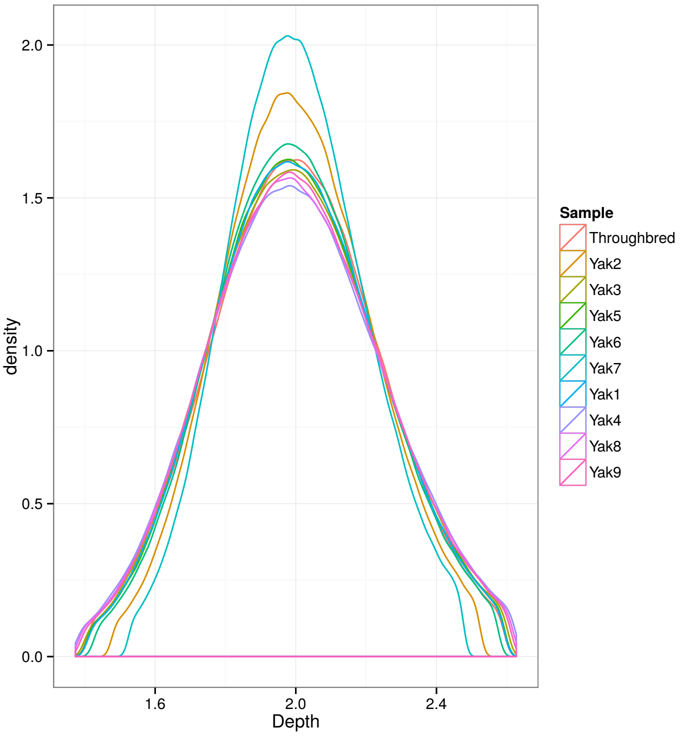


**Figure S4.1. Distribution of copy-numbers called for control regions in modern Yakutian horses and the Thoroughbred Twilight individual underpinning the horse reference genome.** Chromosome X was excluded from these analyses.

## *4.3 Inbreeding and heterozygosity*

### 4.3.1 Genome-wide heterozygosity estimates

The average heterozygosity of each horse described in this study was estimated as the number of segregating sites by calculating the Watterson estimator ($\theta_w$) (45), in 50 kb sliding windows (step-size of 10 Kb). This was done using the program ANGSD (46), which accounts for genotype uncertainties by using genotype likelihoods and priors based on site frequency spectrum (SFS). Priors for the autosomes were constructed using the SFS (47) for the chromosome 22. For high-quality genomes we excluded windows in which less than 45 out of 50 kb (90%) were covered. Additionally, $\theta_w$ was calculated either considering all substitution classes or disregarding transitions, as the latter represent the most typical nucleotide mis-incorporation at damaged sites.

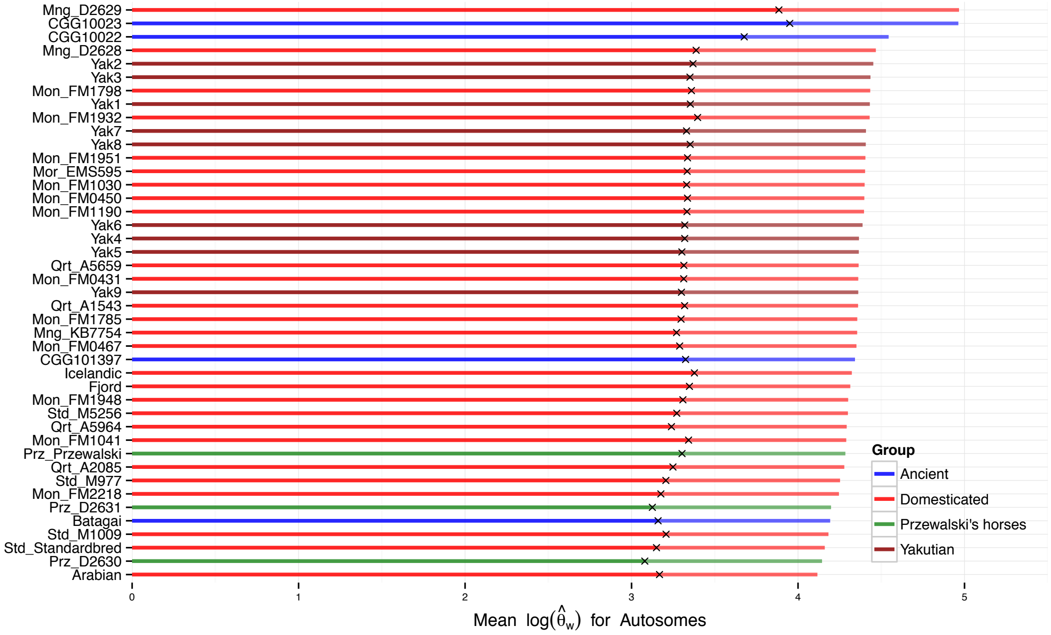


**Figure S4.2. Average autosomal heterozygosity in ancient and modern horses.**

Black dots indicate average log($\theta_W$) values estimated considering both transitions and transversions, while black crosses indicate average log($\theta_W$) values estimated disregarding transitions. Heterozygosity estimates of two Late Pleistocene horse genomes characterized by (16) (specimens CGG10022 and CGG10023), and all the genomes from our comparative panel are shown for comparison.

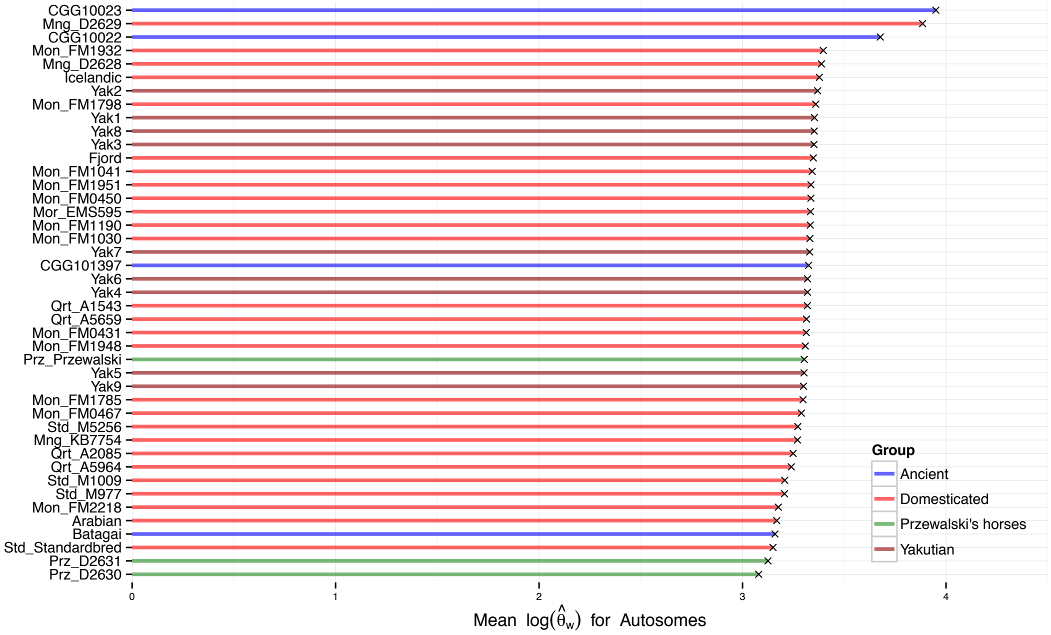


**Figure S4.3. Average autosomal heterozygosity in ancient and modern horses, disregarding transitions.**
See **Figure S4.2** for additional captions.

## 4.3.2 Inbreeding estimates

We estimated inbreeding in Yakutian horses as the proportion of genomic fragments being mostly homozygous, or “homozygous-by-descent” (HBD), following terminologies and methodologies similar to those in (48), and applied to complete horse and equid genomes in (16) and (30). We used the $\theta_W$ estimator across the genome (see **section S4.3.1**) to detect HBD regions. In these analyses, we replaced missing data by the closest neighboring valid values on a per-chromosome basis. We determined the coordinates of the HBD genomic tracks by identifying segments with local changes in the average $\theta_W$ estimates along the genome using the R package ‘changepoints’ ((49); http://cran.r-project.org/web/packages/changepoint/index.html) and the binary segmentation algorithm (method=“BinSeg”) allowing up to 12 breakpoints per chromosome (Q=12). For each segment, we calculated the heterozygosity estimates given by the average of log ($\theta_W$), which were used to weigh genomic track lengths (deduced from their external coordinates) before their density was plotted with the ‘density’ R function (from 100,000 points). A bimodal distribution of the $\theta_W$ estimate values indicates inbreeding in the corresponding individual. We determined the coordinate of the lowest point (pit) between the two modes using numerically estimates of first and second derivatives of the density function (using absolute tolerance of $10^{-7}$). The pit coordinate provided a threshold used to classify genomic segments as showing “high” or “low” heterozygosity (vertical red line). The total size of the “low” heterozygosity regions was divided by the total size of all regions to give the proportion of HBD tracks, or “inbreeding coverage estimate” (value shown adjacent to the vertical line). The average log($\theta_W$) values and positions of the genomic tracks were determined both including (full lines) and excluding transitions (dashed lines) to account for mis-incorporations arising from post-mortem DNA damage in ancient genomes **(FigureS4.4-4.5)**.

Results suggest low levels of inbreeding (4.1-10.1% with transitions; 4.0%-10.4% without transitions) among modern Yakutian horses and low to not existing levels among the ancient horses (CGG101397: 6.7/6.6%; CGG10022: 1.1/0.2%; CGG10023: 0.0/0.0%; Batagai: 9.6/8.3%; with and without transitions respectively; **Figure S4.4**).

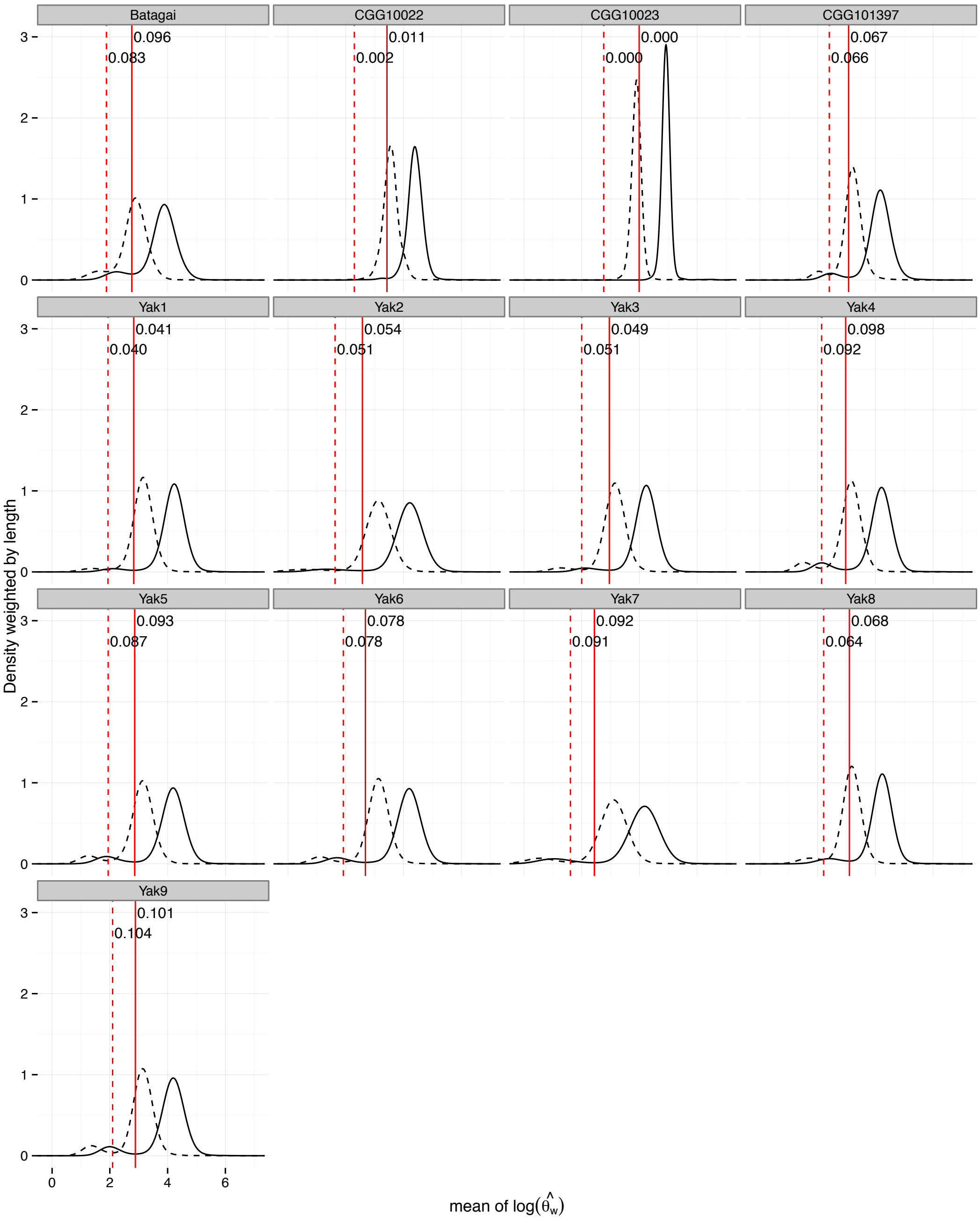


**Figure S4.4. Inbreeding coverage estimates for Yakutian and Late Pleistocene horses.** Bimodal distributions are typical of inbred individuals. The red vertical line indicates the local minim separating the two modes of the distribution, whenever present. The number adjacent to the line is th area under the left peak, and provides an estimate for inbreeding. Solid lines indicate that both transition and transversions are considered, in contrast to dashed lines for which only transversions were used.

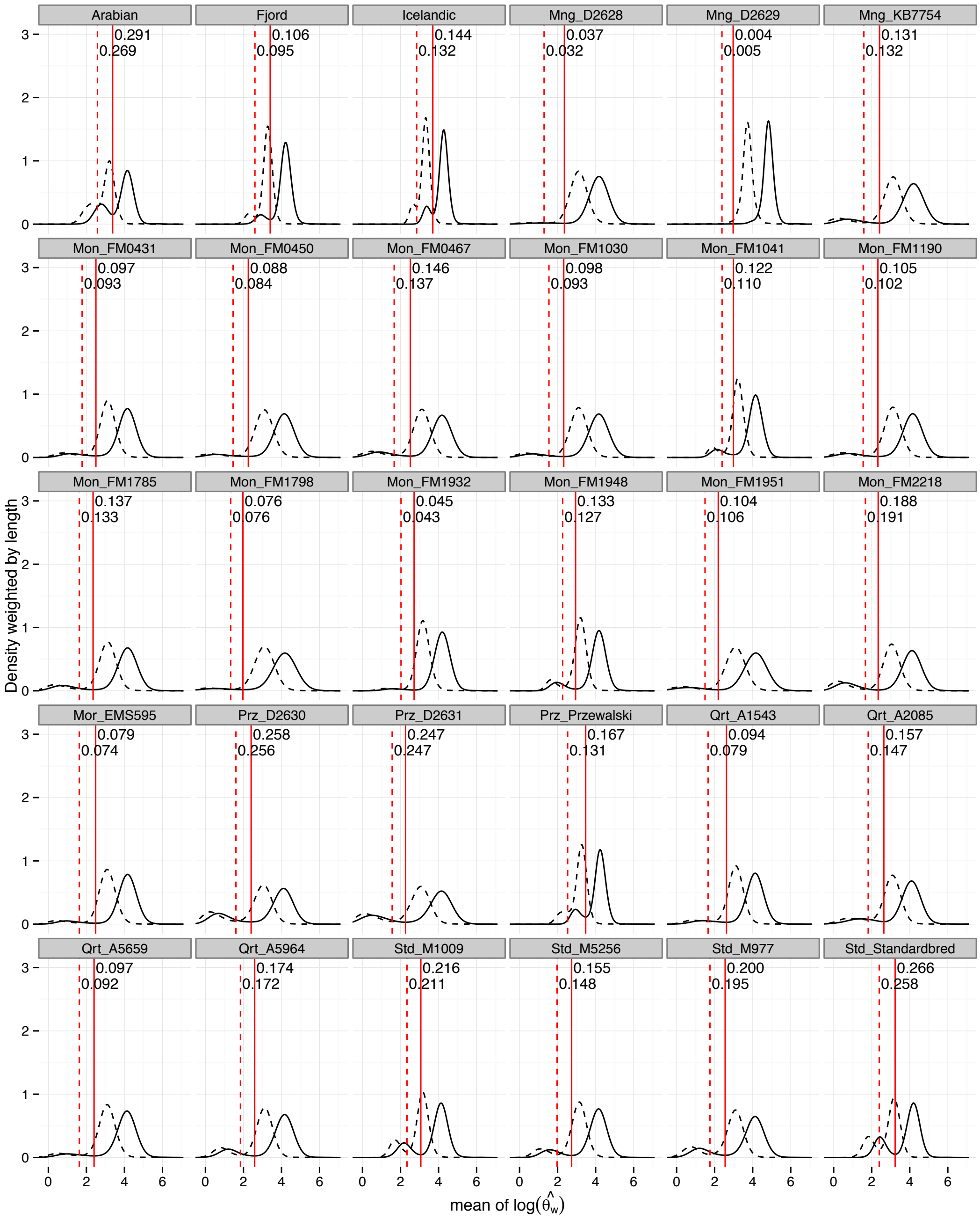


**Figure S4.5. Inbreeding coverage estimates for other domesticated horses.**
See **Figure S4.4** for figure caption.

## *4.4 Supplementary Tables for Section 4*

## Table S4.1. Maximum per-sample depth used to filter base calls.

| Sample | Maximum Depth |
|---|---|
| Batagai | 54 |
| CGG10023 | 35 |
| CGG10022 | 72 |
| CGG101397 | 84 |
| Yak1 | 25 |
| Yak2 | 39 |
| Yak3 | 25 |
| Yak4 | 23 |
| Yak5 | 25 |
| Yak6 | 27 |
| Yak7 | 48 |
| Yak8 | 23 |
| Yak9 | 23 |

| Sample | Maximum Depth |
|---|---|
| Arabian | 39 |
| Fjord | 23 |
| Icelandic | 30 |
| Mng_D2629 | 54 |
| Mng_D2628 | 52 |
| Mon_FM0431 | 65 |
| Mon_FM0450 | 66 |
| Mon_FM0467 | 86 |
| Mon_FM1030 | 109 |
| Mon_FM1041 | 52 |
| Mon_FM1190 | 102 |
| Mon_FM1785 | 80 |
| Mon_FM1798 | 130 |
| Mon_FM1932 | 60 |
| Mon_FM1948 | 51 |
| Mon_FM1951 | 104 |
| Mon_FM2218 | 73 |
| Mor_EMS595 | 95 |
| Prz_D2631 | 56 |
| Prz_Przewalski | 30 |
| Prz_D2630 | 39 |
| Qrt_A1543 | 80 |
| Qrt_A5659 | 83 |
| Qrt_A5964 | 114 |
| Qrt_A2085 | 65 |
| Std_M977 | 82 |
| Std_M5256 | 74 |
| Std_M1009 | 51 |
| Std_Standardbred | 37 |
| Throroughbred | 81 |

## Table S4.2. Classification of functional variants identified across the Yakutian horse genomes.

| | Batagai | | CGG101397 | | Yak1 | | Yak2 | | Yak3 | | Yak4 | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Variant class | Genes | Variants | Genes | Variants | Genes | Variants | Genes | Variants | Genes | Variants | Genes | Variants |
| All variants | 26,697 | 6,762,877 | 26,648 | 5,814,974 | 26,639 | 6,536,920 | 26,653 | 7,084,640 | 26,650 | 6,647,543 | 26,637 | 6,369,435 |
| Not genic | | 5,115,400 | | 4,390,671 | | 4,951,567 | | 5,405,329 | | 5,040,171 | | 4,827,587 |
| Intergenic | | 4,494,040 | | 3,837,388 | | 4,358,504 | | 4,757,343 | | 4,434,041 | | 4,242,032 |
| Upstream | | 337,303 | | 306,259 | | 319,064 | | 350,223 | | 325,963 | | 315,292 |
| Downstream | | 312,317 | | 273,186 | | 300,076 | | 326,709 | | 306,393 | | 296,183 |
| Genic | 21,425 | 1,746,322 | 20,800 | 1,513,251 | 21,249 | 1,679,096 | 21,582 | 1,780,726 | 21,408 | 1,703,005 | 21,145 | 1,634,953 |
| Intron | 16,190 | 1,686,836 | 16,069 | 1,456,391 | 16,080 | 1,623,215 | 16,111 | 1,721,301 | 16,093 | 1,646,300 | 16,035 | 1,580,269 |
| Non-coding Exon | 2,755 | 6,844 | 2,368 | 5,634 | 2,740 | 7,132 | 2,967 | 8,518 | 2,836 | 7,503 | 2,692 | 6,962 |
| 5' UTR | 1,837 | 3,219 | 1,895 | 3,518 | 1,571 | 2,597 | 1,679 | 2,871 | 1,586 | 2,538 | 1,562 | 2,474 |
| 3' UTR | 2,487 | 4,010 | 2,212 | 3,541 | 2,406 | 3,794 | 2,439 | 4,004 | 2,429 | 3,909 | 2,275 | 3,621 |
| Splice Site | 4,613 | 6,985 | 4,553 | 7,002 | 4,357 | 6,369 | 4,589 | 6,937 | 4,317 | 6,383 | 4,205 | 6,157 |
| Mature miRNA | 28 | 29 | 33 | 38 | 36 | 40 | 44 | 47 | 32 | 32 | 38 | 39 |
| Coding Exon | 11,349 | 29,000 | 11,004 | 28,535 | 10,900 | 26,673 | 11,107 | 27,849 | 11,033 | 26,743 | 10,670 | 26,205 |
| Frameshift | 1,588 | 2,080 | 1,799 | 2,675 | 1,272 | 1,503 | 1,463 | 1,801 | 1,242 | 1,482 | 1,237 | 1,466 |
| Synonymous | 10,610 | 25,288 | 10,131 | 23,938 | 10,176 | 23,777 | 10,349 | 24,518 | 10,330 | 23,883 | 9,934 | 23,378 |
| Non-synonymous | 9,051 | 21,853 | 8,736 | 20,837 | 8,799 | 20,956 | 8,808 | 21,933 | 8,827 | 21,346 | 8,618 | 20,734 |
| Stop gain | 185 | 190 | 189 | 194 | 204 | 211 | 202 | 212 | 214 | 224 | 191 | 197 |
| Stop loss | 13 | 13 | 20 | 20 | 15 | 15 | 16 | 16 | 18 | 18 | 13 | 13 |

| | Yak5 | | Yak6 | | Yak7 | | Yak8 | | Yak9 | |
|---|---|---|---|---|---|---|---|---|---|---|
| Variant class | Genes | Variants | Genes | Variants | Genes | Variants | Genes | Variants | Genes | Variants |
| All variants | 26,650 | 6,508,075 | 26,634 | 6,625,881 | 26,655 | 7,098,114 | 26,662 | 6,482,224 | 26,632 | 6,499,579 |
| Not genic | | 4,944,158 | | 5,033,776 | | 5,414,527 | | 4,908,118 | | 4,932,067 |
| Intergenic | | 4,347,364 | | 4,426,444 | | 4,749,072 | | 4,313,267 | | 4,336,606 |
| Upstream | | 321,281 | | 326,825 | | 362,115 | | 320,938 | | 321,250 |
| Downstream | | 301,749 | | 307,257 | | 333,333 | | 300,303 | | 300,417 |
| Genic | 21,225 | 1,658,515 | 21,315 | 1,687,337 | 21,572 | 1,788,196 | 21,275 | 1,668,031 | 21,178 | 1,660,380 |
| Intron | 16,041 | 1,602,973 | 16,072 | 1,630,665 | 16,092 | 1,726,180 | 16,079 | 1,611,908 | 16,013 | 1,604,752 |
| Non-coding Exon | 2,725 | 7,264 | 2,795 | 7,293 | 2,947 | 8,800 | 2,740 | 6,965 | 2,703 | 7,186 |
| 5’ UTR | 1,583 | 2,579 | 1,684 | 2,784 | 1,837 | 3,328 | 1,612 | 2,656 | 1,586 | 2,578 |
| 3’ UTR | 2,352 | 3,756 | 2,358 | 3,823 | 2,440 | 3,992 | 2,389 | 3,782 | 2,302 | 3,656 |
| Splice Site | 4,289 | 6,271 | 4,403 | 6,456 | 4,696 | 7,280 | 4,297 | 6,375 | 4,280 | 6,319 |
| Mature miRNA | 38 | 38 | 36 | 38 | 49 | 52 | 41 | 41 | 41 | 42 |
| Coding Exon | 10,876 | 26,684 | 10,941 | 27,071 | 11,296 | 28,998 | 10,993 | 26,967 | 10,726 | 26,416 |
| Frameshift | 1,306 | 1,566 | 1,366 | 1,654 | 1,646 | 2,146 | 1,284 | 1,532 | 1,277 | 1,518 |
| Synonymous | 10,162 | 23,698 | 10,199 | 23,972 | 10,436 | 25,065 | 10,268 | 24,018 | 10,027 | 23,523 |
| Non-synonymous | 8,659 | 20,614 | 8,713 | 21,160 | 8,933 | 23,006 | 8,739 | 21,089 | 8,603 | 20,965 |
| Stop gain | 185 | 191 | 177 | 183 | 208 | 218 | 205 | 210 | 188 | 198 |
| Stop loss | 14 | 15 | 17 | 18 | 17 | 18 | 17 | 18 | 11 | 11 |

## Table S4.3. Non-synonymous markers specific to Yakutian horses, ordered by specificity (1).

| Gene ID | Transcript ID | Gene Name | Strand | Chr | Pos | NucRef | NucAlt | Codon | CodonPos | AARef | AAAlt | Delta |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ENSECAG00000003780 | ENSECAT00000003667 | *OR6A2* | - | 7 | 76242089 | C | T | GCG | 2 | Arg | His | 0.7188 |
| ENSECAG00000010259 | ENSECAT00000010674 | *TRIM45* | + | 5 | 51983000 | G | A | CGG | 2 | Arg | Gln | 0.6701 |
| ENSECAG00000004104 | ENSECAT00000004094 |  | + | 7 | 76318945 | T | A | ATT | 2 | Ile | Asn | 0.6597 |
| ENSECAG00000019446 | ENSECAT00000020625 | *SH3RF3* | - | 15 | 535504 | C | T | TAC | 3 | Val | Ile | 0.6287 |
| ENSECAG00000010259 | ENSECAT00000010674 | *TRIM45* | + | 5 | 51984595 | G | A | GCA | 1 | Ala | Thr | 0.6188 |
| ENSECAG00000009561 | ENSECAT00000009811 | *FDXACB1* | - | 7 | 20458010 | C | T | GCG | 2 | Arg | His | 0.6042 |
| ENSECAG00000016457 | ENSECAT00000017372 | *E2F8* | - | 7 | 87149752 | C | G | GTC | 3 | Asp | His | 0.6029 |
| ENSECAG00000023844 | ENSECAT00000024681 | *LAMA2* | + | 10 | 75723569 | G | A | CGC | 2 | Arg | His | 0.5972 |
| ENSECAG00000023844 | ENSECAT00000028988 | *LAMA2* | + | 10 | 75723569 | G | A | CGC | 2 | Arg | His | 0.5972 |
| ENSECAG00000011862 | ENSECAT00000012267 | *C11orf1* | + | 7 | 20464163 | C | A | CGT | 1 | Arg | Ser | 0.5956 |
| ENSECAG00000021446 | ENSECAT00000029043 | *GRID1* | + | 1 | 84439712 | A | G | ATT | 1 | Ile | Val | 0.5799 |
| ENSECAG00000013829 | ENSECAT00000014395 | *LRRIQ3* | + | 5 | 88176264 | C | G | ACA | 2 | Thr | Arg | 0.5699 |
| ENSECAG00000017119 | ENSECAT00000017998 | *PIGV* | - | 2 | 28826835 | C | G | CAA | 1 | Leu | Phe | 0.5654 |
| ENSECAG00000025066 | ENSECAT00000027056 | *PLA2G3* | + | 8 | 5892549 | G | A | CGC | 2 | Arg | His | 0.5556 |
| ENSECAG00000000258 | ENSECAT00000000196 | *STX11* | - | 31 | 21771618 | G | T | GAG | 3 | Leu | Ile | 0.5523 |
| ENSECAG00000011135 | ENSECAT00000011552 | *GPATCH4* | + | 5 | 41675513 | C | G | CCC | 2 | Pro | Arg | 0.5486 |
| ENSECAG00000021537 | ENSECAT00000023007 | *FNDC7* | - | 5 | 59259538 | T | C | GAT | 3 | Ile | Val | 0.5389 |
| ENSECAG00000016463 | ENSECAT00000017364 | *TMEM72* | - | 1 | 69377220 | G | C | GTT | 1 | Asn | Lys | 0.5359 |
| ENSECAG00000008441 | ENSECAT00000008613 | *ADORA2B* | - | 11 | 58222333 | G | C | TGG | 3 | Pro | Ala | 0.5278 |
| ENSECAG00000017257 | ENSECAT00000018273 |  | + | 21 | 28288290 | C | T | CGT | 1 | Arg | Cys | 0.5229 |
| ENSECAG00000001015 | ENSECAT00000000878 |  | + | 5 | 38598787 | C | T | CAC | 1 | His | Tyr | 0.5211 |
| ENSECAG00000001626 | ENSECAT00000002010 | *FAM35A* | - | 1 | 83411985 | G | T | AGT | 2 | Thr | Asn | 0.5188 |
| ENSECAG00000000955 | ENSECAT00000000829 |  | - | 7 | 73858338 | G | A | ACG | 3 | Arg | Cys | 0.5147 |
| ENSECAG00000000955 | ENSECAT00000000829 |  | - | 7 | 73858980 | A | G | AAA | 3 | Phe | Leu | 0.5147 |
| ENSECAG00000017255 | ENSECAT00000018215 |  | - | 12 | 20100747 | T | A | CTT | 2 | Lys | Met | 0.5139 |
| ENSECAG00000016678 | ENSECAT00000017506 |  | - | 3 | 69502178 | C | T | CAT | 1 | Met | Ile | 0.5139 |
| ENSECAG00000007957 | ENSECAT00000008003 |  | + | 4 | 96520769 | T | G | CTC | 2 | Leu | Arg | 0.5062 |
| ENSECAG00000024755 | ENSECAT00000026678 |  | - | 11 | 41661616 | C | T | GCG | 2 | Arg | His | 0.5056 |
| ENSECAG00000002127 | ENSECAT00000001972 |  | + | 7 | 72759229 | G | A | TGT | 2 | Cys | Tyr | 0.5035 |
| ENSECAG00000005646 | ENSECAT00000005595 |  | - | 14 | 92740397 | G | C | GAA | 1 | Phe | Leu | 0.5033 |
| ENSECAG00000024306 | ENSECAT00000026216 | *DENND4A* | + | 1 | 126743368 | C | G | ACA | 2 | Thr | Arg | 0.5029 |
| ENSECAG00000017136 | ENSECAT00000018105 | *HEMK1* | - | 16 | 36704036 | A | T | CAT | 2 | Met | Lys | 0.5000 |
| ENSECAG00000000480 | ENSECAT00000002615 | *ABI3BP* | + | 19 | 54500659 | A | G | ATC | 1 | Ile | Val | 0.5000 |
| ENSECAG00000000480 | ENSECAT00000002650 | *ABI3BP* | + | 19 | 54500659 | A | G | ATC | 1 | Ile | Val | 0.5000 |
| ENSECAG00000005434 | ENSECAT00000005420 | *CABS1* | - | 3 | 64622967 | A | G | AAT | 2 | Ile | Thr | 0.5000 |
| ENSECAG00000007941 | ENSECAT00000008326 | *MYO10* | + | 21 | 43825996 | G | A | GTG | 1 | Val | Met | 0.5000 |
| ENSECAG00000012702 | ENSECAT00000013150 | *PPBP* | - | 3 | 61847443 | C | G | TCT | 2 | Arg | Thr | 0.4965 |
| ENSECAG00000007434 | ENSECAT00000007431 | *PRR9* | - | 5 | 44362433 | G | A | CGC | 2 | Ala | Val | 0.4938 |
| ENSECAG00000007434 | ENSECAT00000007431 | *PRR9* | - | 5 | 44362443 | A | C | AGA | 3 | Ser | Ala | 0.4938 |
| ENSECAG00000009561 | ENSECAT00000009811 | *FDXACB1* | - | 7 | 20458221 | T | G | CAT | 3 | Met | Leu | 0.4889 |
| ENSECAG00000012474 | ENSECAT00000013260 | *TTF2* | - | 5 | 52012832 | A | G | CAT | 2 | Met | Thr | 0.4882 |
| ENSECAG00000001007 | ENSECAT00000000986 |  | - | 6 | 25639988 | C | T | GAC | 3 | Val | Ile | 0.4869 |
| ENSECAG00000019154 | ENSECAT00000020422 | *ADAMTSL4* | - | 5 | 46533866 | G | A | CCG | 3 | Arg | Trp | 0.4861 |
| ENSECAG00000009949 | ENSECAT00000010815 | *P2RX7* | + | 8 | 21751816 | G | A | GGT | 1 | Gly | Ser | 0.4853 |
| ENSECAG00000008940 | ENSECAT00000009129 | *TMEM40* | + | 16 | 4588968 | G | C | GGA | 2 | Gly | Ala | 0.4812 |
| ENSECAG00000026818 | ENSECAT00000028860 | *LRP11* | + | 31 | 16993179 | C | T | CGC | 1 | Arg | Cys | 0.4792 |
| ENSECAG00000010626 | ENSECAT00000011724 | *SLC47A2* | + | 11 | 58820685 | T | G | TTC | 1 | Phe | Val | 0.4779 |
| ENSECAG00000014004 | ENSECAT00000014633 | *SAMD7* | + | 19 | 10364690 | A | G | AAT | 2 | Asn | Ser | 0.4779 |
| ENSECAG00000001653 | ENSECAT00000006446 | *POLE* | - | 8 | 29601827 | G | A | CGG | 3 | Pro | Ser | 0.4778 |
| ENSECAG00000001653 | ENSECAT00000006622 | *POLE* | - | 8 | 29601827 | G | A | CGG | 3 | Pro | Ser | 0.4778 |
| ENSECAG00000003630 | ENSECAT00000003501 |  | + | 25 | 343444 | T | C | CTA | 2 | Leu | Pro | 0.4766 |
| ENSECAG00000012975 | ENSECAT00000013485 | *C8orf74* | + | 2 | 58762175 | T | C | TAC | 1 | Tyr | His | 0.4735 |
| ENSECAG00000014941 | ENSECAT00000015992 | *INPP5F* | - | 1 | 12724544 | G | A | CGA | 2 | Ser | Leu | 0.4722 |
| ENSECAG00000005898 | ENSECAT00000005907 |  | + | 10 | 78715436 | A | G | ATC | 1 | Ile | Val | 0.4722 |
| ENSECAG00000009742 | ENSECAT00000009966 | *S100A12* | + | 5 | 44158614 | G | A | GGT | 1 | Gly | Ser | 0.4711 |
| ENSECAG00000024080 | ENSECAT00000025941 |  | - | 7 | 20387970 | T | C | CAT | 3 | Met | Val | 0.4706 |
| ENSECAG00000010626 | ENSECAT00000011724 | *SLC47A2* | + | 11 | 58820680 | A | G | GAC | 2 | Asp | Gly | 0.4688 |

## Table S4.3. Non-synonymous markers specific to Yakutian horses, ordered by specificity (2).

| Gene ID | Transcript ID | Gene Name | Strand | Chr | Pos | NucRef | NucAlt | Codon | CodonPos | AARef | AAAlt | Delta |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ENSECAG00000012702 | ENSECAT00000013150 | *PPBP* | - | 3 | 61847087 | G | A | AGC | 2 | Ala | Val | 0.4688 |
| ENSECAG00000024031 | ENSECAT00000025828 | *KIAA0232* | + | 3 | 114722043 | A | C | ATG | 1 | Met | Leu | 0.4688 |
| ENSECAG00000026868 | ENSECAT00000028994 | *C6orf183* | + | 10 | 58985096 | C | G | GCC | 2 | Ala | Gly | 0.4673 |
| ENSECAG00000018781 | ENSECAT00000019992 | *MCMDC2* | - | 9 | 18407831 | C | T | GCG | 2 | Arg | His | 0.4653 |
| ENSECAG00000019768 | ENSECAT00000021105 | *GRAP2* | + | 28 | 36689689 | G | A | GTG | 1 | Val | Met | 0.4653 |
| ENSECAG00000004408 | ENSECAT00000004329 | *GPR1* | - | 18 | 80490590 | A | G | CAC | 2 | Val | Ala | 0.4647 |
| ENSECAG00000022292 | ENSECAT00000023750 |  | + | 25 | 10628814 | G | A | GGA | 1 | Gly | Arg | 0.4632 |
| ENSECAG00000024126 | ENSECAT00000025990 | *RFC1* | + | 3 | 87992262 | C | G | CTC | 1 | Leu | Val | 0.4625 |
| ENSECAG00000024126 | ENSECAT00000025999 | *RFC1* | + | 3 | 87992262 | C | G | CTC | 1 | Leu | Val | 0.4625 |
| ENSECAG00000000512 | ENSECAT00000000397 | *LBH* | - | 15 | 66450983 | T | C | GTT | 2 | Asn | Ser | 0.4618 |
| ENSECAG00000019204 | ENSECAT00000020316 |  | - | 5 | 40670705 | T | C | CTC | 2 | Glu | Gly | 0.4618 |
| ENSECAG00000005646 | ENSECAT00000005595 |  | - | 14 | 92739894 | C | T | ACT | 2 | Ser | Asn | 0.4588 |
| ENSECAG00000005646 | ENSECAT00000005595 |  | - | 14 | 92740029 | C | A | CCC | 2 | Gly | Val | 0.4588 |
| ENSECAG00000017048 | ENSECAT00000017967 | *CD101* | - | 5 | 52075210 | A | G | GTA | 3 | Tyr | His | 0.4575 |
| ENSECAG00000023120 | ENSECAT00000024726 | *ZFP69B* | - | 2 | 17689694 | T | C | GTT | 3 | Asn | Asp | 0.4563 |
| ENSECAG00000003862 | ENSECAT00000004733 | *NRDE2* | - | 24 | 33697644 | T | G | GTT | 3 | Asn | His | 0.4563 |
| ENSECAG00000014425 | ENSECAT00000015305 | *PIGK* | + | 5 | 85459673 | T | C | GTC | 2 | Val | Ala | 0.4549 |
| ENSECAG00000002749 | ENSECAT00000002789 | *TKTL2* | + | 2 | 71048367 | C | T | CGA | 1 | Arg | Stop | 0.4529 |
| ENSECAG00000002041 | ENSECAT00000001910 | *OR51B5* | - | 7 | 74123183 | A | T | AAA | 2 | Phe | Tyr | 0.4514 |
| ENSECAG00000022217 | ENSECAT00000024081 | *UTP6* | + | 11 | 40582906 | A | C | ATC | 1 | Ile | Leu | 0.4500 |
| ENSECAG00000019457 | ENSECAT00000020821 | *CCDC108* | - | 6 | 8596494 | T | C | TGT | 3 | Thr | Ala | 0.4500 |
| ENSECAG00000009647 | ENSECAT00000011076 | *MGAM* | + | 4 | 94831274 | A | G | ACC | 1 | Thr | Ala | 0.4500 |
| ENSECAG00000026818 | ENSECAT00000028860 | *LRP11* | + | 31 | 16993196 | G | T | TGG | 3 | Trp | Cys | 0.4479 |
| ENSECAG00000024126 | ENSECAT00000025990 | *RFC1* | + | 3 | 87992190 | A | G | ACT | 1 | Thr | Ala | 0.4477 |
| ENSECAG00000024126 | ENSECAT00000025999 | *RFC1* | + | 3 | 87992190 | A | G | ACT | 1 | Thr | Ala | 0.4477 |
| ENSECAG00000024657 | ENSECAT00000026613 | *CHMP7* | + | 2 | 51981404 | T | C | GTT | 2 | Val | Ala | 0.4449 |
| ENSECAG00000007147 | ENSECAT00000008062 | *CAPRIN1* | + | 12 | 950214 | C | A | CCG | 2 | Pro | Gln | 0.4444 |
| ENSECAG00000008488 | ENSECAT00000008751 | *MINA* | + | 19 | 56891113 | C | T | CGC | 1 | Arg | Cys | 0.4444 |
| ENSECAG00000026818 | ENSECAT00000028860 | *LRP11* | + | 31 | 16993218 | A | G | ACC | 1 | Thr | Ala | 0.4444 |
| ENSECAG00000007434 | ENSECAT00000007431 | *PRR9* | - | 5 | 44362551 | T | C | CGT | 3 | Thr | Ala | 0.4444 |
| ENSECAG00000011135 | ENSECAT00000011552 | *GPATCH4* | + | 5 | 41672314 | A | T | ACT | 1 | Thr | Ser | 0.4444 |
| ENSECAG00000017335 | ENSECAT00000017455 | *CDON* | - | 7 | 34819663 | A | G | GAA | 3 | Phe | Leu | 0.4444 |
| ENSECAG00000017335 | ENSECAT00000028894 | *CDON* | - | 7 | 34819663 | A | G | GAA | 3 | Phe | Leu | 0.4444 |
| ENSECAG00000017136 | ENSECAT00000018105 | *HEMK1* | - | 16 | 36705050 | G | T | TGC | 2 | Ala | Glu | 0.4444 |
| ENSECAG00000022378 | ENSECAT00000023985 | *IGSF3* | + | 5 | 52444326 | G | T | GCG | 1 | Ala | Ser | 0.4444 |
| ENSECAG00000022378 | ENSECAT00000023989 | *IGSF3* | + | 5 | 52444326 | G | T | GCG | 1 | Ala | Ser | 0.4444 |
| ENSECAG00000012533 | ENSECAT00000013153 | *RSPH4A* | + | 10 | 65262821 | A | G | GAA | 2 | Glu | Gly | 0.4441 |
| ENSECAG00000006841 | ENSECAT00000006820 |  | - | 1 | 155745453 | A | G | CAA | 2 | Leu | Ser | 0.4437 |
| ENSECAG00000004238 | ENSECAT00000004178 |  | - | 1 | 155180795 | G | A | GGG | 2 | Pro | Leu | 0.4412 |
| ENSECAG00000008618 | ENSECAT00000008785 | *LCA5L* | - | 26 | 35417731 | A | C | AAC | 2 | Val | Gly | 0.4412 |
| ENSECAG00000001007 | ENSECAT00000000986 |  | - | 6 | 25639957 | A | G | CAT | 2 | Met | Thr | 0.4410 |
| ENSECAG00000001007 | ENSECAT00000000986 |  | - | 6 | 25640327 | T | C | GAT | 3 | Ile | Val | 0.4375 |
| ENSECAG00000000355 | ENSECAT00000000276 |  | - | 21 | 4116690 | C | A | CTT | 1 | Lys | Asn | 0.4375 |
| ENSECAG00000001481 | ENSECAT00000001358 | *SAMD9L* | - | 4 | 36789902 | A | T | TAC | 2 | Val | Glu | 0.4375 |
| ENSECAG00000022630 | ENSECAT00000025166 |  | - | 13 | 7486258 | A | G | AGA | 3 | Ser | Pro | 0.4353 |
| ENSECAG00000011433 | ENSECAT00000011865 | *TET2* | - | 2 | 119545075 | G | C | GGC | 2 | Ala | Gly | 0.4324 |
| ENSECAG00000003379 | ENSECAT00000003234 |  | + | 5 | 38662682 | G | T | GCA | 1 | Ala | Ser | 0.4324 |
| ENSECAG00000025154 | ENSECAT00000027160 | *ADPRM* | + | 11 | 53164385 | A | G | ACA | 1 | Thr | Ala | 0.4271 |
| ENSECAG00000020294 | ENSECAT00000021523 | *DCLK3* | + | 16 | 48308428 | C | T | ACG | 2 | Thr | Met | 0.4236 |
| ENSECAG00000024761 | ENSECAT00000026831 | *INOS* | - | 11 | 41896696 | A | G | ACA | 3 | Cys | Arg | 0.4228 |
| ENSECAG00000024761 | ENSECAT00000026843 | *INOS* | - | 11 | 41896696 | A | G | ACA | 3 | Cys | Arg | 0.4228 |
| ENSECAG00000019438 | ENSECAT00000020602 | *OSMR* | - | 21 | 26835081 | T | C | GTA | 2 | Tyr | Cys | 0.4211 |
| ENSECAG00000006186 | ENSECAT00000006135 |  | + | 12 | 18063623 | T | C | CTG | 2 | Leu | Pro | 0.4191 |
| ENSECAG00000005195 | ENSECAT00000005083 |  | + | 17 | 25869085 | C | A | CCT | 1 | Pro | Thr | 0.4184 |
| ENSECAG00000003607 | ENSECAT00000003475 | *OR6K2* | + | 5 | 38692038 | C | T | CCC | 1 | Pro | Ser | 0.4184 |
| ENSECAG00000026868 | ENSECAT00000028994 | *C6orf183* | + | 10 | 58985128 | G | A | GTC | 1 | Val | Ile | 0.4167 |

## Table S4.3. Non-synonymous markers specific to Yakutian horses, ordered by specificity (3).

| Gene ID | Transcript ID | Gene Name | Strand | Chr | Pos | NucRef | NucAlt | Codon | CodonPos | AARef | AAAlt | Delta |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ENSECAG00000013249 | ENSECAT00000013875 | *UTP6* | - | 12 | 16368660 | C | T | GCG | 2 | Arg | His | 0.4167 |
| ENSECAG00000013249 | ENSECAT00000013907 |  | - | 12 | 16368660 | C | T | GCG | 2 | Arg | His | 0.4167 |
| ENSECAG00000012220 | ENSECAT00000013076 |  | + | 18 | 59611848 | A | C | ACT | 1 | Thr | Pro | 0.4167 |
| ENSECAG00000012220 | ENSECAT00000013233 |  | + | 18 | 59611848 | A | C | ACT | 1 | Thr | Pro | 0.4167 |
| ENSECAG00000017484 | ENSECAT00000019252 | *SSFA2* | + | 22 | 19469586 | G | C | AGT | 2 | Ser | Thr | 0.4167 |
| ENSECAG00000017484 | ENSECAT00000019406 | *SSFA2* | + | 22 | 19469586 | G | C | AGT | 2 | Ser | Thr | 0.4167 |
| ENSECAG00000018766 | ENSECAT00000019943 | *C20orf194* | - | 31 | 16485432 | G | C | GTG | 1 | His | Gln | 0.4167 |
| ENSECAG00000001481 | ENSECAT00000001358 | *C20orf194* | - | 4 | 36788507 | C | T | ACG | 2 | Arg | His | 0.4167 |
| ENSECAG00000021886 | ENSECAT00000024105 | *IYD* | + | 5 | 38741054 | G | A | CGT | 2 | Arg | His | 0.4167 |
| ENSECAG00000021886 | ENSECAT00000024105 | *SAMD9L* | + | 5 | 38755214 | C | T | CGC | 1 | Arg | Cys | 0.4167 |
| ENSECAG00000004825 | ENSECAT00000004694 | *SPTA1* | + | 6 | 72700073 | T | G | TTG | 1 | Leu | Val | 0.4167 |
| ENSECAG00000014894 | ENSECAT00000015620 | *SPTA1* | + | 21 | 51321926 | A | G | CAT | 2 | His | Arg | 0.415 |
| ENSECAG00000022607 | ENSECAT00000024108 |  | - | 21 | 25316610 | T | C | TTT | 3 | Lys | Glu | 0.415 |
| ENSECAG00000005754 | ENSECAT00000005662 | *FASTKD3* | + | 10 | 12690122 | G | C | GGG | 2 | Gly | Ala | 0.4147 |
| ENSECAG00000013392 | ENSECAT00000014067 |  | - | 13 | 27721333 | G | C | CGG | 3 | Pro | Ala | 0.4132 |
| ENSECAG00000014894 | ENSECAT00000015620 |  | + | 21 | 51321925 | C | T | CAT | 1 | His | Tyr | 0.4132 |
| ENSECAG00000000971 | ENSECAT00000002158 | *TMC5* | + | 5 | 8511547 | T | C | TCT | 1 | Ser | Pro | 0.4132 |
| ENSECAG00000010114 | ENSECAT00000010916 | *FASTKD3* | + | 1 | 149766002 | G | A | AGC | 2 | Ser | Asn | 0.4125 |
| ENSECAG00000020841 | ENSECAT00000022128 | *SUCO* | - | 16 | 3868218 | C | T | GAC | 3 | Val | Ile | 0.4125 |
| ENSECAG00000015268 | ENSECAT00000016066 | *RASGRP1* | - | 10 | 50124737 | C | T | AAC | 3 | Val | Ile | 0.4118 |
| ENSECAG00000019873 | ENSECAT00000021083 | *MRPS25* | - | 14 | 82927811 | T | C | TAT | 3 | Ile | Val | 0.4118 |
| ENSECAG00000024820 | ENSECAT00000026813 | *FBXL4* | + | 22 | 26511151 | C | A | ACA | 2 | Thr | Lys | 0.4118 |
| ENSECAG00000004038 | ENSECAT00000003883 |  | - | 6 | 39182395 | C | T | TCC | 2 | Gly | Glu | 0.4118 |
| ENSECAG00000004038 | ENSECAT00000003883 | *PHF20* | - | 6 | 39182421 | A | T | ACT | 1 | Ser | Arg | 0.4118 |
| ENSECAG00000015695 | ENSECAT00000017333 |  | + | 6 | 42936427 | C | T | ACA | 2 | Thr | Ile | 0.4111 |
| ENSECAG00000004053 | ENSECAT00000003928 |  | - | 7 | 74793928 | C | T | ACC | 2 | Gly | Asp | 0.4097 |
| ENSECAG00000020783 | ENSECAT00000022047 | *PTPRO* | + | 15 | 61772275 | T | G | TTC | 2 | Phe | Cys | 0.4097 |
| ENSECAG00000008213 | ENSECAT00000008276 |  | - | 4 | 96558253 | T | C | TTG | 2 | Gln | Arg | 0.4097 |
| ENSECAG00000006888 | ENSECAT00000006857 |  | + | 1 | 157012356 | A | G | CAG | 2 | Gln | Arg | 0.4085 |
| ENSECAG00000022192 | ENSECAT00000023632 |  | + | 7 | 1569468 | T | C | ATG | 2 | Met | Thr | 0.4085 |
| ENSECAG00000001095 | ENSECAT00000000914 |  | + | 1 | 48293522 | A | C | GAA | 3 | Glu | Asp | 0.4081 |
| ENSECAG00000000536 | ENSECAT00000001244 | *ZNF77* | - | 1 | 12639680 | C | G | AAC | 3 | Val | Leu | 0.4062 |
| ENSECAG00000003940 | ENSECAT00000003793 |  | - | 11 | 38231191 | C | T | AGC | 3 | Ala | Thr | 0.4062 |
| ENSECAG00000013249 | ENSECAT00000013875 | *SEC23IP* | - | 12 | 16356917 | A | G | AAA | 3 | Phe | Leu | 0.4062 |
| ENSECAG00000013249 | ENSECAT00000013907 |  | - | 12 | 16356917 | A | G | AAA | 3 | Phe | Leu | 0.4062 |
| ENSECAG00000001046 | ENSECAT00000000903 |  | - | 14 | 92773688 | T | C | GTG | 2 | His | Arg | 0.4062 |
| ENSECAG00000021057 | ENSECAT00000023148 |  | - | 14 | 77154512 | A | G | AAC | 2 | Val | Ala | 0.4062 |
| ENSECAG00000008471 | ENSECAT00000008957 |  | - | 18 | 53252113 | A | C | GGA | 3 | Ser | Ala | 0.4062 |
| ENSECAG00000019644 | ENSECAT00000020802 | *GPR98* | - | 26 | 30158575 | C | G | CCT | 1 | Arg | Ser | 0.4062 |
| ENSECAG00000022735 | ENSECAT00000024598 | *GPR155* | + | 26 | 32662948 | A | G | CAG | 2 | Gln | Arg | 0.4062 |
| ENSECAG00000022735 | ENSECAT00000024602 | *DNAJC28* | + | 26 | 32662948 | A | G | CAG | 2 | Gln | Arg | 0.4062 |
| ENSECAG00000000765 | ENSECAT00000000617 | *DOPEY2* | + | 28 | 35568193 | T | A | CTA | 2 | Leu | Gln | 0.4062 |
| ENSECAG00000018887 | ENSECAT00000020177 | *DOPEY2* | - | 20 | 33419570 | G | T | GGC | 2 | Ala | Asp | 0.4059 |
| ENSECAG00000014933 | ENSECAT00000015588 |  | + | 4 | 95785514 | A | C | AGT | 1 | Ser | Arg | 0.4056 |
| ENSECAG00000014933 | ENSECAT00000015588 | *PSMB8* | + | 4 | 95785494 | T | G | ATG | 2 | Met | Arg | 0.4052 |
| ENSECAG00000024156 | ENSECAT00000026060 |  | + | 5 | 51156271 | A | G | AAA | 1 | Lys | Glu | 0.4052 |
| ENSECAG00000015712 | ENSECAT00000016565 |  | + | 7 | 74563791 | C | T | TCT | 2 | Ser | Phe | 0.4029 |
| ENSECAG00000022171 | ENSECAT00000023882 | *SPAG17* | - | 4 | 14999961 | C | T | CTC | 3 | Glu | Lys | 0.4028 |
| ENSECAG00000003379 | ENSECAT00000003234 |  | + | 5 | 38662935 | C | T | GCT | 2 | Ala | Val | 0.4028 |
| ENSECAG00000004790 | ENSECAT00000004685 | *GCK* | + | 5 | 37503723 | G | T | TTG | 3 | Leu | Phe | 0.4020 |
| ENSECAG00000020525 | ENSECAT00000021931 |  | + | 1 | 60935252 | T | C | ATT | 2 | Ile | Thr | 0.4000 |
| ENSECAG00000007563 | ENSECAT00000008109 | *PIGM* | + | 11 | 6361518 | C | T | GCG | 2 | Ala | Val | 0.4000 |
| ENSECAG00000007563 | ENSECAT00000008256 | *NUDT13* | + | 11 | 6361518 | C | T | GCG | 2 | Ala | Val | 0.4000 |
| ENSECAG00000007563 | ENSECAT00000008344 | *RECQL5* | + | 11 | 6361518 | C | T | GCG | 2 | Ala | Val | 0.4000 |
| ENSECAG00000012813 | ENSECAT00000013374 | *RECQL5* | - | 16 | 4787031 | G | A | GGG | 2 | Pro | Leu | 0.4000 |

*Codon on the on the negative strand (-) should be reverse complemented. “Chr#”: chromosome number; “NucRef”: reference nucleotide; “NucAlt”: alternative nucleotide; “CodonPos”: codon position; “AARef”: reference amino-acid; “AAAlt”: alternative amino acid; “Delta”: difference in frequency of the alternative allele between the two populations, positive when specific to the Yakutian horses.

# Table S4.4. GO-term for genes enriched in Yakutian horse specific non-synonymous mutations.

| GO-term | Model organism | Adjusted p-value | Gene ID | Gene Name |
|---|---|---|---|---|
| Disruption of cells of other organism | Human | 0.0299 | ENSG00000089041 | *P2RX7* |
| | | | ENSG00000007171 | *NOS2* |
| | | | ENSG00000163221 | *S100A12* |
| C-terminal protein lipidation | Human | 0.0299 | ENSG00000143315 | *PIGM* |
| | | | ENSG00000142892 | *PIGK* |
| | | | ENSG00000060642 | *PIGV* |
| Killing of cells of other organism | Human | 0.0299 | ENSG00000089041 | *P2RX7* |
| | | | ENSG00000007171 | *NOS2* |
| | | | ENSG00000163221 | *S100A12* |
| Regulation of killing of cells of other organism | Human | 0.0299 | ENSG00000089041 | *P2RX7* |
| | | | ENSG00000007171 | *NOS2* |
| C-terminal protein amino acid modification | Human | 0.0359 | ENSG00000143315 | *PIGM* |
| | | | ENSG00000142892 | *PIGK* |
| | | | ENSG00000060642 | *PIGV* |
| GPI anchor biosynthetic process | Human | 0.0498 | ENSG00000143315 | *PIGM* |
| | | | ENSG00000142892 | *PIGK* |
| | | | ENSG00000060642 | *PIGV* |
| GPI anchor biosynthetic process | Mouse | 0.0015 | ENSMUSG00000050229 | *PIGM* |
| | | | ENSMUSG00000032059 | *ALG9* |
| | | | ENSMUSG00000043257 | *PIGV* |
| | | | ENSMUSG00000039047 | *PIGK* |
| GPI anchor metabolic process | Mouse | 0.0015 | ENSMUSG00000050229 | *PIGM* |
| | | | ENSMUSG00000032059 | *ALG9* |
| | | | ENSMUSG00000043257 | *PIGV* |
| | | | ENSMUSG00000039047 | *PIGK* |
| Membrane lipid biosynthetic process | Mouse | 0.0015 | ENSMUSG00000050229 | *PIGM* |
| | | | ENSMUSG00000032059 | *ALG9* |
| | | | ENSMUSG00000043257 | *PIGV* |
| | | | ENSMUSG00000039047 | *PIGK* |
| Glycolipid biosynthetic process | Mouse | 0.0025 | ENSMUSG00000050229 | *PIGM* |
| | | | ENSMUSG00000032059 | *ALG9* |
| | | | ENSMUSG00000043257 | *PIGV* |
| | | | ENSMUSG00000039047 | *PIGK* |
| Phosphatidylinositol metabolic process | Mouse | 0.0042 | ENSMUSG00000050229 | *PIGM* |
| | | | ENSMUSG00000032059 | *ALG9* |
| | | | ENSMUSG00000043257 | *PIGV* |
| | | | ENSMUSG00000042105 | *INPP5F* |
| | | | ENSMUSG00000039047 | *PIGK* |
| Phosphatidylinositol biosynthetic process | Mouse | 0.0042 | ENSMUSG00000050229 | *PIGM* |
| | | | ENSMUSG00000032059 | *ALG9* |
| | | | ENSMUSG00000043257 | *PIGV* |
| | | | ENSMUSG00000039047 | *PIGK* |
| Protein lipidation | Mouse | 0.0042 | ENSMUSG00000050229 | *PIGM* |
| | | | ENSMUSG00000032059 | *ALG9* |
| | | | ENSMUSG00000043257 | *PIGV* |
| | | | ENSMUSG00000039047 | *PIGK* |
| Glycolipid metabolic process | Mouse | 0.0048 | ENSMUSG00000050229 | *PIGM* |
| | | | ENSMUSG00000032059 | *ALG9* |
| | | | ENSMUSG00000043257 | *PIGV* |
| | | | ENSMUSG00000039047 | *PIGK* |
| Lipoprotein biosynthetic process | Mouse | 0.0048 | ENSMUSG00000050229 | *PIGM* |
| | | | ENSMUSG00000032059 | *ALG9* |
| | | | ENSMUSG00000043257 | *PIGV* |
| | | | ENSMUSG00000039047 | *PIGK* |
| Membrane lipid metabolic process | Mouse | 0.0048 | ENSMUSG00000029468 | *P2RX7* |
| | | | ENSMUSG00000050229 | *PIGM* |
| | | | ENSMUSG00000032059 | *ALG9* |
| | | | ENSMUSG00000043257 | *PIGV* |
| | | | ENSMUSG00000039047 | *PIGK* |
| Mannosyltransferase activity | Mouse | 0.0024 | ENSMUSG00000050229 | *PIGM* |
| | | | ENSMUSG00000032059 | *ALG9* |

## Table S4.5. KEGG pathways for genes enriched in Yakutian horse specific non-synonymous mutations.

| KEGG Pathway | Model organism | Adjusted p-value | Gene ID | Gene Name | Model organism | Adjusted p-value | Gene ID | Gene Name |
|---|---|---|---|---|---|---|---|---|
| Olfactory transduction | Human | <0.0001 | ENSG00000196171 | *OR6K2* | Mouse | <0.0001 | ENSMUSG00000070417 | *OLFR2* |
| | | | ENSG00000176925 | *OR51F2* | | | ENSMUSG00000054526 | *OLFR1500* |
| | | | ENSG00000186943 | *OR13C8* | | | ENSMUSG00000073898 | *OLFR713* |
| | | | ENSG00000177489 | *OR2G2* | | | ENSMUSG00000055033 | *OLFR420* |
| | | | ENSG00000178586 | *OR6B3* | | | ENSMUSG00000093804 | *OLFR1303* |
| | | | ENSG00000166363 | *OR10A5* | | | ENSMUSG00000073965 | *OLFR568* |
| | | | ENSG00000186509 | *OR9Q1* | | | ENSMUSG00000051593 | *OLFR272* |
| | | | ENSG00000184933 | *OR6A2* | | | | |
| Glycosylphosphatidylinositol(GPI)-anchor biosynthesis | Human | 0.0003 | ENSG00000143315 | *PIGM* | Mouse | 0.0001 | ENSMUSG00000050229 | *PIGM* |
| | | | ENSG00000142892 | *PIGK* | | | ENSMUSG00000043257 | *PIGV* |
| | | | ENSG00000060642 | *PIGV* | | | ENSMUSG00000039047 | *PIGK* |
| Metabolic pathways | Human | 0.0027 | ENSG00000143315 | *PIGM* | Mouse | 0.0001 | ENSMUSG00000068587 | *MGAM* |
| | | | ENSG00000177084 | *POLE* | | | ENSMUSG00000020826 | *NOS2* |
| | | | ENSG00000106633 | *GCK* | | | ENSMUSG00000007080 | *POLE* |
| | | | ENSG00000142892 | *PIGK* | | | ENSMUSG00000025519 | *TKTL2* |
| | | | ENSG00000021461 | *CYP3A43* | | | ENSMUSG00000034579 | *PLA2G3* |
| | | | ENSG00000257335 | *MGAM* | | | ENSMUSG00000032059 | *ALG9* |
| | | | ENSG00000151005 | *TKTL2* | | | ENSMUSG00000054417 | *CYP3A44* |
| | | | ENSG00000007171 | *NOS2* | | | ENSMUSG00000041798 | *GCK* |
| | | | ENSG00000100078 | *PLA2G3* | | | ENSMUSG00000039047 | *PIGK* |
| | | | ENSG00000060642 | *PIGV* | | | ENSMUSG00000050229 | *PIGM* |
| | | | | | | | ENSMUSG00000043257 | *PIGV* |
| Galactose metabolism | Human | 0.0095 | ENSG00000257335 | *MGAM* | Mouse | 0.0053 | ENSMUSG00000068587 | *MGAM* |
| | | | ENSG00000106633 | *GCK* | | | ENSMUSG00000041798 | *GCK* |
| Linoleic acid metabolism | Human | 0.0096 | ENSG00000021461 | *CYP3A43* | Mouse | 0.0060 | ENSMUSG00000034579 | *PLA2G3* |
| | | | ENSG00000100078 | *PLA2G3* | | | ENSMUSG00000054417 | *CYP3A44* |
| DNA replication | Human | 0.0113 | ENSG00000035928 | *RFC1* | Mouse | 0.0053 | ENSMUSG00000007080 | *POLE* |
| | | | ENSG00000177084 | *POLE* | | | ENSMUSG00000029191 | *RFC1* |
| Toxoplasmosis | | 0.0117 | ENSG00000196569 | *LAMA2* | Mouse | 0.0053 | ENSMUSG00000020826 | *NOS2* |
| | | | ENSG00000007171 | *NOS2* | | | ENSMUSG00000034579 | *PLA2G3* |
| | | | ENSG00000100078 | *PLA2G3* | | | ENSMUSG00000019899 | *LAMA2* |
| Nucleotide excision repair | Human | 0.0127 | ENSG00000035928 | *RFC1* | Mouse | 0.0060 | ENSMUSG00000007080 | *POLE* |
| | | | ENSG00000177084 | *POLE* | | | ENSMUSG00000029191 | *RFC1* |
| Starch and sucrose metabolism | Human | 0.0169 | ENSG00000257335 | *MGAM* | Mouse | 0.0060 | ENSMUSG00000068587 | *MGAM* |
| | | | ENSG00000106633 | *GCK* | | | ENSMUSG00000041798 | *GCK* |
| Calcium signaling pathway | Human | 0.0182 | ENSG00000089041 | *P2RX7* | Mouse | 0.0068 | ENSMUSG00000029468 | *P2RX7* |
| | | | ENSG00000170425 | *ADORA2B* | | | ENSMUSG00000020826 | *NOS2* |
| | | | ENSG00000007171 | *NOS2* | | | ENSMUSG00000018500 | *ADORA2B* |

## Table S4.6. Wikipathways for genes enriched in Yakutian horse specific non-synonymous mutations.

| Wikipathway | Model organism | Adjusted p-value | Gene ID | Gene Name |
|---|---|---|---|---|
| DNA Replication | Human | 0.0094 | ENSG00000035928 | *RFC1* |
| | | | ENSG00000177084 | *POLE* |
| GPCRs, Class A Rhodopsin-like | Human | 0.0094 | ENSG00000166363 | *OR10A5* |
| | | | ENSG00000170425 | *ADORA2B* |
| | | | ENSG00000183671 | *GPR1* |
| | | | ENSG00000184933 | *OR6A2* |
| GPCRs, Other | Human | 0.0339 | ENSG00000166363 | *OR10A5* |
| | | | ENSG00000164199 | *GPR98* |
| DNA Replication | Human | 0.0107 | ENSMUSG00000007080 | *POLE* |
| | | | ENSMUSG00000029191 | *RFC1* |
| Odorant GPCRs | Mouse | 0.0107 | ENSMUSG00000056115 | *TAS2R134* |
| | | | ENSMUSG00000041762 | *GPR155* |
| | | | ENSMUSG00000046856 | *GPR1* |
| GPCRs, Class A Rhodopsin-like | Mouse | 0.0108 | ENSMUSG00000070417 | *OLFR2* |
| | | | ENSMUSG00000073898 | *OLFR713* |
| | | | ENSMUSG00000018500 | *ADORA2B* |
| MAPK signaling pathway | Mouse | 0.0394 | ENSMUSG00000041798 | *GCK* |
| | | | ENSMUSG00000027347 | *RASGRP1* |

## Table S4.7. Phenotypes for genes enriched in Yakutian horse specific non-synonymous mutations.

| Phenotype | Model organism | Adjusted p-value | Gene ID | Gene Name |
|---|---|---|---|---|
| Abnormal bone healing | Mouse | 0.0312 | ENSMUSG00000020826 | *NOS2* |
| | | | ENSMUSG00000040297 | *SUCO* |
| Increased T cell number | Mouse | 0.0312 | ENSMUSG00000042351 | *GRAP2* |
| | | | ENSMUSG00000020826 | *NOS2* |
| | | | ENSMUSG00000040943 | *TET2* |
| | | | ENSMUSG00000026532 | *SPNA1* |
| | | | ENSMUSG00000027347 | *RASGRP1* |
| | | | ENSMUSG00000001123 | *LGALS9* |
| Increased circulating prolactin level | Mouse | 0.0312 | ENSMUSG00000020826 | *NOS2* |
| | | | ENSMUSG00000019899 | *LAMA2* |
| Decreased hematopoietic cell number | Mouse | 0.0312 | ENSMUSG00000042351 | *GRAP2* |
| | | | ENSMUSG00000020826 | *NOS2* |
| | | | ENSMUSG00000022146 | *OSMR* |
| | | | ENSMUSG00000019899 | *LAMA2* |
| | | | ENSMUSG00000001123 | *LGALS9* |
| | | | ENSMUSG00000040943 | *TET2* |
| | | | ENSMUSG00000086564 | *CD101* |
| | | | ENSMUSG00000040297 | *SUCO* |
| | | | ENSMUSG00000026532 | *SPNA1* |
| | | | ENSMUSG00000027347 | *RASGRP1* |
| Increased prolactin level | Mouse | 0.0333 | ENSMUSG00000020826 | *NOS2* |
| | | | ENSMUSG00000019899 | *LAMA2* |
| Abnormal hematopoietic cell number | Mouse | 0.0347 | ENSMUSG00000029468 | *P2RX7* |
| | | | ENSMUSG00000042351 | *GRAP2* |
| | | | ENSMUSG00000020826 | *NOS2* |
| | | | ENSMUSG00000022146 | *OSMR* |
| | | | ENSMUSG00000019899 | *LAMA2* |
| | | | ENSMUSG00000001123 | *LGALS9* |
| | | | ENSMUSG00000040943 | *TET2* |
| | | | ENSMUSG00000086564 | *CD101* |
| | | | ENSMUSG00000040297 | *SUCO* |
| | | | ENSMUSG00000026532 | *SPNA1* |
| | | | ENSMUSG00000027347 | *RASGRP1* |
| Abnormal leukocyte cell number | Mouse | 0.0357 | ENSMUSG00000029468 | *P2RX7* |
| | | | ENSMUSG00000042351 | *GRAP2* |
| | | | ENSMUSG00000020826 | *NOS2* |
| | | | ENSMUSG00000019899 | *LAMA2* |
| | | | ENSMUSG00000001123 | *LGALS9* |
| | | | ENSMUSG00000040943 | *TET2* |
| | | | ENSMUSG00000086564 | *CD101* |
| | | | ENSMUSG00000040297 | *SUCO* |
| | | | ENSMUSG00000026532 | *SPNA1* |
| | | | ENSMUSG00000027347 | *RASGRP1* |

## Table S4.8. Diseases for genes enriched in Yakutian horse specific non-synonymous mutations.

| Disease | Model organism | Adjusted p-value | Gene ID | Gene Name |
|---|---|---:|---|---|
| Immune System Diseases | Human | 0.0122 | ENSG00000204264 | *PSMB8* |
| | | | ENSG00000132256 | *TRIM5* |
| | | | ENSG00000183671 | *GPR1* |
| | | | ENSG00000089041 | *P2RX7* |
| | | | ENSG00000035928 | *RFC1* |
| | | | ENSG00000007171 | *NOS2* |
| | | | ENSG00000172575 | *RASGRP1* |
| | | | ENSG00000163221 | *S100A12* |
| Leukemia | Human | 0.0122 | ENSG00000145623 | *OSMR* |
| | | | ENSG00000089041 | *P2RX7* |
| | | | ENSG00000168769 | *TET2* |
| | | | ENSG00000151490 | *PTPRO* |
| | | | ENSG00000035928 | *RFC1* |
| | | | ENSG00000132704 | *FCRL2* |
| | | | ENSG00000132256 | *TRIM5* |
| Lymphatic Diseases | Human | 0.0122 | ENSG00000089041 | *P2RX7* |
| | | | ENSG00000151490 | *PTPRO* |
| | | | ENSG00000035928 | *RFC1* |
| | | | ENSG00000132704 | *FCRL2* |
| | | | ENSG00000135604 | *STX11* |
| | | | ENSG00000172575 | *RASGRP1* |
| Inflammation | Human | 0.0214 | ENSG00000145623 | *OSMR* |
| | | | ENSG00000089041 | *P2RX7* |
| | | | ENSG00000170425 | *ADORA2B* |
| | | | ENSG00000007171 | *NOS2* |
| | | | ENSG00000163221 | *S100A12* |
| | | | ENSG00000163736 | *PPBP* |
| Histiocytosis, Langerhans-Cell | Human | 0.0227 | ENSG00000134256 | *CD101* |
| | | | ENSG00000135604 | *STX11* |
| Thrombocytosis | Human | 0.0227 | ENSG00000088854 | *C20orf194* |
| | | | ENSG00000168769 | *TET2* |
| Nelson syndrome | Human | 0.0227 | ENSG00000021461 | *CYP3A43* |
| | | | ENSG00000257335 | *MGAM* |
| | | | ENSG00000138434 | *SSFA2* |
| | | | ENSG00000128519 | *TAS2R16* |
| | | | ENSG00000163328 | *GPR155* |
| | | | ENSG00000183671 | *GPR1* |
| | | | ENSG00000163736 | *PPBP* |
| Trisomy | Human | 0.0235 | ENSG00000213626 | *LBH* |
| | | | ENSG00000142197 | *DOPEY2* |
| | | | ENSG00000035928 | *RFC1* |
| Leishmaniasis | Human | 0.0235 | ENSG00000171916 | *LGALS9C* |
| | | | ENSG00000007171 | *NOS2* |
| Bronchitis | Human | 0.0235 | ENSG00000204264 | *PSMB8* |
| | | | ENSG00000111834 | *RSPH4A* |
| | | | ENSG00000007171 | *NOS2* |
| | | | ENSG00000163736 | *PPBP* |

## Table S4.9. Loci underlying Mendelian traits in horses.

| Chrom. | Coordinate | Gene | Phenotype | Reference |
|---|---|---|---|---|
| 1 | 74,842,283 | *ACTN2* | Racing performance | Hill et al. 2010 (50) |
| 1 | 108,249,293 | *TRPM1* | Leopard complex spotting and cation channel congenital stationary night blindness | Bellone et al. 2010 (51) |
| 1 | 128,056,148 | *PPIB* | Hereditary equine dermal isomerase B asthenia | Tryon et al. 2007 (52) |
| 1 | 138,235,715 | *MYO5A* | Lavender foal syndrome | Brooks et al. 2010 (53) |
| 2 | 13,074,277 | *TOE1* | Cerebellar abiotrophy | Brault et al. 2011 (54) |
| 3 | 32,772,871 | *COX4/1* | Racing performance | Hill et al. 2010 (50) |
| 3 | 36,259,552 | *MC1R (1)* | Chestnut coat color | Marklund et al. 1996 (55) |
| 3 | 36,259,554 | *MC1R (2)* | Chestnut coat color | Wagner et al. 2000 (56) |
| 3 | 77,735,520 | *KIT (1)* | Sabino spotting | Brooks et al. 2005 (57) |
| 3 | 77,740,161 | *KIT (2)* | Tobiano spotting | Brooks et al. 2002 (58) |
| 3 | 105,547,002 | *LCORL/NCAPG* | Larger body size | Makvandi-Nejad et al. 2012 (59), Signer-Hasler et al. 2012 (40) |
| 4 | 38,697,145 | *PON1* | Racing performance | Hill et al. 2010 (50) |
| 4 | 38,969,307 | *PDK4 (1)* | Racing performance | Hill et al. 2010 (50) |
| 4 | 38,973,231 | *PDK4 (2)* | Racing performance | Hill et al. 2010 (50) |
| 4 | 40,279,726 | *ACN9* | Racing performance | Hill et al. 2010 (50) |
| 4 | 96,375,588 | *CLCN1* | Congenital myotonia | Wijnberg et al. 2012 (60) |
| 5 | 20,256,789 | *LAMC2* | Junctional epidermolysis bullosa | Spirito et al. 2002 (61) |
| 6 | 11,429,753 | *PAX3* | Splashed white coat | Hauswirth et al. 2012, Hauswirth et al. 2013 (62, 63) |
| 6 | 73,665,304 | *PMEL17* (also known as *SIL11*) | Silver coat color | Brunberg et al. 2006 (64) |
| 6 | 81,481,065 | *HMGA2* | Larger body size | Makvandi-Nejad et al. 2012 (59) |
| 9 | 35,528,429 | *DNAPK* | Severe combined immunodeficiency | Shin et al. 1997 (65) |
| 9 | 74,795,013 | *ZFAT (1)* | Wither height | Signer-Hasler et al. 2012 (40) |
| 9 | 74,795,089 | *ZFAT (2)* | Wither height | Signer-Hasler et al. 2012 (40) |
| 9 | 74,795,236 | *ZFAT (3)* | Wither height | Signer-Hasler et al. 2012 (40) |
| 9 | 74,798,143 | *ZFAT (4)* | Wither height | Signer-Hasler et al. 2012 (40) |
| 9 | 75,550,059 | *ZFAT (5)* | Larger body size | Makvandi-Nejad et al. 2012 (59) |
| 10 | 9,554,699 | *RYR1* | Malignant hyperthermia | Aleman et al. 2004 (66) |
| 10 | 15,884,567 | *CKM* | Racing performance | Gu et al. 2010 (67) |
| 10 | 18,940,324 | *GYS1* | Polysaccharide storage myopathy | McCue et al. 2012 (68) |
| 11 | 15,500,439 | *SCN4A* | Equine hyperkalemic periodic paralysis | Cannon et al. 1995 (69) |
| 11 | 19,184,674 | *ITGA2B* | Glanzmann Thrombasthenia | Christopherson et al. 2007 (70) |
| 11 | 23,259,732 | *LASP1* | Larger body size | Makvandi-Nejad et al. 2012 (59) |
| 14 | 3,761,254 | *PROP1 (1)* | Dwarfism | Orr et al. 2010 (71) |
| 14 | 3,761,355 | *PROP1 (2)* | Dwarfism | Orr et al. 2010 (71) |
| 14 | 5,418,619 | Intergenic | Dwarfism | Orr et al. 2010 (71) |
| 14 | 26,701,092 | *SLC36A1* | Champagne dilution | Cook et al. 2008 (72) |
| 14 | 27,991,841 | *SLC26A2* | Autosomal recessively inherited chondrodysplasia | Hansen et al. 2007 (73) |
| 16 | 20,103,081 | *MITF (1)* | Macchiato, hearing loss | Hauswirth et al. 2012, Hauswirth et al. 2013 (62, 63) |
| 16 | 20,105,348 | *MITF (2)* | Splashed white coat | Hauswirth et al. 2012, Hauswirth et al. 2013 (62, 63) |
| 16 | 20,117,302 | *MITF (3)* | Splashed white coat | Hauswirth et al. 2012, Hauswirth et al. 2013 (62, 63) |
| 17 | 50,624,658 | *EDNRB* | Lethal white foal syndrome | Yang et al. 1998 (74) |
| 18 | 66,493,737 | *MSTN* | Optimum racing position | Tozaki et al. 2010, Hill et al. 2012 (75, 76) |
| 21 | 30,666,626 | *SLC45A2* (also known as *MATP*) | Cream coat color | Mariat et al. 2003 (77) |
| 22 | 22,684,390 | *COX4/2* | Racing performance | Gu et al. 2010 (67) |
| 22 | 25,168,579 | *ASIP* | Black and bay color | Rieder et al. 2001 (78) |
| 23 | 22,999,655 | *DMRT3* | Pattern of locomotion (altered gait) | Andersson et al. 2012 (79) |
| 26 | 30,660,224 | *SLC5A3* | Foal immunodeficiency syndrome | Fox-Clipsham et al. 2011 (80) |
| X | 49,635,250 | *AR* | Androgen insensitivity syndrome (AIS) | Revay et al. 2012 (81) |
| X | 122,833,887 | *IKBKG* | Incontinentia pigmenti | Towers et al. 2013 (82) |

## Table S4.10. Mendelian variant genotypes in Late Pleistocene horses, ancient and present-day horses from Yakutia.

| Gene | Mutation | 22 | 23 | Batagai | 37 | Yak1 | Yak2 | Yak3 | Yak4 | Yak5 | Yak6 | Yak7 | Yak8 | Yak9 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *ACTN2* | A>G | . | A4 | . | A/G | . | A/G | . | G/G | G/G | G/G | A/G | A4,G1 | A6 |
| *TRPM1* | C>T | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *PPIB* | G>A | . | . | . | . | . | . | G4 | G5 | G6 | . | . | . | . |
| *MYO5A* | 1 DEL | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *TOE1* | C>T | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *COX4/1* | C>T | . | C1,T4 | C/T | T/T | T/T | T/T | T/T | C/T | C/T | T/T | . | C/T | T/T |
| *MC1R (1)* | C>T | . | . | . | . | T/T | C/T | T/T | . | . | T/T | C/T | T/T | . |
| *MC1R (2)* | G>A | . | . | . | . | A/G | . | A/G | . | . | . | . | . | . |
| *KIT (1)* | A>T | . | A1 | . | A6 | . | . | . | A4 | . | . | . | . | A7 |
| *KIT (2)* | G>C | . | . | . | . | . | . | . | . | . | . | . | G6 | . |
| *LCORL / NCAPG* | T>C | . | T7 | . | . | . | C4,T14 | . | T7 | . | . | . | . | . |
| *PON1* | C>T | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *PDK4 (1)* | C>A | . | C5 | . | A/C | . | . | . | A/C | . | . | . | . | . |
| *PDK4 (2)* | G>A | . | G3 | G5 | . | . | . | . | . | . | . | . | . | . |
| *ACN9* | C>T | C/T | C/T | C/T | C/T | T/T | . | T/T | C/T | . | . | C/T | C/T | C11,T11 |
| *CLCN1* | A>C | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *LAMC2* | 1C INS | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *PAX3* | C>T | . | . | . | . | C3 | . | C2 | . | . | . | . | . | . |
| *PMEL17* | G>A | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *HMGA2* | C>T | C2 | C4 | T/T | T3 | T7 | T/T | T/T | C/T | T/T | T6 | T/T | T3 | C/T |
| *DNAPK* | 5 DEL | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *ZFAT (1)* | C>T | C/T | C1,T1 | . | C/T | C4,T2 | . | T3 | C/T | C/T | C/T | . | . | T/T |
| *ZFAT (2)* | C>A | A/C | C1 | . | A/C | A/C | . | A4 | A/C | A/C | A/C | . | . | A/A |
| *ZFAT (3)* | G>A | A/G | A1,G4 | . | A4,G23 | A1,C2,G2 | . | A4,C1 | A/G | A2,G11 | A/G | . | . | A6 |
| *ZFAT (4)* | G>A | A/G | A/G | . | A/G | A2,G1 | . | A4 | A2,G7 | A/G | A/G | . | . | A/A |
| *ZFAT (5)* | C>T | . | . | . | . | C/T | C/T | C/T | C/T | . | . | . | C/T | . |
| *RYR1* | C>G | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *CKM* | G>A | . | A/G | A/A | A/G | G7 | . | A3,G1 | A/G | A/G | A/A | A/G | G6 | G4 |
| *GYS1* | C>T | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *SCN4A* | C>T | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *ITGA2B* | 10 DEL | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *LASP1* | G>A | . | G4 | . | A/G | A3 | G3 | A1,G2 | G1 | G1 | . | A4,G3 | G3 | G1 |
| *PROP1 (1)* | G>C | C/G | C4,G2 | C/C | C/G | C5 | C/G | . | C/G | C/G | C/C | C/G | . | . |
| *PROP1 (2)* | T>C | C/C | C4,T2 | C/C | C/T | C/C | C/T | . | . | C/T | C/T | . | . | . |
| Intergenic | G>A | . | . | . | . | A4,G2 | A/G | G8,T1 | G3 | . | . | . | . | . |
| *SLC36A1* | G>C | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *SLC26A2* | G>A | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *MITF (1)* | T>C | . | . | . | . | . | . | . | . | . | . | . | . | T5 |
| *MITF (2)* | 5 DEL | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *MITF (3)* | 11 DEL | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *EDNRB* | GA>CT | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *MSTN* | T>C | T7 | T2 | . | T2 | . | C/T | . | T4 | . | . | . | . | T7 |
| *SLC45A2* | G>A | . | . | . | . | . | . | . | . | . | . | . | G7 | . |
| *COX4/2* | C>T | . | . | C/T | C/T | . | . | C4 | . | T/T | C/T | . | C/T | T/T |
| *ASIP* | 11 DEL | . | . | . | . | +/? | +/- | +/+ | +/- | . | . | +/+ | . | . |
| *DMRT3* | C>A | . | . | . | . | . | . | C4 | . | . | . | . | . | . |
| *SLC5A3* | C>T | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *AR* | A>G | . | . | . | . | . | . | . | A3 | . | . | . | . | A4 |
| *IKBKG* | C>T | . | . | . | . | . | . | . | . | . | . | . | . | C4 |

“INS”: insertion; “DEL”: bp deletion; “22”: CGG10022; “23”: CGG10023”; “97”: CGG101397. A dot (.) stands for same alleles as the common allele, otherwise alternate allele associated with the Mendelian trait. A positive sign (+) refers to the presence of indels, while a minus sign (-) stands for the absence of indels.

**Table S4.11. Mendelian variant genotypes in Late Pleistocene and domesticated horses.**

| Gene | Mutation | FM0431 | FM0450 | FM0467 | FM1030 | FM1041 | FM1190 | FM1785 | FM1798 | FM1932 | FM1948 | FM1951 | FM2218 | EMS595 | A1543 | A5659 | A5964 | A2085 | M977 | M5256 | M1009 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *ACTN2* | A>G | . | . | A4 | . | . | . | . | . | . | . | . | . | . | . | . | A/G | . | . | . | . |
| *TRPM1* | C>T | C/T | C5 | . | . | C1 | C/T | . | C/T | . | C4 | . | . | C3 | . | . | . | . | . | . | . |
| *PPIB* | G>A | G6 | G4 | . | . | G4 | G3 | G3 | . | . | . | G3 | . | . | G1 | G4 | . | NA | . | G4 | G4 |
| *MYO5A* | 1 DEL | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *TOE1* | C>T | C4 | . | . | . | C2 | . | . | . | . | . | . | C4 | . | . | C4 | . | . | C4 | . | C3 |
| *COX4/1* | C>T | C3,T2 | T/T | T/T | C/T | C1,T3 | C/T | T/T | C/T | C/T | C1,T2 | . | C1 | C/T | C4 | C1,T2 | . | C2,T3 | C/T | T4 | C2 |
| *MC1R (1)* | C>T | C1 | C2 | T/T | C1,T4 | C5,T2 | C/T | C5 | T/T | . | C2,T2 | . | C/T | T2 | T2 | T3 | T/T | C1,T1 | . | . | C1 |
| *MC1R (2)* | G>A | NA | G1 | . | G3 | . | . | G5 | . | . | G4 | . | . | G3 | G3 | G4 | . | G2 | . | . | G2 |
| *KIT (1)* | A>T | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | A3 | . | . | . | A5 |
| *KIT (2)* | G>C | . | . | . | . | G6 | . | . | . | . | . | . | . | . | G3 | . | . | . | . | . | G3 |
| *LCORL / NCAPG* | T>C | . | . | C/T | C/T | . | . | C/T | . | . | . | . | . | . | . | . | . | . | C/T | C/T | . |
| *PON1* | C>T | C4 | . | . | . | C3 | . | . | . | . | . | C4 | . | . | T2 | . | C1 | C/T | C6 | C2 | T2 |
| *PDK4 (1)* | C>A | . | . | A/C | A/C | A2 | A/A | . | A/C | . | . | A/C | . | . | . | A/C | C6 | A/C | . | A2,C4 | . |
| *PDK4 (2)* | G>A | A/G | . | A2,G10 | A/G | A1,G4 | A/A | . | A/A | . | G2 | A/A | . | A/A | A/G | A/G | . | . | . | A/G | . |
| *ACN9* | C>T | . | C/T | T/T | . | C3,T2 | T/T | . | C/T | T6 | . | . | T/T | . | . | T/T | T/T | C/T | C/T | T/T | C/T |
| *CLCN1* | A>C | A6 | . | . | . | . | . | . | . | . | A4 | . | . | . | . | . | . | A4 | . | . | A3 |
| *LAMC2* | 1C INS | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *PAX3* | C>T | C2 | C1 | . | . | C3 | . | C2 | . | . | C2 | . | . | C3 | C2 | . | . | C3 | C4 | C1 | C3 |
| *PMEL17* | G>A | . | . | . | . | G5 | . | . | . | G2 | . | . | . | . | . | . | . | G3 | . | . | G3 |
| *HMGA2* | C>T | T/T | T/T | T/T | C/T | T/T | T/T | T/T | T/T | T/T | . | C/T | C/T | T/T | T/T | T/T | C/T | C/T | T/T | T/T | T5 |
| *DNAPK* | 5 DEL | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *ZFAT (1)* | C>T | C/T | C/T | . | . | T2 | C/T | . | T/T | C2,T1 | C1 | . | . | C/T | C/T | . | C/T | C/T | . | . | C2 |
| *ZFAT (2)* | C>A | A/C | A/C | . | . | A/A | A/C | . | A/A | A1,C3 | A1,C5 | . | . | A/C | A/C | . | A/C | A/C | . | . | C6 |
| *ZFAT (3)* | G>A | A3,G1 | A/G | . | . | A3 | A/G | . | A/A | A4,G2 | A/G | . | . | A/G | A1,G3 | . | A/G | A5 | . | . | . |
| *ZFAT (4)* | G>A | A1,G7 | A2,G2 | . | . | A4 | A/G | . | A/A | A/G | A/G | . | . | A2,G1 | A3,G2 | G1 | A/G | A1,G2 | . | . | . |
| *ZFAT (5)* | C>T | . | . | . | . | . | C/T | C/T | C/T | . | C/T | . | . | . | . | . | . | C2,T1 | . | C/T | T3 |
| *RYR1* | C>G | . | C2 | . | . | . | . | . | . | C4 | . | . | . | C2 | C2 | . | . | NA | . | C3 | C1 |
| *CKM* | G>A | A/G | A/G | . | G3 | G3 | A/G | A/G | . | G4 | A/G | A/G | . | A/A | A6 | A/A | A/G | A1,G2 | A1,G2 | A4 | A1,G2 |
| *GYS1* | C>T | . | . | . | . | C2,T1 | . | . | . | C6 | C4 | C5 | . | . | C3 | . | . | . | . | . | C3 |
| *SCN4A* | C>T | C3 | C5 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C4 |
| *ITGA2B* | 10 DEL | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *LASP1* | G>A | A2 | NA | A2 | A1 | A2 | A/A | G1 | A1 | G1 | NA | A1 | NA | A/A | A3 | G4 | A1,G5 | G1 | G3 | G2 | A1 |
| *PROP1 (1)* | G>C | C/G | G4 | C/G | . | G5 | . | . | C/G | C/C | . | . | C/G | G4 | . | G4 | . | C3,G2 | G5 | . | G2 |
| *PROP1 (2)* | T>C | . | . | C/T | . | T2 | . | . | . | C/C | . | . | C/T | . | T3 | . | . | . | T4 | T1 | T3 |
| Intergenic | G>A | A1,G4 | . | A1,G1 | A/G | A/G | A/G | . | A/G | A1 | G2 | A/G | . | . | . | A1,G1 | A/G | NA | G4 | . | G1 |
| *SLC36A1* | G>C | . | . | . | . | G4 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *SLC26A2* | G>A | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *MITF (1)* | T>C | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | T4 |
| *MITF (2)* | 5 DEL | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *MITF (3)* | 11 DEL | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *EDNRB* | GA>CT | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . |
| *MSTN* | T>C | . | C/T | C/T | C/T | . | . | . | . | C/T | . | . | . | C2,T9 | C/T | C/C | C1 | C7 | . | . | . |
| *SLC45A2* | G>A | . | . | . | . | . | . | . | . | . | G3 | . | . | . | A/A | . | . | G3 | . | . | G4 |
| *COX4/2* | C>T | T5 | C/T | C/T | . | C2,T2 | C/T | . | . | C4 | C3 | C3 | . | C1 | T/T | C2 | T/T | C3,T2 | C/T | C/T | C3 |
| *ASIP* | 11 DEL | . | . | . | +/- | . | +/- | +/- | . | . | . | +/- | . | +/- | +/- | . | +/- | +/- | . | . | . |
| *DMRT3* | C>A | C3 | C5 | . | . | C4 | . | . | . | . | C3 | . | . | C3 | C3 | . | . | NA | A4 | A3 | A1 |
| *SLC5A3* | C>T | C4 | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | . | C5 |
| *AR* | A>G | NA | A3 | A3 | . | A2 | A3 | A5 | . | A2 | A2 | . | A3 | . | . | . | . | A2 | . | A1 | A1 |
| *IKBKG* | C>T | C4 | C3 | C1 | . | C1 | C4 | . | C5 | C2 | . | C2 | C2 | . | C3 | C3 | . | C1 | C3 | C4 | C4 |

“INS”: insertion; “DEL”: bp deletion; “FM0431”: “Mon_FM0431”; “FM0450”: “Mon_FM0450”; “FM0467”: “Mon_FM0467”; “FM1030”: “Mon_FM1030”; “FM1041”: “Mon_FM1041”; “FM1190”: “Mon_ FM1190”; “FM1190”; “Mon_FM1190”; “FM1785”; “Mon_FM1785”; “FM1798”, “Mon_FM1798”; “FM1932”; “Mon_FM1932”; “FM1948”; “Mon_FM1948”; “FM1951”, “Mon_FM1951”; “FM2218”; “Mon_FM2218”; “EMS595”; “Mor_EMS595”; “A1543”: “Qrt_A1543”; “A5659”: “Qrt_A5659”; “A5964”: “Qrt_A5964”; “A2085”: “Qrt_A2085”; “M977”: “Std_M977”; “M5256”: “Std_M5256”; “M1009”: “Std_M1009”. A dot (.) stands for same alleles as the common allele, otherwise alternate allele associated with the Mendelian trait. A positive sign (+) refers to the presence of indels, while a minus sign (-) stands for the absence of indels.

## Table S4.12. Genes contained in segmental duplications in Yakutian horses and encoding proteins of known functions (1).

| Chr# | Gene ID | #Carriers | #Females | #Males | Gene Name | Gene Description |
|---|---|---|---|---|---|---|
| 1 | ENSECAG00000012176 | 7 | 4 | 3 | *SNX6* | sorting nexin 6 |
| 19 | ENSECAG00000008906 | 7 | 4 | 3 | *PARP15* | poly (ADP-ribose) polymerase family, member 15 |
| 29 | ENSECAG00000024146 | 7 | 4 | 3 | *UCMA* | upper zone of growth plate and cartilage matrix associated |
| 9 | ENSECAG00000007583 | 5 | 4 | 1 | *LY6K* | lymphocyte antigen 6 complex, locus K |
| 18 | ENSECAG00000009694 | 5 | 3 | 2 | *IWS1* | IWS1 homolog (S. cerevisiae) |
| 3 | ENSECAG00000000727 | 4 | 4 | 0 | *HTRA3* | HtrA serine peptidase 3 |
| 3 | ENSECAG00000020879 | 4 | 4 | 0 | *SH3TC1* | SH3 domain and tetratricopeptide repeats 1 |
| 12 | ENSECAG00000020462 | 4 | 3 | 1 | *GSTP1* | glutathione S-transferase pi 1 |
| 14 | ENSECAG00000013199 | 4 | 3 | 1 | *PCDHA1* | protocadherin alpha 1 |
| 19 | ENSECAG00000024988 | 4 | 3 | 1 | *PARP14* | poly (ADP-ribose) polymerase family, member 14 |
| 25 | ENSECAG00000017095 | 4 | 2 | 2 | *CAMSAP1* | calmodulin regulated spectrin-associated protein 1 |
| 25 | ENSECAG00000021229 | 4 | 4 | 0 | *SURF6* | surfeit 6 |
| 28 | ENSECAG00000022371 | 4 | 2 | 2 | *APOBEC3Z1B* | Equus caballus apolipoprotein B mRNA-editing enzyme-catalytic polypeptide-like 3Z1b (APOBEC3Z1B), mRNA. |
| Un | ENSECAG00000004570 | 4 | 3 | 1 | *OR4C3* | olfactory receptor, family 4, subfamily C, member 3 |
| X | ENSECAG00000000012 | 4 | 4 | 0 | *PRKX* | protein kinase, X-linked |
| X | ENSECAG00000000026 | 4 | 4 | 0 | *XG* | Xg blood group |
| X | ENSECAG00000003414 | 4 | 4 | 0 | *ZBED1* | zinc finger, BED-type containing 1 |
| X | ENSECAG00000005751 | 4 | 4 | 0 | *GYG2* | glycogenin 2 |
| X | ENSECAG00000009174 | 4 | 4 | 0 | *ARSD* | arylsulfatase D |
| X | ENSECAG00000010384 | 4 | 4 | 0 | *ARSE* | arylsulfatase E (chondrodysplasia punctata 1) |
| X | ENSECAG00000012190 | 4 | 4 | 0 | *ARSH* | arylsulfatase family, member H |
| X | ENSECAG00000014736 | 4 | 4 | 0 | *OFD1* | oral-facial-digital syndrome 1 |
| X | ENSECAG00000015651 | 4 | 4 | 0 | *ARSF* | arylsulfatase F |
| X | ENSECAG00000017487 | 4 | 4 | 0 | *MXRA5* | matrix-remodelling associated 5 |
| X | ENSECAG00000022025 | 4 | 4 | 0 | *SCML2* | sex comb on midleg-like 2 (Drosophila) |
| X | ENSECAG00000022295 | 4 | 4 | 0 | *AKAP17A* | A kinase (PRKA) anchor protein 17A |
| X | ENSECAG00000022808 | 4 | 4 | 0 | *ASMT* | acetylserotonin O-methyltransferase |
| 2 | ENSECAG00000021856 | 3 | 3 | 0 | *PADI2* | Equus caballus peptidyl arginine deiminase, type II (PADI2), mRNA. |
| 2 | ENSECAG00000024136 | 3 | 2 | 1 | *RHOBTB2* | Rho-related BTB domain containing 2 |
| 7 | ENSECAG00000005132 | 3 | 1 | 2 | *OR2D3* | olfactory receptor, family 2, subfamily D, member 3 |
| 7 | ENSECAG00000011501 | 3 | 1 | 2 | *ZNF215* | zinc finger protein 215 |
| 8 | ENSECAG00000014402 | 3 | 2 | 1 | *MPHOSPH9* | M-phase phosphoprotein 9 |
| 8 | ENSECAG00000020493 | 3 | 2 | 1 | *TOP3B* | topoisomerase (DNA) III beta |
| 10 | ENSECAG00000001193 | 3 | 2 | 1 | *EQUCABV1R918* | Equus caballus vomeronasal 1 receptor equCabV1R918 (EQUCABV1R918), mRNA. |
| 10 | ENSECAG00000009053 | 3 | 1 | 2 | *ZNF304* | zinc finger protein 304 |
| 13 | ENSECAG00000015801 | 3 | 2 | 1 | *POLR3E* | polymerase (RNA) III (DNA directed) polypeptide E (80kD) |
| 16 | ENSECAG00000009171 | 3 | 2 | 1 | *NUP210* | nucleoporin 210kDa |
| 20 | ENSECAG00000013685 | 3 | 1 | 2 | *TNXB* | tenascin XB |
| X | ENSECAG00000007323 | 3 | 3 | 0 | *MTMR1* | myotubularin related protein 1 |
| 2 | ENSECAG00000012961 | 2 | 0 | 2 | *HSPG2* | heparan sulfate proteoglycan 2 |
| 3 | ENSECAG00000014798 | 2 | 1 | 1 | *ZNF19* | zinc finger protein 19 |
| 5 | ENSECAG00000024827 | 2 | 1 | 1 | *ATP1A2* | ATPase, Na+/K+ transporting, alpha 2 polypeptide |
| 6 | ENSECAG00000004076 | 2 | 2 | 0 | *OR10A7* | olfactory receptor, family 10, subfamily A, member 7 |
| 8 | ENSECAG00000011312 | 2 | 0 | 2 | *RTDR1* | rhabdoid tumor deletion region gene 1 |
| 8 | ENSECAG00000018855 | 2 | 2 | 0 | *EP400NL* | EP400 N-terminal like |

## Table S4.12. Genes contained in segmental duplications in Yakutian horses and encoding proteins of known functions (2).

| Chrom.# | Gene ID | #Carriers | #Females | #Males | Gene Name | Gene Description |
|---|---|---|---|---|---|---|
| 10 | ENSECAG00000011643 | 2 | 2 | 0 | *KLK1E2* | Equus caballus glandular kallikrein precursor (KLK1E2), mRNA. |
| 10 | ENSECAG00000018268 | 2 | 2 | 0 | *KLK4* | kallikrein-related peptidase 4 |
| 10 | ENSECAG00000021708 | 2 | 1 | 1 | *FAM26F* | family with sequence similarity 26, member F |
| 12 | ENSECAG00000002053 | 2 | 1 | 1 | *OR5M8* | olfactory receptor, family 5, subfamily M, member 8 |
| 12 | ENSECAG00000013637 | 2 | 2 | 0 | *SLC22A24* | solute carrier family 22, member 24 |
| 23 | ENSECAG00000007210 | 2 | 0 | 2 | *CTSL* | cathepsin L |
| 30 | ENSECAG00000000283 | 2 | 1 | 1 | *F13B* | coagulation factor XIII, B polypeptide |
| X | ENSECAG00000000478 | 2 | 2 | 0 | *CA5B* | carbonic anhydrase VB, mitochondrial |
| X | ENSECAG00000002779 | 2 | 2 | 0 | - | Histone H2A |
| X | ENSECAG00000012150 | 2 | 2 | 0 | *SHROOM2* | shroom family member 2 |
| 1 | ENSECAG00000007462 | 1 | 1 | 0 | *RHOU* | ras homolog family member U |
| 1 | ENSECAG00000008327 | 1 | 0 | 1 | *PI4K2A* | phosphatidylinositol 4-kinase type 2 alpha |
| 1 | ENSECAG00000016495 | 1 | 0 | 1 | *NUSAP1* | nucleolar and spindle associated protein 1 |
| 2 | ENSECAG00000008307 | 1 | 1 | 0 | *UBE4B* | ubiquitination factor E4B |
| 3 | ENSECAG00000018979 | 1 | 1 | 0 | *GRK4* | G protein-coupled receptor kinase 4 |
| 5 | ENSECAG00000021098 | 1 | 0 | 1 | *SLAMF6* | SLAM family member 6 |
| 6 | ENSECAG00000012364 | 1 | 1 | 0 | *MYEOV2* | myeloma overexpressed 2 |
| 7 | ENSECAG00000012108 | 1 | 1 | 0 | *PPP2R1B* | protein phosphatase 2, regulatory subunit A, beta |
| 7 | ENSECAG00000022192 | 1 | 1 | 0 | *ZNF77* | zinc finger protein 77 |
| 8 | ENSECAG00000002578 | 1 | 0 | 1 | *PES1* | pescadillo ribosomal biogenesis factor 1 |
| 8 | ENSECAG00000016242 | 1 | 0 | 1 | - | Tuftelin-interacting protein 11 |
| 9 | ENSECAG00000000565 | 1 | 1 | 0 | *NSMAF* | neutral sphingomyelinase (N-SMase) activation associated factor |
| 9 | ENSECAG00000015053 | 1 | 1 | 0 | *SDCBP* | Equus caballus syndecan binding protein (syntenin) (SDCBP), mRNA. |
| 10 | ENSECAG00000005998 | 1 | 0 | 1 | *TAAR6* | trace amine associated receptor 6 |
| 10 | ENSECAG00000016657 | 1 | 0 | 1 | *ZNF211* | zinc finger protein 211 |
| 11 | ENSECAG00000008934 | 1 | 1 | 0 | *KRT35* | keratin 35 |
| 11 | ENSECAG00000011417 | 1 | 1 | 0 | *KRT32* | keratin 32 |
| 11 | ENSECAG00000022015 | 1 | 0 | 1 | *ALOX12* | arachidonate 12-lipoxygenase |
| 12 | ENSECAG00000020705 | 1 | 1 | 0 | *DUSP8* | dual specificity phosphatase 8 |
| 15 | ENSECAG00000023970 | 1 | 1 | 0 | *MERTK* | c-mer proto-oncogene tyrosine kinase |
| 16 | ENSECAG00000011706 | 1 | 1 | 0 | *DENND6A* | DENN/MADD domain containing 6A |
| 16 | ENSECAG00000015620 | 1 | 0 | 1 | *ZNF850* | zinc finger protein 850 |
| 19 | ENSECAG00000011230 | 1 | 0 | 1 | *SENP2* | SUMO1/sentrin/SMT3 specific peptidase 2 |
| 21 | ENSECAG00000003563 | 1 | 1 | 0 | *BST2* | bone marrow stromal cell antigen 2 |
| 22 | ENSECAG00000006225 | 1 | 1 | 0 | *EDN3* | endothelin 3 |
| 23 | ENSECAG00000002305 | 1 | 1 | 0 | - | Equus caballus interferon-alpha-4 (LOC100052921), mRNA. |
| 23 | ENSECAG00000007878 | 1 | 0 | 1 | *FRMD3* | FERM domain containing 3 |
| 24 | ENSECAG00000015948 | 1 | 1 | 0 | *ACOT4* | acyl-CoA thioesterase 4 |
| 24 | ENSECAG00000016290 | 1 | 1 | 0 | *ACOT6* | acyl-CoA thioesterase 6 |
| 25 | ENSECAG00000017439 | 1 | 1 | 0 | *SPATA31E1* | SPATA31 subfamily E, member 1 |
| Un | ENSECAG00000000039 | 1 | 0 | 1 | *PFKFB1* | 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 1 |
| Un | ENSECAG00000007364 | 1 | 1 | 0 | *COMT* | Equus caballus catechol-O-methyltransferase (COMT), transcript variant MB-COMT, mRNA. |
| Un | ENSECAG00000013435 | 1 | 1 | 0 | *OAS1* | Equus caballus 2',5'-oligoadenylate synthetase 1, 40/46kDa (OAS1), mRNA. |
| Un | ENSECAG00000026951 | 1 | 1 | 0 | *ARVCF* | armadillo repeat gene deleted in velocardiofacial syndrome |
| X | ENSECAG00000009398 | 1 | 1 | 0 | *STS* | Equus caballus steroid sulfatase (microsomal), isozyme S (STS), mRNA. |
| X | ENSECAG00000017572 | 1 | 1 | 0 | *ZNF280C* | zinc finger protein 280C |

## Table S4.13. KEGG and Wikipathway pathways for genes enriched in Yakutian horse segmental duplications.

| Database | Pathways | Model organism | Adjusted p-value | Gene ID | Gene Name |
|---|---|---|---|---|---|
| KEGG | Olfactory transduction | Human | 0.00002 | ENSG00000175619 | *OR4B1* |
| | | | | ENSG00000258817 | *OR4C13* |
| | | | | ENSG00000166368 | *OR2D2* |
| | | | | ENSG00000178358 | *OR2D3* |
| | | | | ENSG00000179919 | *OR10A7* |
| | | | | ENSG00000177693 | *OR4F4* |
| | | | | ENSG00000181371 | *OR5M8* |
| | | Mouse | 0.00280 | ENSMUSG00000043267 | *OLFR1031* |
| | | | | ENSMUSG00000094970 | *OLFR943* |
| | | | | ENSMUSG00000073896 | *OLFR716* |
| | | | | ENSMUSG00000075063 | *OLFR142* |
| | | | | ENSMUSG00000061798 | *OLFR1204* |
| | | | | ENSMUSG00000061195 | *OLFR1289* |
| KEGG | Steroid hormone biosynthesis | Human | 0.01110 | ENSG00000109193 | *SULT1E1* |
| | | | | ENSG00000179142 | *CYP11B2* |
| | | Mouse | 0.00280 | ENSMUSG00000029272 | *SULT1E1* |
| | | | | ENSMUSG00000022589 | *CYP11B2* |
| KEGG | Wnt signaling pathway | Human | 0.02940 | ENSG00000137713 | *PPP2R1B* |
| | | | | ENSG00000163904 | *SENP2* |
| | | Mouse | 0.01210 | ENSMUSG00000022855 | *SENP2* |
| | | | | ENSMUSG00000032058 | *PPP2R1B* |
| KEGG | Ubiquitin mediated proteolysis | Human | 0.02940 | ENSG00000008853 | *RHOBTB2* |
| | | | | ENSG00000130939 | *UBE4B* |
| | | Mouse | 0.01170 | ENSMUSG00000028960 | *UBE4B* |
| | | | | ENSMUSG00000022075 | *RHOBTB2* |
| KEGG | RNA transport | Human | 0.02940 | ENSG00000132182 | *NUP210* |
| | | | | ENSG00000163904 | *SENP2* |
| | | Mouse | 0.01250 | ENSMUSG00000030091 | *NUP210* |
| | | | | ENSMUSG00000022855 | *SENP2* |
| KEGG | Fatty acid metabolism | Mouse | 0.00280 | ENSMUSG00000029455 | *ALDH2* |
| | | | | ENSMUSG00000030861 | *ACADSB* |
| KEGG | Valine, leucine and isoleucine degradation | Mouse | 0.00280 | ENSMUSG00000029455 | *ALDH2* |
| | | | | ENSMUSG00000030861 | *ACADSB* |
| KEGG | Metabolic pathways | Mouse | 0.01070 | ENSMUSG00000029455 | *ALDH2* |
| | | | | ENSMUSG00000030861 | *ACADSB* |
| | | | | ENSMUSG00000000320 | *ALOX12* |
| | | | | ENSMUSG00000025178 | *PI4K2A* |
| | | | | ENSMUSG00000022589 | *CYP11B2* |
| Wikipathways | Metapathway biotransformation | Human | 0.01160 | ENSG00000084207 | *GSTP1* |
| | | | | ENSG00000109193 | *SULT1E1* |
| | | | | ENSG00000179142 | *CYP11B2* |
| Wikipathways | Lymphocyte TarBase | Human | 0.02120 | ENSG00000104689 | *TNFRSF10A* |
| | | | | ENSG00000137575 | *SDCBP* |
| | | | | ENSG00000160886 | *LY6K* |
| | | | | ENSG00000129515 | *SNX6* |
| Wikipathways | Epithelium TarBase | Human | 0.02310 | ENSG00000137575 | *SDCBP* |
| | | | | ENSG00000160886 | *LY6K* |
| | | | | ENSG00000129515 | *SNX6* |

## Table S4.14. Phenotypes for genes enriched in Yakutian horse segmental duplications.

| Database | Phenotype | Model organism | Adjusted p-value | Gene ID | Gene Name |
|---|---|---|---|---|---|
| PheWAS | Circumscribed scleroderma | Human | 0.00480 | ENSG00000104689 | *TNFRSF10A* |
| | | | | ENSG00000153208 | *MERTK* |
| PheWAS | Other hemoglobinopathies | Human | 0.00480 | ENSG00000116574 | *RHOU* |
| | | | | ENSG00000104689 | *TNFRSF10A* |
| PheWAS | Disturbances in tooth eruption | Human | 0.00480 | ENSG00000116574 | *RHOU* |
| | | | | ENSG00000153208 | *MERTK* |
| PheWAS | Disorders of tooth development | Human | 0.00520 | ENSG00000116574 | *RHOU* |
| | | | | ENSG00000153208 | *MERTK* |
| PheWAS | Hereditary hemolytic anemias | Human | 0.00970 | ENSG00000116574 | *RHOU* |
| | | | | ENSG00000104689 | *TNFRSF10A* |
| PheWAS | Chronic airway obstruction | Human | 0.03380 | ENSG00000116574 | *RHOU* |
| | | | | ENSG00000104689 | *TNFRSF10A* |
| PheWAS | Pneumonia | Human | 0.03380 | ENSG00000104689 | *TNFRSF10A* |
| | | | | ENSG00000153208 | *MERTK* |
| PheWAS | Type 1 diabetes | Human | 0.03590 | ENSG00000116574 | *RHOU* |
| | | | | ENSG00000153208 | *MERTK* |
| PheWAS | Disorders of menstruation | Human | 0.04080 | ENSG00000116574 | *RHOU* |
| | | | | ENSG00000153208 | *MERTK* |
| Phenotype | Abnormality of temperature regulation | Human | 0.02120 | ENSG00000196177 | *ACADSB* |
| | | | | ENSG00000142798 | *HSPG2* |
| | | | | ENSG00000018625 | *ATP1A2* |
| | | | | ENSG00000179142 | *CYP11B2* |
| Phenotype | Increased circulating interleukin-12b level | Mouse | 0.01900 | ENSMUSG00000022074 | *TNFRSF10B* |
| | | | | ENSMUSG00000015314 | *SLAMF6* |
| Phenotype | Abnormal circulating interleukin-12b level | Mouse | 0.04620 | ENSMUSG00000022074 | *TNFRSF10B* |
| | | | | ENSMUSG00000015314 | *SLAMF6* |

## Table S4.15. Phenotypes for genes enriched in Yakutian horse segmental duplications (Humans as model organisms).

| Disease | Adjusted p-value | Gene ID | Gene Name |
|---|---|---|---|
| Urogenital Neoplasms | 0.02500 | ENSG00000167749 | *KLK4* |
| | | ENSG00000170801 | *HTRA3* |
| | | ENSG00000108839 | *ALOX12* |
| | | ENSG00000084207 | *GSTP1* |
| | | ENSG00000109193 | *SULT1E1* |
| Essential hypertension | 0.02500 | ENSG00000125388 | *GRK4* |
| | | ENSG00000167748 | *KLK1* |
| | | ENSG00000179142 | *CYP11B2* |
| Hypertension, Renal | 0.02500 | ENSG00000125388 | *GRK4* |
| | | ENSG00000167748 | *KLK1* |
| | | ENSG00000179142 | *CYP11B2* |
| Demyelinating Diseases | 0.03110 | ENSG00000153208 | *MERTK* |
| | | ENSG00000051825 | *MPHOSPH9* |
| | | ENSG00000117115 | *PADI2* |
| Beckwith-Wiedemann Syndrome | 0.03110 | ENSG00000157429 | *ZNF19* |
| | | ENSG00000149054 | *ZNF215* |
| Multiple Sclerosis | 0.03110 | ENSG00000153208 | *MERTK* |
| | | ENSG00000051825 | *MPHOSPH9* |
| | | ENSG00000117115 | *PADI2* |
| Hypertension | 0.03110 | ENSG00000125388 | *GRK4* |
| | | ENSG00000167748 | *KLK1* |
| | | ENSG00000179142 | *CYP11B2* |
| Ototoxicity | 0.03110 | ENSG00000084207 | *GSTP1* |
| | | ENSG00000109193 | *SULT1E1* |
| Cholestasis | 0.03110 | ENSG00000132182 | *NUP210* |
| | | ENSG00000067066 | *SP100* |
| Brenner tumour of ovary | 0.03110 | ENSG00000167749 | *KLK4* |
| | | ENSG00000167748 | *KLK1* |
| | | ENSG00000108839 | *ALOX12* |

# 5 Section 5: Phylogenomic inference

We reconstructed the phylogenetic relationships amongst the Yakutian horses sequenced here, by comprehensively comparing their genetic variability against a set of previously sequenced genomes from both, present-day and Late Pleistocene specimens (13, 16). More specifically, we compared their mitochondrial, Y-chromosome and nuclear genomes.

## *5.1 Phylogenetic inference from mitochondrial sequences*

### 5.1.1 Dataset and substitution models

Following the procedure described in **section S2.5,** we mapped the sequencing reads of 12 modern **(Table S2.5)** and 9 ancient **(Table S2.6)** Yakutian horses against the reference mitochondrial genome (Accession Nb. NC_001640; (83)). Samples for which the average depth-of-coverage was inferior to 4X were disregarded, resulting in a total of 16 mitochondrial genomes analyzed. BAM alignments were converted into pileup format, and subsequently used in bcftools (84) to call genotypes (depth ≥ 3, and baseQ ≥ 30). The final mtDNA consensus sequences were generated using a strict majority rule at each position (disregarding 16,129-16,360 bp corresponding to tandem repeats), which resulted in an average depth-of-coverage of 30.6-3,277.4X. The final consensus sequences were submitted to GenBank, under Accession numbers KT368723-KT368738.

We aligned our 16 mitochondrial genome sequences to the collection of 105-horse mitochondrial genomes that was previously presented in (16**).** This collection is comprised by ancient and modern specimens, and encompasses most of the mitochondrial diversity present in horses **(Table S5.1)**.

We applied ModelGenerator v0.851 (85) to determine which nucleotide substitution model best fitted the mitochondrial data. The mitochondrial sites were first split into six categories, including the control region, rRNAs, tRNAs, and the three codon positions (detailed in **Table S5.2**). We also generated a global alignment corresponding to the concatenation of the six categories. For each site category, as well as the global alignment, the best-fit nucleotide substitution model was selected by means of the Bayesian Information Criterion (BIC) (86), and used for the corresponding Maximum Likelihood (from the global alignment) and Bayesian (from the six partitions) phylogenetic inferences.

### 5.1.2 Maximum likelihood phylogeny

We inferred the maximum likelihood (ML) phylogenetic tree from the global mitochondrial alignment by using PhyML3.0 (87) with the corresponding best-fit substitution model (TrN+I+Γ8). Node robustness was assessed using three statistics based on approximate Likelihood Ratio Tests (88): (i) Approximate-Unbiased test (aLRT); (ii) aLRT parametric Chi2-based (Chi2-aLRT); and (iii) aLRT non-parametric branch support based on a Shimodaira-Hasegawa-like procedure (SH-aLRT) (89). The unrooted phylogenetic tree, drawn using Dendroscope 3 (90), is shown in **Figure S5.1**.

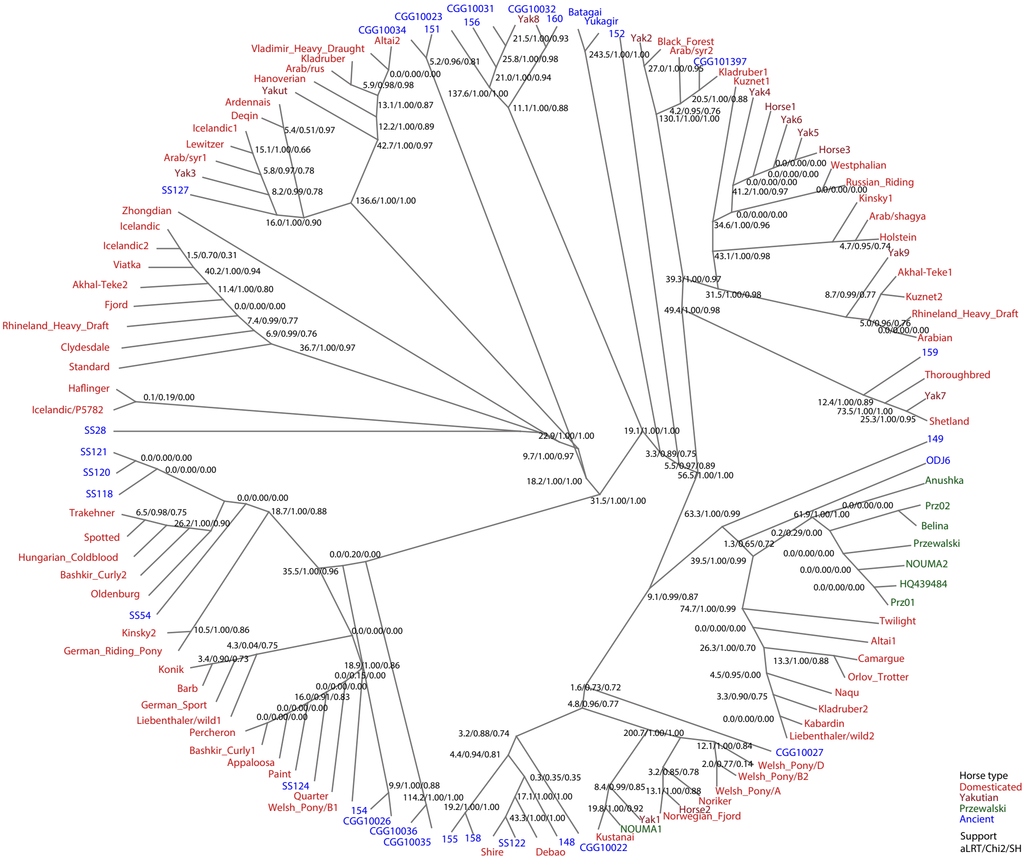


**Figure S5.1. Mitochondrial ML phylogeny of ancient and modern horses.**
The mitochondrial horse genomes considered are described in **(Table S5.1)** and the phylogeny was estimated using PhyML 3.0 (87). Node support values (aLRT/Chi2-aLRT/SH-aLRT) are indicated at each node. Samples are color-coded as ancient (in blue), domestic (red), Przewalski's (green), and modern Yakutian horses (brown).

### 5.1.3 Bayesian phylogeny

We also performed Bayesian phylogenetic inference using BEAST 1.8.0 (91). The six site categories were treated as unlinked partitions. Divergence dates were estimated using a log-uncorrelated molecular clock model, with radiocarbon dates (or stratigraphic context information) of the ancient specimens employed for tip calibration **(Table S5.1)**. The Bayesian phylogenetic inference was run for 150 millions generations, sampling the chain every 1,000 (thin-in interval), and discarding the first 10% chains (burn-in). We analysed the MCMC samples with TRACER 1.5 (92), which indicated convergence and adequate mixing of the Markov chains. We thus used the TREEANNOTATOR 1.7.5 program to summarize the MCMC samples as the maximum clade credibility topology (91). The tree, drawn with Figtree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/), is shown in **(Figure S5.2)** with the corresponding node support values (posterior probabilities).

We also evaluated the fit of three demographic models to the mitochondrial data, including a constant population size, a Bayesian Skyline (93) and a Bayesian Skyride (94). We then compared their marginal likelihoods through the Bayes factors, which were calculated in TRACER v1.5 (92) via importance sampling (1,000 bootstraps), and using the harmonic mean of the sampled likelihoods. The comparison supported the Bayesian Skyline as the best-fit demographic model to the mtDNA data **(Table S5.3)**. We therefore reconstructed the Bayesian Skyline demographic profile ($N_e$ x generation time) with TRACER 1.5 (92), and plotted its log10-transformed value using the ggplot2 (95) package of the R programming language **(Figure S5.3)**.

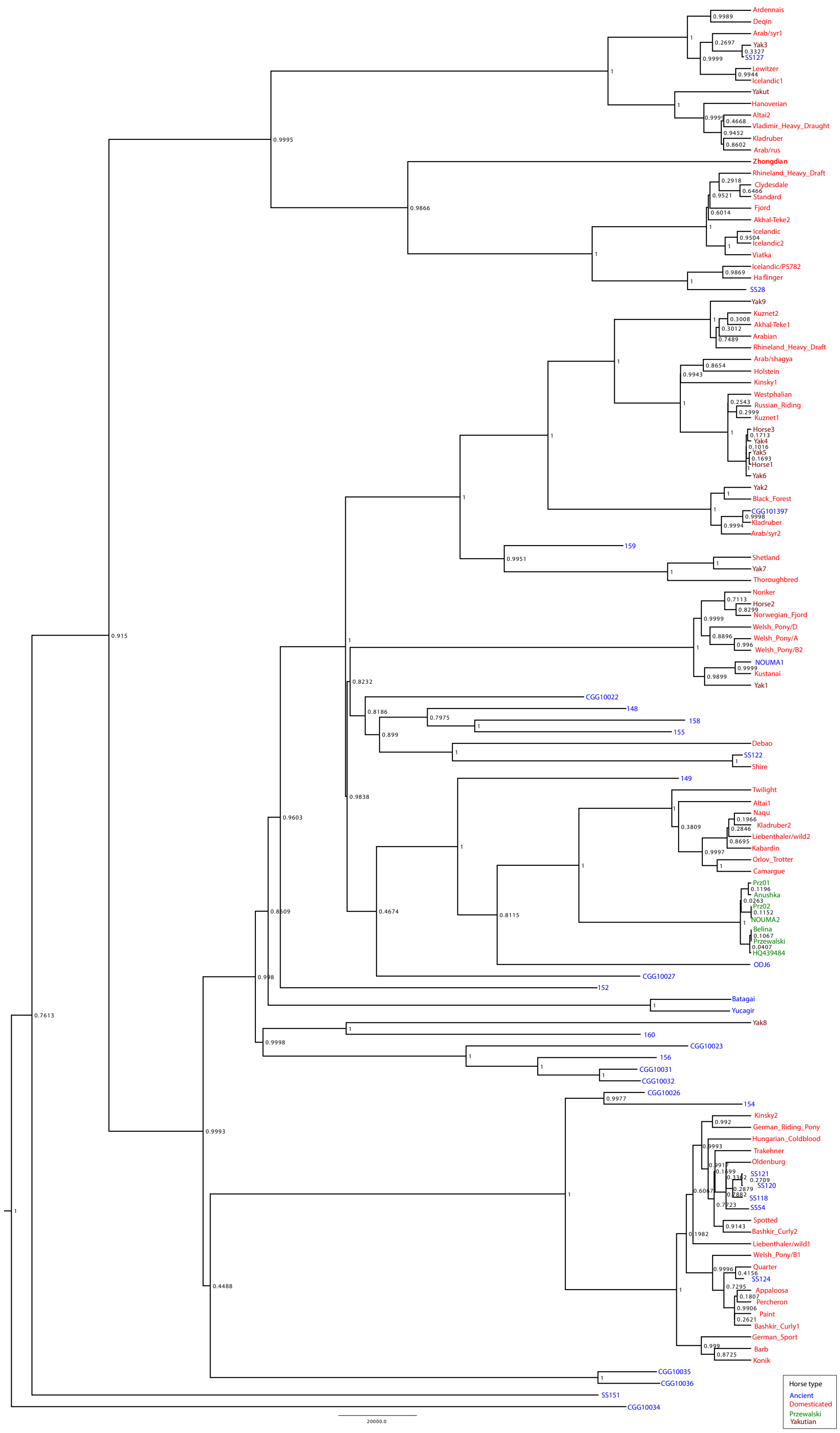


**Figure S5.2. Mitochondrial Bayesian phylogeny of ancient and modern horses.**
The mitochondrial horse genomes considered are described in **(Table S5.1)** and the phylogeny was estimated using BEAST and six unlinked site partitions **(Table S5.2)**. The node labels depict the clade support (posterior probabilities). Samples are color-coded as ancient (in blue), domestic (red), Przewalski's (green), and modern Yakutian horses (brown).

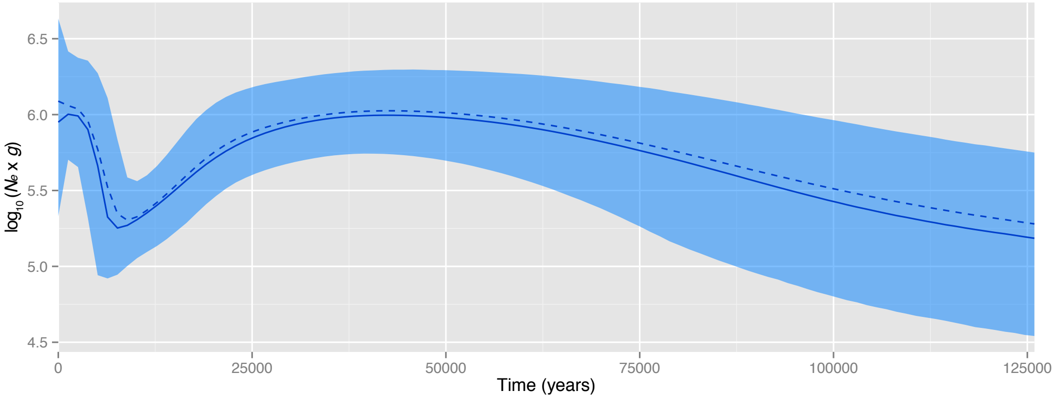


**Figure S5.3: Bayesian Skyline plot reconstructed from mitochondrial genome data.**
The y-axis provides a $\log_{10}$-transformed measure of the product of the effective population size ($N_e$) and the generation time ($g$). The full and dashed blue lines represent the median and the mean of demographic traces, respectively. The blue shaded area delimits the 95% confidence intervals of the mean.

## *5.2 Y-chromosome phylogenetic inference*

We first prepared FASTA files corresponding to the Y-chromosome contigs characterised by Wallner et al. (96) and Lippold et al. (97). We excluded sequence G72337.1 as it overlaps between both studies, resulting in a total of 193,857bp.

We mapped the sequencing reads of six Yakutian horse samples, including four modern stallions (Yak1, Yak4, Yak8 and Yak9) and the two ancient specimens, Batagai and CGG101397, against these Y-chromosome contigs. To assure we retained Y-specific sequencing reads, we followed a two-step mapping approach. The first step involved mapping against the Y-chromosome contigs (without filtering for quality and PCR duplicates). This provided an initial list of Y-chromosome read candidates that were further mapped against both the EquCab2.0 nuclear and the Y-chromosome sequences. Only candidate reads mapping again uniquely to the Y-chromosome were considered to be Y-specific, provided they showed a minimal mapping quality of 25 and were not PCR duplicates.

We called genotypes as described in **section S4.1.1**, except that a minimum depth of 4 (as the Y-chromosome is haploid) and a maximum depth of 50 were required. Based on the genomic coordinates of the reference sequence, we then merged the Y-linked genotypes with the sequence information previously characterized for two modern domesticated horses (Standarbred and Icelandic; from (13) and (16) and one Late Pleistocene stallion (CGG10023, from (16)).

The phylogenetic relationships among these samples was inferred by Maximum Likelihood (ML) with PhyML3.0 (87)**,** under the GTR substitution model with eight rate categories, as estimated by ModelGenerator v0.851 (85). Node support was assessed using three statistics based on two approximate Likelihood Ratio Tests (88): the aLRT parametric Chi2-based (Chi2-aLRT), and the aLRT non-parametric branch support based on a Shimodaira-Hasegawa-like procedure (SH-aLRT) (89). The unrooted phylogenetic tree, drawn using Dendroscope 3 (90), is shown in **Figure S5.4**.

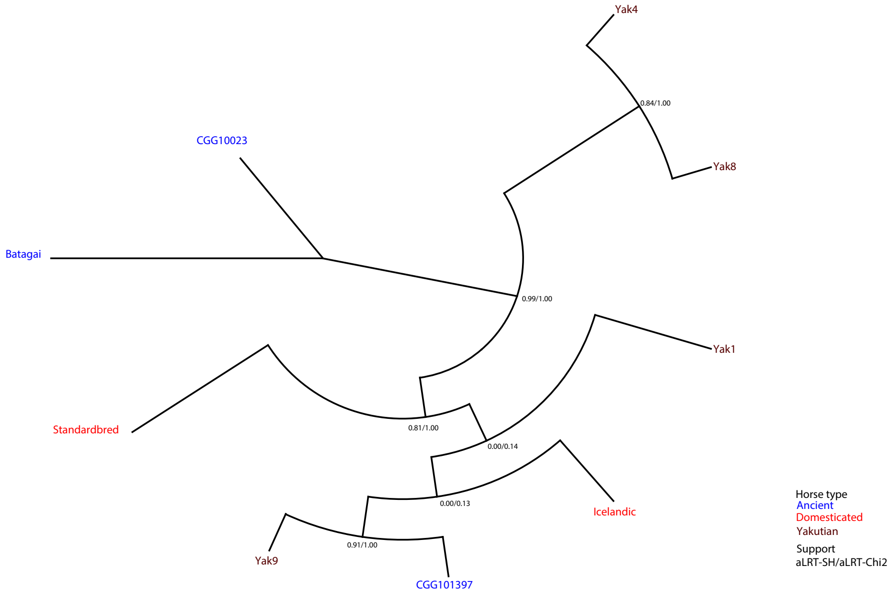


**Figure S5.4. Y-chromosome Maximum Likelihood phylogeny.**
Phylogenetic inference was performed using PhyML 3.0 (87). Node support values (Chi2-aLRT/SH-aLRT) are indicated at each node. Samples are color-coded as ancient (in blue), domestic (red), Przewalski's (green), and modern Yakutian horses (brown).

### *5.3 Phylogenetic inference using a super-matrix of nuclear coding sequences*

We inferred the phylogenetic relationship amongst the 9 modern and 2 ancient Yakutian samples (Batagai and CGG101397), by comparing their nuclear genetic variability with that present in Przewalski's and domesticated horses **(Table S5.4)** (13, 16, 30). The low-coverage genomes of other Yakutian horses were not compatible with this analysis **(Table S2.5** and **S2.6).**

More specifically, we used the VCF genotyping calls produced in **section S4.1** to build a partitioned super-matrix based on the 50% longest protein-coding genes annotated in EquCab v2.76 (98). Support was calculated based on 100 bootstrap trees. Maximum Likelihood phylogenetic inference was performed using ExaML v2.04 (http://sco.h-its.org/exelixis/software.html) and RAxML v8.1.3 (99), with default parameters. The resulting tree was rooted using the divergence from the outgroup *Equus africanus somaliensis* (Somali wild ass) **(Figure S5.5).**

We found that the specimens Batagai (dating ~5,200 BP ago) and CGG101397 (dating to the 19th century) did not cluster together. Instead, sample CGG101397 grouped within the diversity of modern Yakutian horses (which is also forming a monophyletic clade, included within that of modern domesticated horses), while Batagai clustered with two previously characterised Late Pleistocene horses (CGG10022 and CGG10023). Przewalski's horses were found to represent a third phylogenetic cluster.

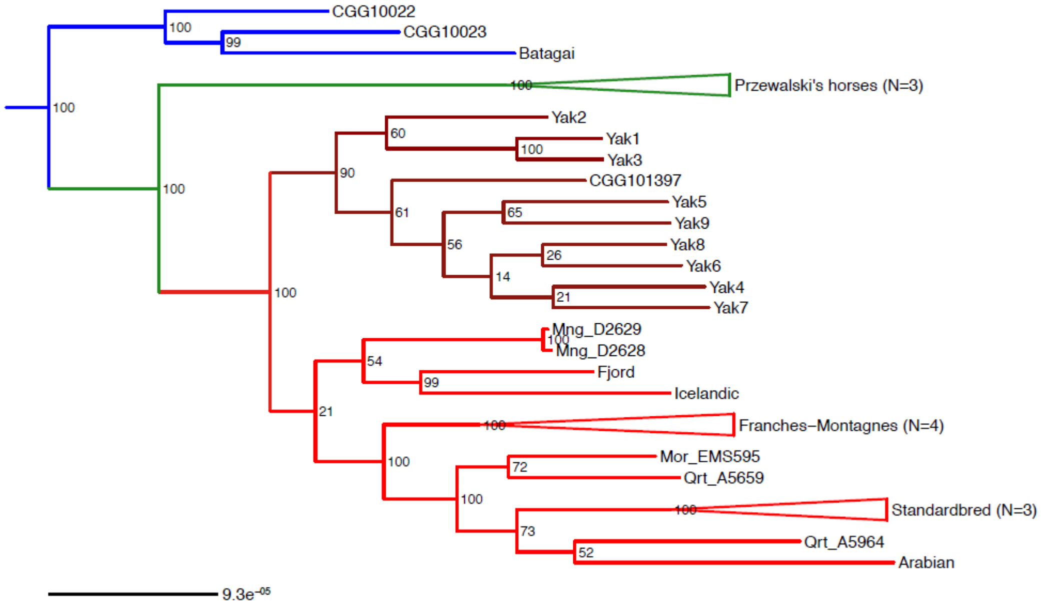


**Figure S5.5. Exome-based phylogenetic inference.**
Individual samples are color-coded as: ancient horses (blue), modern domesticated horses (red), Przewalski's horses (green) and modern Yakutian horses (brown). Node labels show the corresponding support values using 100 bootstrap pseudo-replicates. For improving readability, we collapsed all individuals from the same breeds that clustered together. The number of horses (*N*) in collapsed tips is listed after the breed name. The topology presented is rooted using the divergence from the outgroup *Equus africanus somaliensis* (Somali wild ass, not shown), previously characterized by (30).

## 5.4 *Supplementary Tables for Section 5*

## Table S5.1. Complete mitochondrial genome sequences used for phylogenetic inference.

| Horse type/Sample name | Accession number | Reference | Tip Age for Calibration in BEAST (yBP) |
|---|---|---|---|
| Ancient/Batagai | KT368725 | This study | 5,088 |
| Ancient/CGG101397 | KT368726 | This study | 100 |
| Ancient/Yukagir | KT368723 | This study | 5,384 |
| Ancient/ODJ6 | KT368724 | This study | 225 |
| Yakutian/Yak1 | KT368730 | This study | 0 |
| Yakutian/Yak2 | KT368731 | This study | 0 |
| Yakutian/Yak3 | KT368732 | This study | 0 |
| Yakutian/Yak4 | KT368733 | This study | 0 |
| Yakutian/Yak5 | KT368734 | This study | 0 |
| Yakutian/Yak6 | KT368735 | This study | 0 |
| Yakutian/Yak7 | KT368736 | This study | 0 |
| Yakutian/Yak8 | KT368737 | This study | 0 |
| Yakutian/Yak9 | KT368738 | This study | 0 |
| Yakutian/Horse2 | KT368727 | This study | 0 |
| Yakutian/Horse1 | KT368728 | This study | 0 |
| Yakutian/Horse3 | KT368729 | This study | 0 |
| Przewalski/Belina | AP012267 | Goto et al. 2011 (100) | 0 |
| Przewalski/Anushka | AP012268 | Goto et al. 2011 (100) | 0 |
| Przewalski/NOUMA2 | AP013095 | n/a | 0 |
| Przewalski/NOUMA1 | NC_024030 | n/a | 0 |
| Przewalski/Prz01 | JN398402 | Achilli et al. 2012 (101) | 0 |
| Przewalski/Prz02 | JN398403 | Achilli et al. 2012 (101) | 0 |
| Zhongdian | EF597512 | Xu et al. 2007 (102) | 0 |
| Naqu | EF597513 | Xu et al. 2007 (102) | 0 |
| Deqin | EF597514 | Xu et al. 2007 (102) | 0 |
| Debao | EU939445 | Jiang et al. 2010 (103) | 0 |
| Akhal-Teke1 | HQ439441 | Lippold et al. 2011 (97) | 0 |
| Akhal-Teke2 | HQ439442 | Lippold et al. 2011 (97) | 0 |
| Altai1 | HQ439443 | Lippold et al. 2011 (97) | 0 |
| Altai2 | HQ439444 | Lippold et al. 2011 (97) | 0 |
| Kladruber | HQ439445 | Lippold et al. 2011 (97) | 0 |
| Appaloosa | HQ439446 | Lippold et al. 2011 (97) | 0 |
| Arabian/rus | HQ439447 | Lippold et al. 2011 (97) | 0 |
| Arabian/syr1 | HQ439448 | Lippold et al. 2011 (97) | 0 |
| Arabian/syr2 | HQ439449 | Lippold et al. 2011 (97) | 0 |
| Ardennais | HQ439450 | Lippold et al. 2011 (97) | 0 |
| Bashkir Curly1 | HQ439451 | Lippold et al. 2011 (97) | 0 |
| Bashkir Curly2 | HQ439452 | Lippold et al. 2011 (97) | 0 |
| Barb | HQ439453 | Lippold et al. 2011 (97) | 0 |
| Camargue | HQ439454 | Lippold et al. 2011 (97) | 0 |
| Clydesdale | HQ439455 | Lippold et al. 2011 (97) | 0 |

| | | | |
|---|---|---|---|
| German Sport | HQ439456 | Lippold et al. 2011 (97) | 0 |
| Hanoverian | HQ439457 | Lippold et al. 2011 (97) | 0 |
| Holstein | HQ439458 | Lippold et al. 2011 (97) | 0 |
| Oldenburg | HQ439459 | Lippold et al. 2011 (97) | 0 |
| Westphalian | HQ439460 | Lippold et al. 2011 (97) | 0 |
| German Riding Pony | HQ439461 | Lippold et al. 2011 (97) | 0 |
| Thoroughbred | HQ439462 | Lippold et al. 2011 (97) | 0 |
| Norwegian Fjord | HQ439463 | Lippold et al. 2011 (97) | 0 |
| Haflinger | HQ439464 | Lippold et al. 2011 (97) | 0 |
| Icelandic1 | HQ439465 | Lippold et al. 2011 (97) | 0 |
| Icelandic2 | HQ439466 | Lippold et al. 2011 (97) | 0 |
| Yakut | HQ439467 | Lippold et al. 2011 (97) | 0 |
| Kabardin | HQ439468 | Lippold et al. 2011 (97) | 0 |
| Kinsky1 | HQ439469 | Lippold et al. 2011 (97) | 0 |
| Kinsky2 | HQ439470 | Lippold et al. 2011 (97) | 0 |
| Kladruber1 | HQ439471 | Lippold et al. 2011 (97) | 0 |
| Kladruber2 | HQ439472 | Lippold et al. 2011 (97) | 0 |
| Konik | HQ439473 | Lippold et al. 2011 (97) | 0 |
| Kuznet1 | HQ439474 | Lippold et al. 2011 (97) | 0 |
| Kuznet2 | HQ439475 | Lippold et al. 2011 (97) | 0 |
| Kustanai | HQ439476 | Lippold et al. 2011 (97) | 0 |
| Lewitzer | HQ439477 | Lippold et al. 2011 (97) | 0 |
| Liebenthaler/wild1 | HQ439478 | Lippold et al. 2011 (97) | 0 |
| Liebenthaler/wild2 | HQ439479 | Lippold et al. 2011 (97) | 0 |
| Noriker | HQ439480 | Lippold et al. 2011 (97) | 0 |
| Orlov Trotter | HQ439481 | Lippold et al. 2011 (97) | 0 |
| Paint | HQ439482 | Lippold et al. 2011 (97) | 0 |
| Percheron | HQ439483 | Lippold et al. 2011 (97) | 0 |
| Przewalski | HQ439484 | Lippold et al. 2011 (97) | 0 |
| Rhineland Heavy Draft | HQ439485 | Lippold et al. 2011 (97) | 0 |
| Russian Riding Horse | HQ439486 | Lippold et al. 2011 (97) | 0 |
| Black Forest | HQ439487 | Lippold et al. 2011 (97) | 0 |
| Arab/shagya | HQ439488 | Lippold et al. 2011 (97) | 0 |
| Shetland | HQ439489 | Lippold et al. 2011 (97) | 0 |
| Shire | HQ439490 | Lippold et al. 2011 (97) | 0 |
| Rhineland Heavy Draft | HQ439491 | Lippold et al. 2011 (97) | 0 |
| Spotted | HQ439492 | Lippold et al. 2011 (97) | 0 |
| Trakehner | HQ439493 | Lippold et al. 2011 (97) | 0 |
| Hungarian Cold Blood | HQ439494 | Lippold et al. 2011 (97) | 0 |
| Viatka | HQ439495 | Lippold et al. 2011 (97) | 0 |
| WelshPony/D | HQ439496 | Lippold et al. 2011 (97) | 0 |
| WelshPony/A | HQ439497 | Lippold et al. 2011 (97) | 0 |
| WelshPony/B1 | HQ439498 | Lippold et al. 2011 (97) | 0 |
| WelshPony/B2 | HQ439499 | Lippold et al. 2011 (97) | 0 |
| Vladimir Heavy Draught | HQ439500 | Lippold et al. 2011 (97) | 0 |
| Przewalski | n/a | Orlando et al. 2013 (13) | 0 |
| Arabian | n/a | Orlando et al. 2013 (13) | 0 |
| Fjord | n/a | Orlando et al. 2013 (13) | 0 |
| Twilight | n/a | Orlando et al. 2013 (13) | 0 |
| Standard | n/a | Orlando et al. 2013 (13) | 0 |
| Icelandic/P5782 | n/a | Orlando et al. 2013 (13) | 0 |

|  |  |  |  |
|---|---|---|---|
| Icelandic | n/a | Orlando et al. 2013 (13) | 0 |
| Quarter | n/a | Orlando et al. 2013 (13) | 0 |
| Ancient/CGG10022 | n/a | Schubert et al. 2014 (16) | 42,692 |
| Ancient/CGG10023 | n/a | Schubert et al. 2014 (16) | 16,099 |
| Ancient/CGG10026 | n/a | Orlando et al. 2013 (13) | 27,230 |
| Ancient/CGG10027 | n/a | Orlando et al. 2013 (13) | 28,336 |
| Ancient/CGG10031 | n/a | Orlando et al. 2013 (13) | 29,081 |
| Ancient/CGG10032 | n/a | Orlando et al. 2013 (13) | 28,198 |
| Ancient/CGG10034 | n/a | Orlando et al. 2013 (13) | 31,890 |
| Ancient/CGG10035 | n/a | Orlando et al. 2013 (13) | 24,104 |
| Ancient/CGG10036 | n/a | Orlando et al. 2013 (13) | 23,244 |
| Ancient/148 | n/a | Orlando et al. 2013 (13) | 31,917 |
| Ancient/149 | n/a | Orlando et al. 2013 (13) | 18,471 |
| Ancient/151 | n/a | Orlando et al. 2013 (13) | 39,022 |
| Ancient/152 | n/a | Orlando et al. 2013 (13) | 39,311 |
| Ancient/154 | n/a | Orlando et al. 2013 (13) | 2,235 |
| Ancient/155 | n/a | Orlando et al. 2013 (13) | 20,273 |
| Ancient/156 | n/a | Orlando et al. 2013 (13) | 24,154 |
| Ancient/158 | n/a | Orlando et al. 2013 (13) | 16,880 |
| Ancient/159 | n/a | Orlando et al. 2013 (13) | 32,667 |
| Ancient/160 | n/a | Orlando et al. 2013 (13) | 28,242 |
| Ancient/SS127 | n/a | Sawyer et al. 2012 (104) | 1,815 |
| Ancient/SS28 | n/a | Sawyer et al. 2012 (104) | 1,250 |
| Ancient/SS54 | n/a | Sawyer et al. 2012 (104) | 550 |
| Ancient/SS118 | n/a | Sawyer et al. 2012 (104) | 2,150 |
| Ancient/SS120 | n/a | Sawyer et al. 2012 (104) | 2,150 |
| Ancient/SS121 | n/a | Sawyer et al. 2012 (104) | 2,150 |
| Ancient/SS122 | n/a | Sawyer et al. 2012 (104) | 2,150 |
| Ancient/SS124 | n/a | Sawyer et al. 2012 (104) | 1,815 |

yBP, years Before Present; cal. yBP, calibrated years Before Present.

**Table S5.2. Best-fit nucleotide substitution model, according to the Bayesian Information Criterion (BIC) in ModelGenerator v0.851.**

| Partition | #Sites* | Model BIC |
|---|---|---|
| 1st codon position | 3,626 | HKY+I |
| 2nd codon position | 3,623 | HKY+I |
| 3rd codon position | 3,623 | TrN+Γ8 |
| Control Region | 965 | TrN+I+Γ8 |
| rRNA | 2,561 | HKY+I+Γ8 |
| tRNA | 1,521 | TrN+I+Γ8 |
| Global alignment | 15,919 | TrN+I+Γ8 |

Number of sites considered in each partition/dataset, based on the horse reference mtDNA genome (Accession number NC_001640.1). The global dataset corresponds to the concatenate of the other six categories; HKY, Hasegawa-Kishino-Yano substitution model (89). TrN, Tamura-Nei subsitution model (105). I, Invariant sites; Γ8, 8 Gamma categories to model of substitution rate variation among sites.

**Table S5.3. Bayes Factor for three demographic models fitted in the BEAST phylogenetic analyses.**

| Tree model | Bayes factor (ln P (model \| data)) |
|---|---|
| Constant population size | -30,841.2 |
| Skyline | -30,857.7 |
| Skyride | -30,851.2 |

## Table S5.4. Samples used for phylogenetic inference based on nuclear coding sequences.

| Sample | Horse type | Reference |
|---|---|---|
| Arabian | Arabian | Orlando et al. 2013 (13) |
| Batagai | Ancient Yakutia | this study |
| CGG10022 | Ancient Taymyr peninsula | Schubert et al. 2014 (16) |
| CGG10023 | Ancient Taymyr peninsula | Schubert et al. 2014 (16) |
| CGG101397 | Ancient Yakutia | this study |
| Fjord | Fjord | Orlando et al. 2013 (13) |
| Icelandic | Icelandic | Orlando et al. 2013 (13) |
| Mng_D2628 | Mongolian | Do et al. 2014 (19) |
| Mng_D2629 | Mongolian | Do et al. 2014 (19) |
| Mon_FM1030 | Franches-Montagnes | this study |
| Mon_FM1785 | Franches-Montagnes | this study |
| Mon_FM1932 | Franches-Montagnes | this study |
| Mon_FM1951 | Franches-Montagnes | this study |
| Mor_EMS595 | Morgan | this study |
| Prz_D2630 | Przewalski | Do et al. 2014 (19) |
| Prz_D2631 | Przewalski | Do et al. 2014 (19) |
| Prz_Przewalski | Przewalski | Orlando et al. 2013 (13) |
| Qrt_A5659 | Quarter | this study |
| Qrt_A5964 | Quarter | this study |
| Std_M5256 | Standardbred | this study |
| Std_M977 | Standardbred | this study |
| Std_Standardbred | Standardbred | Orlando et al. 2013 (13) |
| Yak1 | Yakutian | this study |
| Yak2 | Yakutian | this study |
| Yak3 | Yakutian | this study |
| Yak4 | Yakutian | this study |
| Yak5 | Yakutian | this study |
| Yak6 | Yakutian | this study |
| Yak7 | Yakutian | this study |
| Yak8 | Yakutian | this study |
| Yak9 | Yakutian | this study |

# 6 Section 6: Demographic history of the Yakutian horses

In this section, we addressed a long-standing debate about the demographic history of the Yakutian horses, namely whether they arrived with the Yakut people a few centuries ago, or rather descend or were admixed with wild Late Pleistocene horses. We tested these hypotheses by applying a range of methods (i) characterizing the effective population size ($N_e$) changes of Yakutian horses over time, (ii) estimating their genetic distances with other populations/breeds, as well as (iii) detecting the presence of admixture events.

## *6.1 Reconstructing their long and short-term effective population sizes*

### 6.1.1 Pairwise Sequentially Markov Coalescent

We used PSMC (106) to trace $N_e$ changes over the last two million years. The method operates on a single diploid genome, exploiting the local density of heterozygous positions along unlinked loci to date the distribution of coalescent events (and thus $N_e$ in a temporal scale).

Because deep sequencing is required to accurately call heterozygous positions, a minimum coverage of 20X is required for appropriate PSMC inferences (13). Two Yakutian genomes in our data set meet this requirement, namely the historical sample CGG101397 (20.24x) and the modern Yakutian horse Yak7 (21.63x) (**Tables S2.5 and S2.6**). The genomes of Batagai and Yak2 were sequenced at an average coverage of 18.29X and 18.38X, respectively, which is sub-optimal for PSMC inference. In order to include these genomes in the analyses, we therefore applied the correction method based on uniform false negative rates (uFNR) of heterozygous calls, devised in (13) and suggested in (106). Briefly, Yak7 (which has the highest genome coverage) was first randomly down-sampled to an average depth-of-coverage of 18.3X (corresponding to the average genome coverage observed for both ancient Yakutian genomes). We then tested a different set of uFNR of heterozygous calls (from 1% to 30%), searching for the one that minimizes the differences with the PSMC profile recovered from the original Yak7 sequence. We found that PSMC inference on 18.3X depth-of-coverage datasets could be satisfactorily corrected with ~2% as uFNR of heterozygous calls.

PSMC analyses were performed as described in the Supplementary section S9.2 of (13), except that we defined a per-sample maximum coverage threshold to avoid overestimating heterozygosity in the presence of unidentified repetitive regions (see **section S4.2**). Briefly, consensus FASTA sequences were generated for autosomal chromosomes from BAM files. SNPs were then called with samtools v0.1.18 (84) and filtered with the 'vcf2fq' command from vcfutils.pl, using the following settings:

- Minimum depth-of-coverage: 8X.
- Maximum depth-of-coverage: 0.995 quantile of the coverage distribution.
- InDels filtered in a windows size of 5 bp.
- Minimum RMS mapping quality: 10.
- Sites with genotyping quality scores inferior to 35 were also excluded, but during the conversion to the PSMC input file format.

PSMC inference was run using the parameters recommended in (13) and (16), including *Number of iterations* = 25; *maximum $2N_0$ coalescent time* = 15; *initial $\theta_0/\rho$* = 5. We evaluated the variance of the PSMC reconstructions, by splitting chromosomal sequences into shorter segments of 500 Kb and generating 100 genome bootstrap pseudo-replicates (with replacement). Unscaled PSMC profiles are shown in **Figure S6.1**. For scaling PSMC profiles, we used a calibration point of T = 4.5 million years (Myr), a mutation rate of $\mu = 7.242\times10^{-9}$ per site per generation, and a generation time of g = 8 years. These parameters are in agreement with the recent findings from Orlando and colleagues (13) and (16). Scaled PSMC profiles are shown in **Figure 3A**.

PSMC profiles reported virtually identical demographic trajectories for modern Yakutian horses and the historical Yakutian horse CGG101397. In contrast, the 5.2 kyr-old Yakutian sample Batagai showed a similar pattern to that observed for the Late Pleistocene horse CGG10022, which was originally analysed in (16) and is included here for comparative purposes **(Figure 3A)**.

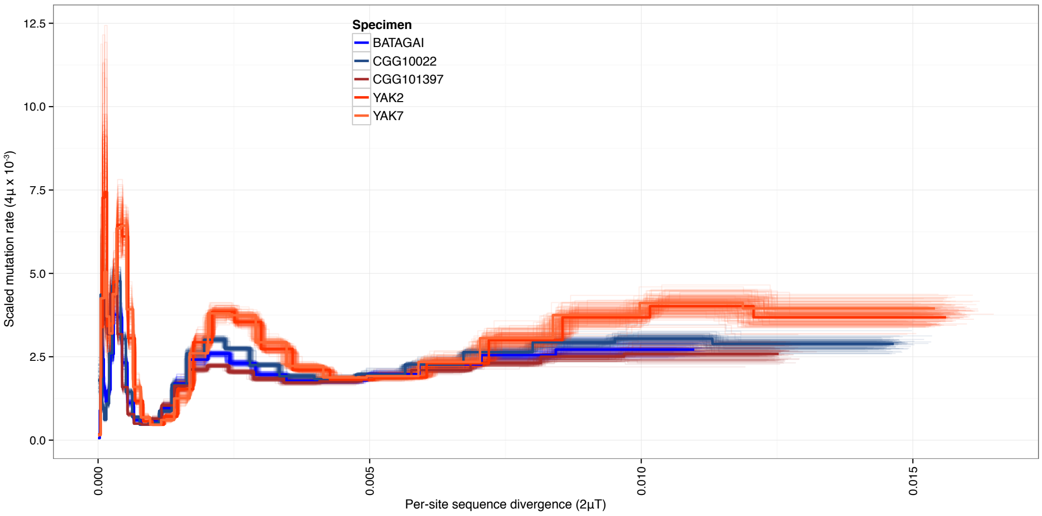


**Figure S6.1. Unscaled PSMC profiles.**
Bootstrap pseudo-replicates are indicated with thin lines around the corresponding PSMC profile.

### 6.1.2 Diffusion approximation for demographic inference

We reconstructed the recent demographic history of Yakutian horses using the software *dadi* (version 1.6.3), where the estimation of demographic parameters is based on the diffusion approximation to the site frequency spectrum (SFS) (107).

We first considered the SFS calculated from 4d-fold degenerate sites (see definition in **section S7.2.1**), as a proxy for neutrally evolving and approximately independent sites. We used the reference (Thoroughbred; **Table S2.4**), Mongolian (Mng_D2628; **Table S2.4**), and Donkey (*Equus africanus somaliensis*) outgroup species to infer derived alleles. After masking monomorphic sites, the unfolded SFS resulted in 9,381 to 9,573 entries, depending on the ancestral sequence used. As additional analyses, we also considered the SFS computed across the whole genome, comprised of a total of 2,843,575 to 2,933,002 SNPs.

We fixed the ancient demographic history of Yakutian horses from the model inferred by PSMC analyses (see **section S6.1.1**). The ancestral effective population size was therefore set to 136,337, and it changed instantaneously over time until approximately 18,000 yBP, as detailed in **Table S6.1**, which was calibrated considering a generation time of 8 years. We then tested three demographic models of recent evolution for Yakutian horses (from 18,000 yBP to the present):

**- Model PSMC**: at 18,545 yBP (see **Table S6.1**) the effective population size drops instantaneously to 5,107, exactly as estimated by PSMC analyses;

**- Model one epoch**: is the same as Model PSMC, but assuming that at 5,500 yBP (domestication time; (108), the effective population size exponentially decays to $N_2$, which represents the current size ($N_1$ and $N_2$ are free parameters to be estimated);

**- Model three epochs**: the same as the one epoch model, but assuming that at a putative founder time (700 yBP; (1, 2)) the effective population size instantaneously drops to *N3* and exponential recovers to *N4*, which represents the current size. $N_1$, $N_2$, $N_3$, $N_4$ are free parameters to be estimated.

During each model optimization, we also estimated an additional parameter, i.e. the probability of misidentification ($P_{misid}$) of the derived state based on the ancestral sequence used. For each model, we ran the program 20 times from different initial values to ensure convergence, and retained the parameters set with the highest likelihood **(Table S6.2)**. To compare fitting between models, we performed a likelihood ratio test (LRT) with two degrees of freedom, equal to the difference of free parameters.

Our analyses show that the model assuming constant size starting ~18,000 yBP, as estimated by PSMC analyses, provides a poor fit to the data, suggesting that PSMC performs poorly for recent times (**Table S6.2**). Conversely, imposing a population size change at the domestication time greatly improves the model fitting and the current effective population size is estimated at approximately 3,000, irrespective of the ancestral sequence used (**Table S6.2**). Furthermore, when assuming a bottleneck event at a hypothetical founder time (700 yBP), we observe a significant increase in the likelihood ($P$ = 0.041, 0.045, 0.020 using the reference, Mongolian, or Donkey sequence as ancestral state, respectively). Current effective population size was estimated to be between 5,600 and 7,900 (**Table S6.2**).

We finally tested whether a different assumption on the founder event provided a better fit to the data. We therefore used as initial values the parameters for the best-fit model and let vary the founder (bottleneck) event at different arbitrary times: 5,000, 4,000, 3,000, 2,000, 1,000, 500 and 200 yBP. To improve the optimization, we fixed $N_1$ and $N_2$ parameters (old population size changes) as estimated in the best-fit models. We recorded the best model likelihood at these different times (**Tables S6.3**).

We observed that more recent times appear to provide greater model likelihood (**Table S6.3, Fig. S1**). These results are not an artefact of sequencing heterogeneity, as this trend is observed even when the whole-genome data set is used (**Table S6.3**). We could not pinpoint a precise timing for the founder event, as FS-based methods (like the one adopted here) may have low power for characterising bottlenecks (109). However, these results show that it is more likely that this event occurred within the last 1,000 years, in agreement with historical records (1, 2).

### *6.2 Genome projections of Yakutian and other domesticated horses*

We next applied the method of projecting test genomes onto reference populations (110). The projection value *w* summarizes the past relationship between a single test genome and a reference population, by exploiting differences in their derived allele frequencies. Projections of $w = 1$ across all mutation frequency classes indicate that the test genome is randomly sampled from the reference population. Otherwise, it highlights a more complex demographic history of the tested individual and/or the reference population, being especially sensitive to recent $N_e$ changes and small amounts of admixture. For this analysis, we compared the genomes detailed in **Table S6.4**, and grouped them according to their corresponding breeds/population.

### 6.2.1 Test horse genomes and reference panels

We first considered the refDOM panel, which includes all of the 27 horse genomes from non-Yakutian domesticated breeds. By including a high number of individuals, this panel may provide greater resolution of the projection for rare alleles. However, it is composed of genomes from several horse breeds/types, which violates the random mating assumption made for the reference population. Therefore, to avoid the confounding effects originating from the refDOM structure, we focused our subsequent projections onto refFM and refYAK populations.

In panel refFM, we only considered a subset of refDOM, exclusively comprised of 12 Franches-Montagnes horses. Whereas panel refYAK grouped all of the nine modern Yakutian horse genomes characterized in this study.

### 6.2.2 Projection results

*Projections of modern Yakutian horse genomes onto the reference panels, refYAK and refFM*

Projections of modern Yakutian horses onto refYAK mostly lie around $w = 1$ **(Figure S6.2)**, which corresponds to the expectation of a random mating population, including a slight departure for rare alleles (110). For three genomes, namely Yak1, Yak3 and notably Yak2, the projections rose above the $w = 1$ line for alleles found at a low frequency, suggesting they experienced reduced levels of admixture, which is in agreement with our admixture test (see **section S6.5**).

All nine Yakutian horse genomes have similar projections relative to refFM, with minimum projection values (MPV) below one (mean MPV = 0.7724), and a standard deviation (sd) of 0.0092 (**Table S6.5**), suggesting that modern Yakutian and FM horses do not belong to the same breed, as expected.

*Projections of Przewalski's and Yakutian horse genomes onto refFM*

When projected onto refFM, the mean MPV is interestingly lower for Przewalski's horses (mean MPV = 0.6762; sd = 0.0109) than for modern Yakutian horses (mean MPV = 0.7724; sd = 0.0092; **Table S6.5**), indicating that Franches-Montagnes horses are more closely related to modern Yakutian horses than to the Przewalski's horses (**Figure S6.3** and **S6.4**). This is consistent with the phylogenetic position of modern Yakutian horses within the domesticated horse clade, and outside the monophyletic group of Przewalski's horses (see **section S5.3**).

*Projections of non-Yakutian modern horse genomes onto refYAK*

Further evidence that the modern Yakutian horses belong to the clade of domesticated horses is provided by the reciprocal projections onto refYAK (**Figure S6.5)**, which report relatively high MPVs when testing either the Franches-Montagnes horse genomes (mean MPV = 0.8038; sd = 0.0062) or non-Franches-Montagnes domesticated horses (MPV = 0.8073; sd = 0.0246) (**Table S6.5**). These MPVs are higher than the MPVs observed when projecting the Przewalski's horse genomes onto refYAK (mean MPV = 0.6926; sd = 0.0009; **Table S6.5)**, in line with their early phylogenetic divergence (see **section S5.3)**. Although the two mean MPVs obtained when testing Franches-Montagnes and non-Franches-Montagnes are comparable, the standard deviation is about four fold larger for non-Franches-Montagnes, nicely reflecting their heterogeneous genetic background, consisting of a mixture from eight different domesticated breeds.

*Projections of ancient horse genomes onto the reference panels, refYAK and refFM*

Projections of test ancient genomes onto refYAK show that CGG101397 on the one hand, and the other three surveyed ancient horses on the other hand (namely, CGG10022, CGG10023 and Batagai), have very different demographic histories (**Figure S6.6**). The genome of specimen CGG101397 shows mean MPVs of 0.8719 and 0.7637, when projected onto refYAK and refFM panels, respectively **(Table S6.5)**. This suggests a closer relationship to modern Yakutian horses than to the Franches-Montagnes horses. These findings are in agreement with the results from phylogenetic analyses (see **section S5.3**), where CGG101397 clustered within the modern Yakutian diversity, supporting genetic discontinuity in the horse population of Yakutia, with the ancient population represented by sample Batagai (~5.2kyr sample) being replaced by the population of present-day domesticated horses.

The projections of the CGG10022, CGG10023 and Batagai ancient horse genomes onto refYAK (mean MPV = 0.5977; sd = 0.0085) and refFM (mean MPV = 0.5955; sd = 0.0050) are very similar. The extremely reduced variance observed in the projections of these three genomes further suggests that they share a similar demographic history, despite spanning a ~40 kyr-long temporal range. The mean MPV obtained for the projections of the CGG10022, CGG10023 and Batagai genomes onto refYAK (mean MPV = 0.5977) is much lower than those obtained for the modern domesticated horses (mean MPV = 0.8038-0.8073), and even than those observed when projecting Przewalski´s horses (mean MPV = 0.6926) **(Table S6.5).** This, again, supports that Late Pleistocene horses and the Batagai sample diverged prior to the most recent common ancestor of Przewalski's, Yakutian (including CGG101397), and other domesticated horses.

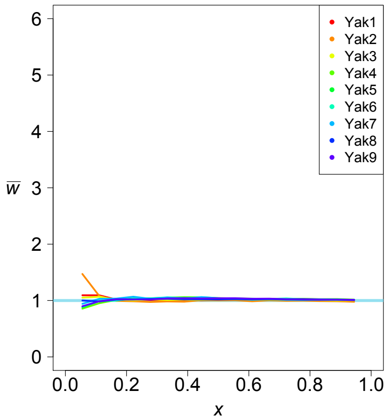


**Figure S6.2. Projections of modern Yakutian horses onto the refYAK panel.**
The *x*-axis represents the categories of derived allele frequencies, while *w* their corresponding projection (a *w* value above 1 indicates that the test genome has more alleles at that frequency than reference panel, and *vice versa*).

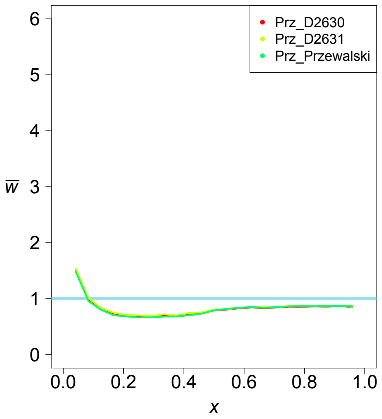


**Figure S6.3. Projections of Przewalski's horses onto the refFM panel.**
See **Figure S6.2** for captions.

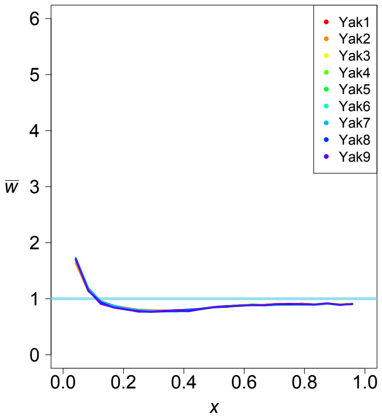


**Figure S6.4. Projections of modern Yakutian horses onto the refFM panel.**
See **Figure S6.2** for captions.

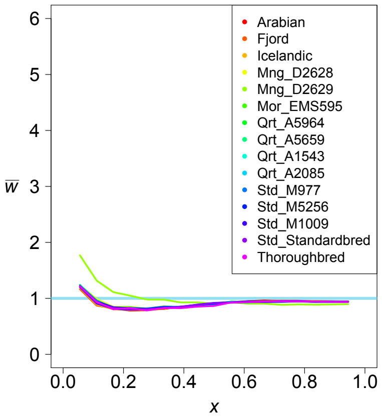


**Figure S6.5. Projections of non-Yakutian and non-Franches-Montagnes modern horses onto the refYAK panel.**

See **Figure S6.2** for captions.

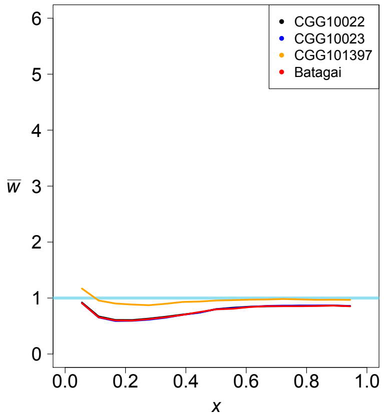


**Figure S6.6. Projections of ancient horse genomes onto the refYAK panel.**
See **Figure S6.2** for captions.

## *6.3 Principal component analysis*

Principal Component Analysis (PCA) was used to investigate the genetic distances amongst ancient (including the two Late Pleistocene horses CGG10022 and CGG10023, and the two ancient Yakutian horses Batagai and CGG101397) and modern horses (including Przewalski's horses and modern domesticated horses, either Yakutian or not).

### 6.3.1 PCA based on genotype calls

VCF files encoding individual-based genotypes were collected (see **section 4.1.1**), and merged with bcftools (111). We restricted the analysis to biallelic sites called for every individual (-g ^miss), with a minimum frequency for the alternate allele of –q 0.00001. This merged VCF file, containing a total of 356,720 variants, was converted into plink format with vcftools v0.1.12 (111). This plink file was used as input for the EIGENSTRAT program vEIG5.0.1 (112), which was run with no outlier iterations (numoutlieriter: 0). Finally, the first three principal components, which explain 27.71% of the total variance, were plotted with ggplot2 (95) in R 3.02.

We found three main clusters (**Figure S6.7)**. The first included all Przewalski's horses, while the second included the two Late Pleistocene horses and the ancient Yakutian sample Batagai. The third cluster consisted in all other horses, including the historical Yakutian horse from the 19th century (CGG101397), which grouped together within the diversity of modern domesticated horses, especially within the Yakutian breed. This is in line with our phylogenomic inference based on the exome (see **section S5.3**) and the topology recovered from TreeMix analyses (see **section S6.5.2**).

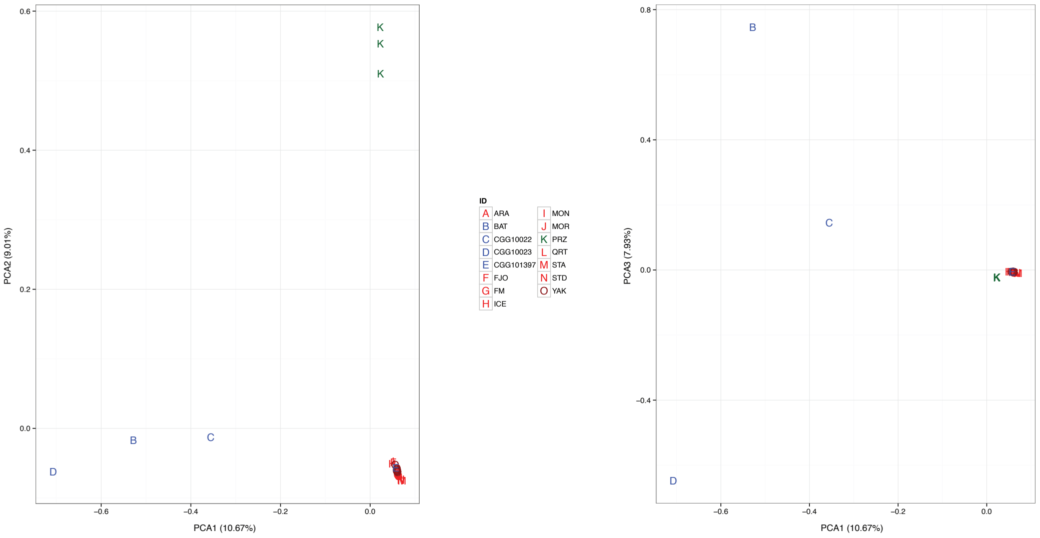


**Figure S6.7. PCA plot depicting the genetic affinities of ancient and modern horses, based on genotype calls.**
Left: first two principal components. Right: first and third principal components. The proportion of the variance explained by each principal component is indicated on each axis, between parentheses. For clarity, labels only refer to breeds or populations, except for Yakutian and Late Pleistocene horses.

### 6.3.2 PCA in a genotype likelihood framework

We also performed PCA with the ngsCovar program of the ngsTools package (113), which accounts for genotype uncertainties. Briefly, priors for the allele frequencies of each position (saf files) were estimated with ANGSD v0.615 (46). For that, we assumed Hardy-Weinberg equilibrium, and considered reads showing a mapping quality greater than or equal to 25 (-minMapQ), as well as sites showing a minimum base quality of 20 (-minBQ) and for which data was available in at least five individuals (-minInd). The saf files were then used as input for ngsCovar.

PCA plots were generated for the first three components, which explained 13.33% of the total variance (**Figure S6.8**). Although we consistently recovered the same three main clusters as in the SNP-based analysis (**Figure S6.7)**, the PCA relying on genotype likelihoods revealed finer sub-clustering patterns, especially the separation of domesticated horses into their corresponding breeds/populations. For example, modern Yakutian horses form a clearly separated subgroup together with the historical horse CGG101397, with Yak2 being genetically closer to the Mongolian horses.

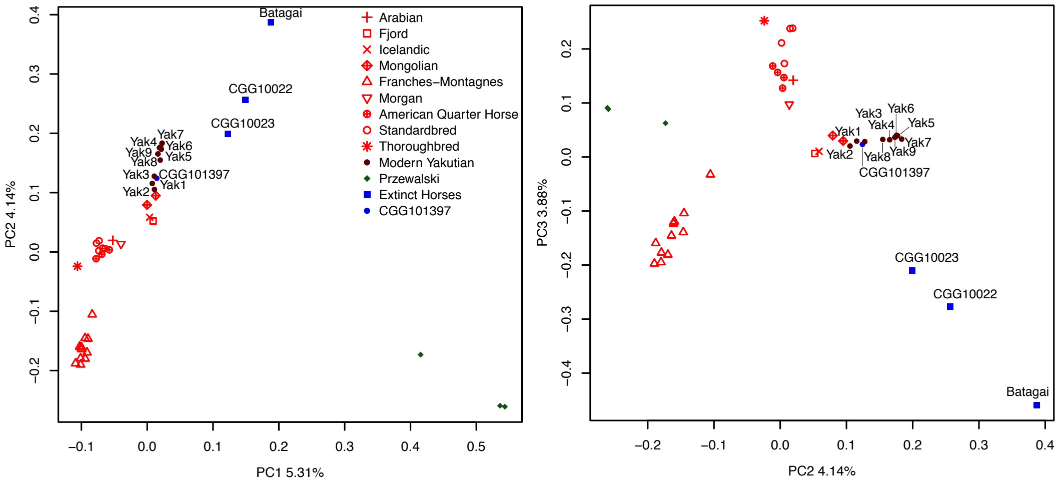


**Figure S6.8. PCA plot depicting the genetic affinities of ancient and modern horses, based on genotype likelihoods.**
The analysis was based on genotype likelihoods and 43 genomes representative of present-day Yakutian horses, nine domestic breeds, the Przewalski's horse population, CGG101397 and extinct horses. The fraction of the total variance explained by each of the three principal components is indicated on the corresponding axes. Left: first two principal components. Right: second and third principal components.

As the over-representation of one group of related individuals can impact the overall structure reflected in PCA, we also ran PCA on a subset of genomes, considering a maximum three individuals per population/breed of the comparative panel (**Figure 2**). These were selected to represent the genomes characterised with the highest average depth-of-coverage: Yak1, Yak2, Yak3, Yak4, Yak5, Yak6, Yak7, Yak8, Yak9, Arabian, Fjord, Icelandic, Mng_D2628, Mng_D2629, Mon_FM1951, Mon_FM1798, Mor_EMS595, Qrt_A5659, Qrt_A5964, Std_M977, Std_Standardbred, Throughbred, Prz_D2630, Prz_D2631, Prz_Przewalski, Batagai, CGG10022, CGG10023, and CGG101397. The resulting structure was in agreement with that from the analyses presented above, showing 1) three clusters represented by domesticated horses, Przewalski's horses, and extinct horses, respectively, 2) close affinities between the historical (CGG101397) and the modern Yakutian horses, and 3) close genetic proximity between Mongolian horses and historical/modern Yakutian horses.

## *6.4 Admixture tests*

We explored whether modern and ancient Yakutian horses showed evidence of genetic admixture using a range of admixture tests based on D-statistics (114) f3-statistics (115), Tree-Graph reconstructions in TreeMix (116), and NGSadmix (117)**.**

### 6.4.1 D-statistcs

The D-statistics, also know as ABBA-BABA test (114), assesses whether the genetic data of three taxa deviate from their tree-like relationship. Briefly, given three taxa ($H_1$, $H_2$ and $H_3$), an outgroup (O) and a tree topology ((($H_1$, $H_2$), $H_3$), O)), the D-statistics quantifies the occurrence of two incomplete sorting patterns, so called ABBA and BABA events, where A and B refer to ancestral (namely, identical to O) and derived allelic states. ABBA events occur when $H_2$ and $H_3$ share the derived allele (B) while the ancestral allele (A) is carried by $H_1$. Conversely, in BABA events the ancestral allele (A) is carried by $H_2$ while $H_1$ and $H_3$ share the derived one. Under the null hypothesis that the tree is correct and there is no gene flow connecting $H_3$ to either $H_1$ or $H_2$, the ABBA and BABA events result from incomplete lineage sorting, and, thus, occur with equal frequency. Enrichment of ABBA sites (or BABA), therefore, indicates the possible presence of gene flow between the $H_2$ and $H_3$) lineages ($H_1$ and $H_3$, respectively).

Such deviations from equal proportion of ABBA-BABA events can however result from ancestral substructure prior to the divergence of $H_1$, $H_2$ and $H_3$, as well as to heterogeneities in genome-wide sequencing error rates (118). The latter was ruled out by only using samples with large differences in sequencing error rates as $H_3$. Moreover, to reduce the effect of base calling errors in ancient genomes (and to a lower extent those of some modern Yakutian horse genomes also showing an excess of GC→AT substitutions, see **section S2.6**), we restricted the calculation of D-statistics to transversions.

To comprehensively assess the amount of gene flow between horse lineages, we calculated multiple D-statistics, corresponding to different taxa combinations as $H_1$, $H_2$ or $H_3$ compatible with the inferred tree topology (**Figure S5.5**). As outgroup species, we used *Equus africanus somaliensis* from (30), because it was re-sequenced at higher depth-of-coverage than the domestic donkey (*Equus asinus asinus*) originally reported and used by Orlando and colleagues (13). The full list of surveyed scenarios is illustrated in **(Figure S6.9)**.

Our significance assessment of D-statistics follows (13), where we applied a 10Mb block jackknife procedure to accommodate for the large levels of linkage disequilibrium in horses. The D-statistics significance was expressed as Z-scores, which are generally considered significant when their absolute values are higher than 3. However, in order to control the family-wise error rate arising from the large number of tests performed, we corrected the Z-scores for multiple testing using the function 'p.adjust' with the Holm correction (Holm 1979) in R. A test was considered significant when showing an adjusted p-value smaller than 0.05.

When one ancient horse was considered as $H_3$ and two Yakutian or two non-Yakutian domesticated horses as $H_1$ and $H_2$, the vast majority of the D-statistics were not significantly different from zero (**Figure S6.9A-B)**. Conversely, the topology tested was rejected when including Przewalski's horses as $H_1$ or $H_2$ **(Figure S6.9C-D**). Interestingly, all D-statistics calculated for the tree topology with non-Yakutian domesticated horses (**Figure S6.9D**) were significantly positive (2.95 < Z-scores < 22.91), in agreement with the findings from Schubert and colleagues based on a more limited genome dataset (16). This confirms that the extinct population of ancient horses (represented by Batagai, CGG10022 and CGG10023) contributed genetically to the population of domesticated horses, prior to or in the early stages of horse domestication.

The tree topology was rejected when non-Yakutian and Yakutian modern domesticated horses were placed as $H_1$ and $H_2$ and CGG101397 as $H_3$ **(Figure S6.9E)**. This was not the case when any other ancient horse (Batagai, CGG10022, and CGG10023) was used as $H_3$ **(Figure S6.9E)**. This not only agrees with a high genetic affinity between CGG101397 and modern Yakutian horses, but also rejects genetic continuity from more ancient Yakutian horses (here represented with the 5.2-kyr old Batagai sample) to modern Yakutian horses (see **sections S5.3, S6.3 and S6.4**). Finally, the tree topology could not be rejected when Przewalski's horses were placed as $H_3$ and modern Yakutian horses were placed as $H_1/H_2$, suggesting that no Yakutian individual is particular closer to Przewalski's horses **(Figure S6.9F)**.

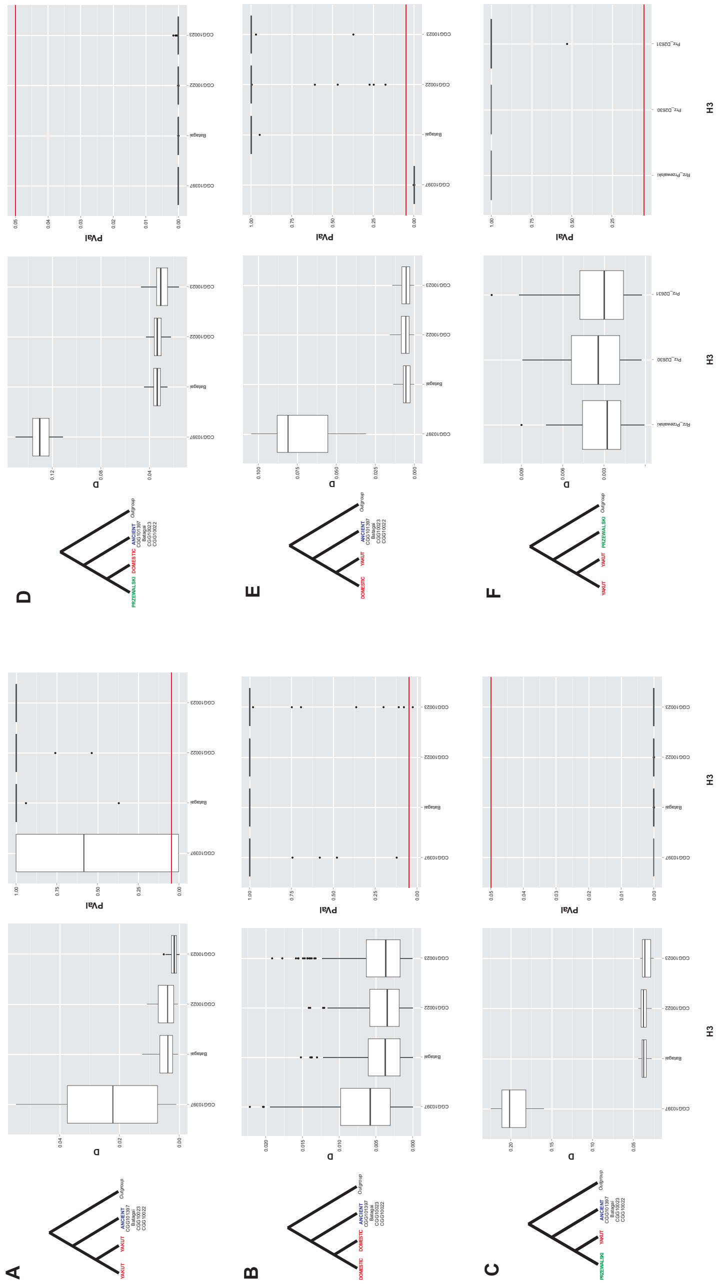


**Figure S6.9. Admixture tests based on D-statistics and transversions.**

The combinations tested are shown in the first and fourth columns. The D-statistics and p-values are indicated in the second/fifth columns and in the third/sixth columns, respectively. The red line indicates the p-value significance threshold of 0.05, following correction for multiple testing. In panels A-E, we tested ancient specimens (as $H_3$) against modern horses (considered as $H_1$ and $H_2$), while in panel F, Przewalski's horses were tested (as $H_3$) against our panel of domesticated breeds (considered as $H_1$ and $H_2$), including Yakutian horses. Ancient: Batagai, CGG10022, CGG10023, CGG101397. Domestic: Arabian, Fjord, Icelandic, Mng_D2629, Mng_D2628, Mon_FM0431, Mon_FM0450, Mon_FM0467, Mon_FM1030, Mon_FM1041, Mon_FM1190, Mon_FM1785, Mon_FM1798, Mon_FM1932, Mon_FM1948, Mon_FM1951, Mon_FM2218, Mor_EMS595, Qrt_A1543, Qrt_A5659, Qrt_A5964, Qrt_A2085, Std_Standardbred, Std_M977, Std_M5256, Std_M1009, Thoroughbred. Przewalski: Prz_Przewalski, Prz_D2631, Prz_D2630. Yakut: Yak1, Yak2, Yak3, Yak4, Yak5, Yak6, Yak7, Yak8, Yak9. Outgroup: *Equus africanus somaliensis*.

### 6.4.2 $f_3$-statistics and TreeMix analyses

The plink file already generated for the PCA based on genotype calls (see **section S6.3.1**) was converted into a TreeMix input file (116). For that, we used the Python plink2treemix.py script provided by the TreeMix package (https://code.google.com/p/treemix/), grouping samples as indicated in **Table S6.6**. We ran TreeMix using a block resampling procedure of groups of 1,500 variants (-k 1,500).

*$f_3$-statistics*

We calculated the $f_3(C;A,B)$-statistic (115)**,** which assesses whether a population C is the result of an admixture of ancestral populations A and B (**Figure S6.10**). In this section, we use the notation from (115) for the expected value of $f_3$:

$$E\,[f_3(C;A,B)] = c+\alpha^2 d+\beta^2 e -\alpha\beta(g+f)$$

where $f$, $g$, $d$, $e$ and $c$ represent the branch lengths illustrated in **Figure S6.10**, while $\alpha$ and $\beta$ the corresponding admixture proportions. If the $f_3$-statistic is negative, it supports a complex population history whereby population C descends from both A and B populations. However, we found no significant results when ancient specimens were placed as the test population C (**Table S6.7**), suggesting an absence of admixture for each combination tested.

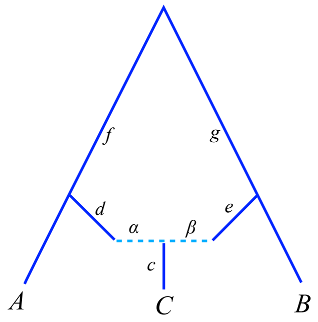


**Figure S6.10. Schematic illustration depicting the notation used to estimate $f_3$-statistics values.** For three hypothetical populations (*A, B* and *C*) branch lengths (*f, g, d, e* and *c)* and admixture proportions (*α* and *β*). Migration edges are depicted with a light blue dashed line.

*TreeMix: estimation of the population trees with admixture*

We used TreeMix (116) to infer horse population splits, and subsequent admixture events. TreeMix analyses were run considering up to four migration events (-m 0-4). We performed a round of global rearrangements after all populations were added (-global), and placed the root of the inferred trees at the root of the clade represented by the ancient specimens CGG10022, CGG10023, Batagai (option -root CGG10022, CGG10023, Batagai). The resulting trees were plotted using the supplied TreeMix R functions. The total fraction of the variance explained by each migration model was estimated with the 'get_f()' R function.

TreeMix recovers the same tree topologies as the one inferred from our phylogenomic analysis (see **section S5.3; Figures 1B** and **S6.11**), regardless of the number of migration events considered. The Late Pleistocene horses (CGG10022 and CGG10023) cluster with the ancient Yakutian sample Batagai, forming a sister lineage clearly separated from the modern horses and CGG101397. The majority of the allele frequency variation (variance = 98.6669%) is solely explained by the tree topology, with migration events only providing a marginal increment (from up to 0.26%). The first migration edge is between the Prz_Przewalski and the population ancestral to the domesticated breeds (0.17% increase in vairance). The variance explained by the additional migration edges is much smaller than the first migration edge (<0.09%), and simply involves admixture within modern Yakutian horses, probably reflecting their shared ancestry.

### 6.4.3 Genetic Clustering based on genotype likelihoods

The genetic structure of our panel of modern and ancient horses was investigated using the function "NGSAdmix" implemented in ANGSD, which handles genotype uncertainties in a Maximum Likelihood framework (46, 117). The analyses were run considering two to 20 ancestral populations (-K), a minimum minor allele frequency of 5% (-minMaf) and a maximum number of Expectation Maximization iterations of 5,000 (-maxiter) (**Figure S6.12**). We retained K = 6 as the plausible number of ancestral populations, as it retrieves the known genetic structure among present-day domesticated horse breeds, including: (K1) Arabian horses, Morgan horses, Thoroughbred horses, Standardbred horses, American Quarter horses, in accordance with the latter three descending from Arabian horses; (K2) Franches-Montagnes; (K3) Przewalski´s horses; (K4) Mongolian horses, Icelandic horses and Norwegian Fjord horses, in agreement with the genetic affinities reported in (13, 16, 119), and in our phylogenetic results (see **section S5**). Importantly, the last two clusters consist of (K5) Batagai and the Late Pleistocene horses CGG10022 and CGG10023, and (K6) all modern Yakutian horses, including sample CGG101397. No ancestry component is shared between clusters (K5) and (K6), confirming that samples Batagai and CGG101397 reflect two different population ancestries, in line with the results from our other analyses. Of note, some horses such as Yak2 seem to share some of their ancestry with modern Mongolian horses, as also suggested by our genome projections (see **section S6.2**).

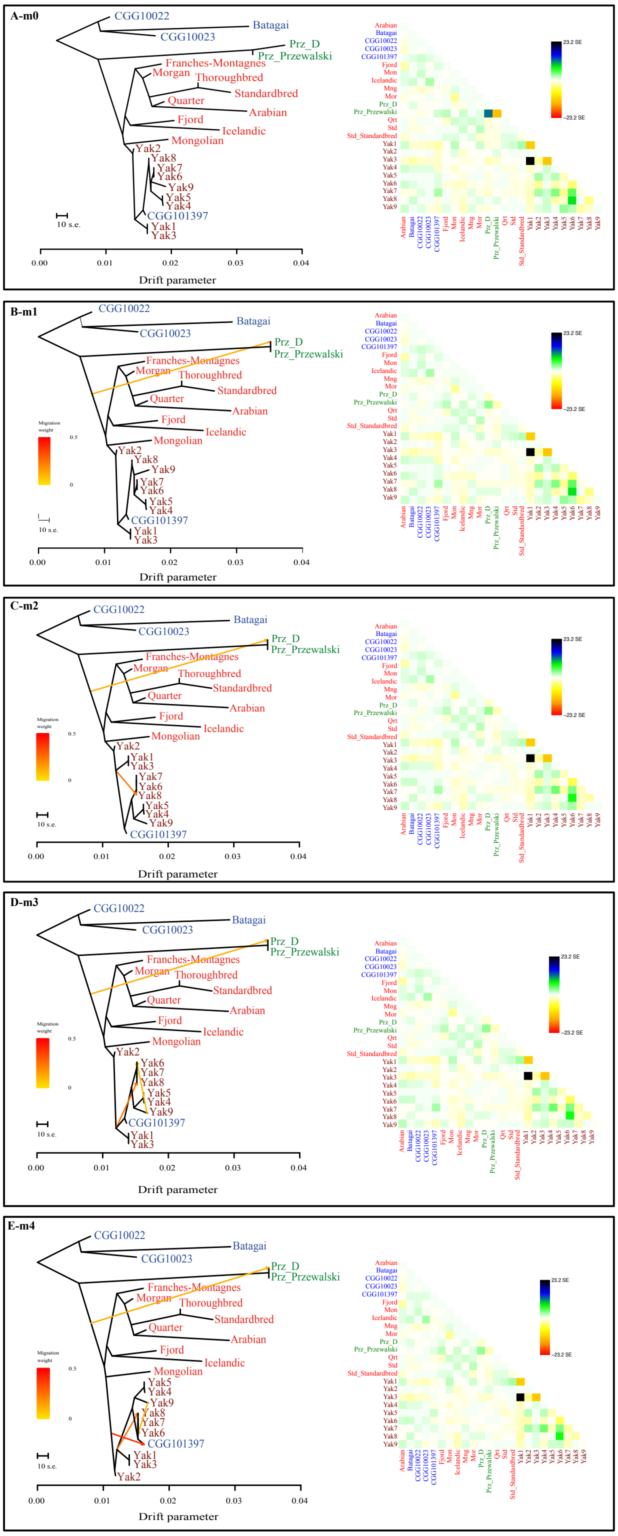


**Figure S6.11. TreeMix population splits between modern and ancient horses.**
Left: population relationships as inferred from Treemix analyses, considering 0-4 possible migration edges (m0-m4). Right: Residuals of the covariance matrix. Individuals belonging to the same domesticated breed were grouped together, and branch labels are provided in **Table S6.4**. Proportion of the variance explained by the tree: A. 98.6669%; B. 98.8381%; C. 98.9042%; D. 98.9049%; E. 98.9270%.

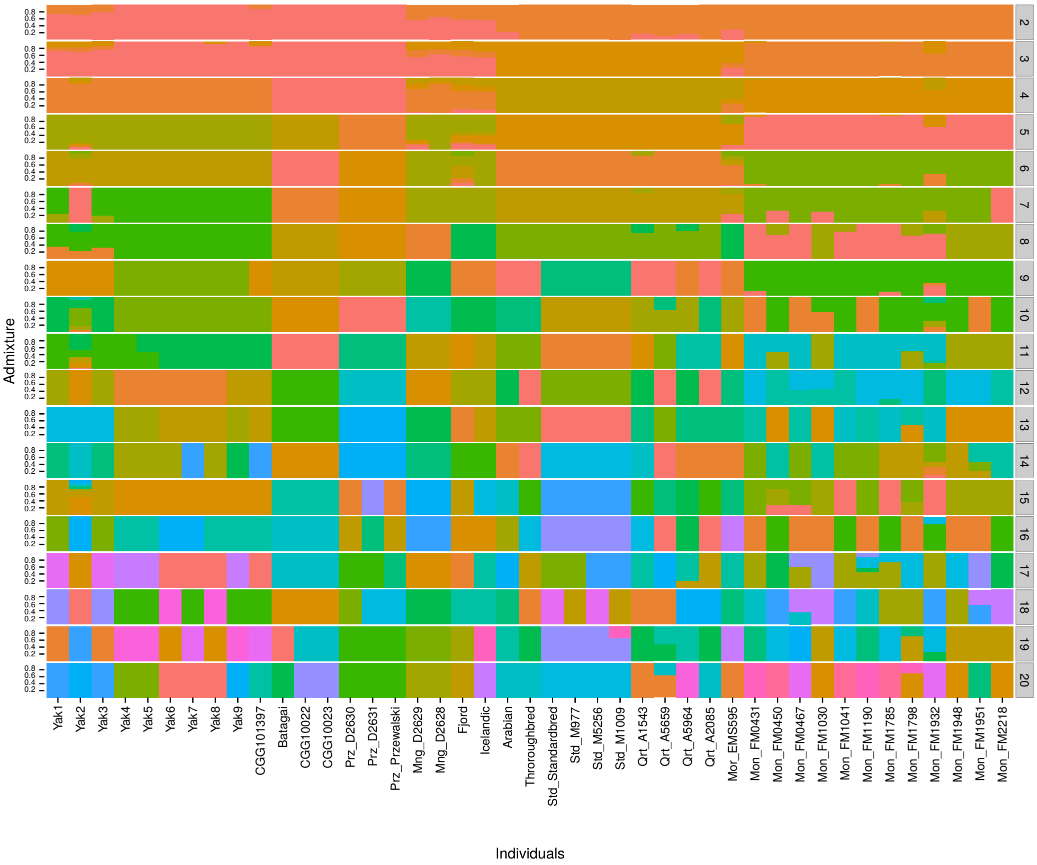


**Figure S6.12. Admixture plot representing from *K*=2 to 20 ancestry components.**

## *6.5 Supplementary Tables for Section 6*

### Table S6.1. Ancestral population size changes estimated from PSMC analyses, used as non-free (fixed) parameters in the estimation of recent history

| Time (yBP) | Effective population size* |
|---:|---:|
| 8509835 | 136338 |
| 6009540 | 142062 |
| 5047854 | 129485 |
| 4238467 | 105545 |
| 3557261 | 80974 |
| 2983935 | 66227 |
| 2501405 | 64800 |
| 2095293 | 76820 |
| 1753495 | 101314 |
| 1465827 | 129322 |
| 1223716 | 131287 |
| 1019947 | 89860 |
| 848449 | 43198 |
| 704110 | 20871 |
| 582631 | 16067 |
| 480389 | 23631 |
| 394339 | 55064 |
| 321917 | 135346 |
| 260963 | 221906 |
| 209664 | 219181 |
| 166488 | 151155 |
| 130149 | 110403 |
| 99566 | 128007 |
| 73826 | 174071 |
| 52162 | 147017 |
| 33930 | 146775 |
| 18584 | n.a.[#] |

[#] parameter to be estimated

**Table S6.2. Demographic parameters estimated using *dadi* and the reference, Mongolian, and donkey genome sequence as ancestral state.**

| Outgroup | Parameters | Model PSMC | Model one epoch | Model three epochs |
|---|---|---|---|---|
| **Throroughbred (EquCab2.0 Reference sequence)** | Log-likelihood | -867.86 | -124.25 | -121.05 |
| | $N_1$ | 5,107* | 16,352 | 48,152 |
| | $N_2$ | - | 3,072 | 17,878 |
| | $N_3$ | - | - | <15 |
| | $N_4$ | - | - | 6,582 |
| | $P_{misid}$ | <0.001 | 0.012 | 0.012 |
| **Mng_D2628 (Mongolian)** | Log-likelihood | -872.43 | -126.49 | -123.39 |
| | $N_1$ | 5,107* | 16,339 | 47,958 |
| | $N_2$ | - | 3,059 | 17,763 |
| | $N_3$ | - | - | <15 |
| | $N_4$ | - | - | 7,968 |
| | $P_{misid}$ | <0.001 | 0.0093 | 0.0097 |
| ***Equus africanus somaliensis* (donkey)** | Log-likelihood | -872.43 | -126.49 | -123.39 |
| | $N_1$ | 5,107* | 16,339 | 47,958 |
| | $N_2$ | - | 3,059 | 17,763 |
| | $N_3$ | - | - | <15 |
| | $N_4$ | - | - | 7,968 |
| | $P_{misid}$ | <0.001 | 0.0093 | 0.0097 |

* parameter set as fixed value

**Table S6.3. Likelihood surface assuming different values for the founder effect.**

| Data set | Founder time (yBP) | Reference | Mongolian | Donkey |
|---|---:|---:|---:|---:|
| 4d-fold sites | 5,000 | -127.17 | -129.48 | -119.55 |
| | 4,000 | -124.94 | -127.25 | -116.67 |
| | 3,000 | -123.20 | -125.52 | -114.44 |
| | 2,000 | -121.95 | -124.28 | -112.84 |
| | 1,000 | -121.18 | -123.52 | -111.84 |
| | 500 | -120.98 | -123.32 | -111.57 |
| | 200 | -120.90 | -123.24 | -111.47 |
| Genome-wide SFS | 5,000 | -16,183.96 | -14,444.93 | -13,312.14 |
| | 4,000 | -15,531.13 | -13,757.58 | -12,390.05 |
| | 3,000 | -15,034.58 | -13,225.16 | -11,663.36 |
| | 2,000 | -14,691.63 | -12,845.11 | -11,125.64 |
| | 1,000 | -14,499.94 | -12,615.29 | -10,771.32 |
| | 500 | -14,460.18 | -12,556.14 | -10,661.38 |
| | 200 | -14,454.12 | -12,538.36 | -10,616.59 |

## Table S6.4. Horse genomes used for projection analyses.

| Horse ID | Horse type | Reference | Reference panel(s) |
|---|---|---|---|
| Yak1 | Yakutian | this study | refYAK |
| Yak2 | Yakutian | this study | refYAK |
| Yak3 | Yakutian | this study | refYAK |
| Yak4 | Yakutian | this study | refYAK |
| Yak5 | Yakutian | this study | refYAK |
| Yak6 | Yakutian | this study | refYAK |
| Yak7 | Yakutian | this study | refYAK |
| Yak8 | Yakutian | this study | refYAK |
| Yak9 | Yakutian | this study | refYAK |
| CGG10022 | Ancient Taymyr peninsula | Orlando et al. 2013 (13) | n/a |
| CGG10023 | Ancient Taymyr peninsula | Schubert et al. 2014 (16) | n/a |
| CGG101397 | Ancient Yakutia | this study | n/a |
| Batagai | Ancient Yakutia | this study | n/a |
| Prz_D2630 | Przewalski | Do et al. 2014 (19) | n/a |
| Prz_D2631 | Przewalski | Do et al. 2014 (19) | n/a |
| Prz_Przewalski | Przewalski | Orlando et al. 2013 (13) | n/a |
| Arabian | Arabian | Orlando et al. 2013 (13) | refDOM |
| Fjord | Norwegian Fjord | Orlando et al. 2013 (13) | refDOM |
| Icelandic | Icelandic | Orlando et al. 2013 (13) | refDOM |
| Mng_D2628 | Mongolian | Do et al. 2014 (19) | refDOM |
| Mng_D2629 | Mongolian | Do et al. 2014 (19) | refDOM |
| Mor_EMS595 | Morgan | this study | refDOM |
| Qrt_A5964 | Quarter | this study | refDOM |
| Qrt_A5659 | Quarter | this study | refDOM |
| Qrt_A1543 | Quarter | this study | refDOM |
| Qrt_A2085 | Quarter | this study | refDOM |
| Std_M977 | Standardbred | this study | refDOM |
| Std_M5256 | Standardbred | this study | refDOM |
| Std_M1009 | Standardbred | this study | refDOM |
| Std_Standardbred | Standardbred | Orlando et al. 2013 (13) | refDOM |
| Thoroughbred | Thoroughbred | Wade et al. 2009 (29) | refDOM |
| Mon_FM1798 | Franches-Montagnes | this study | refDOM, refFM |
| Mon_FM1932 | Franches-Montagnes | this study | refDOM, refFM |
| Mon_FM1948 | Franches-Montagnes | this study | refDOM, refFM |
| Mon_FM2218 | Franches-Montagnes | this study | refDOM, refFM |
| Mon_FM1190 | Franches-Montagnes | this study | refDOM, refFM |
| Mon_FM1041 | Franches-Montagnes | this study | refDOM, refFM |
| Mon_FM1951 | Franches-Montagnes | this study | refDOM, refFM |
| Mon_FM0467 | Franches-Montagnes | this study | refDOM, refFM |
| Mon_FM_431 | Franches-Montagnes | this study | refDOM, refFM |
| Mon_FM1785 | Franches-Montagnes | this study | refDOM, refFM |
| Mon_FM1030 | Franches-Montagnes | this study | refDOM, refFM |
| Mon_FM0450 | Franches-Montagnes | this study | refDOM, refFM |

**Table S6.5. Minimum projection values.**

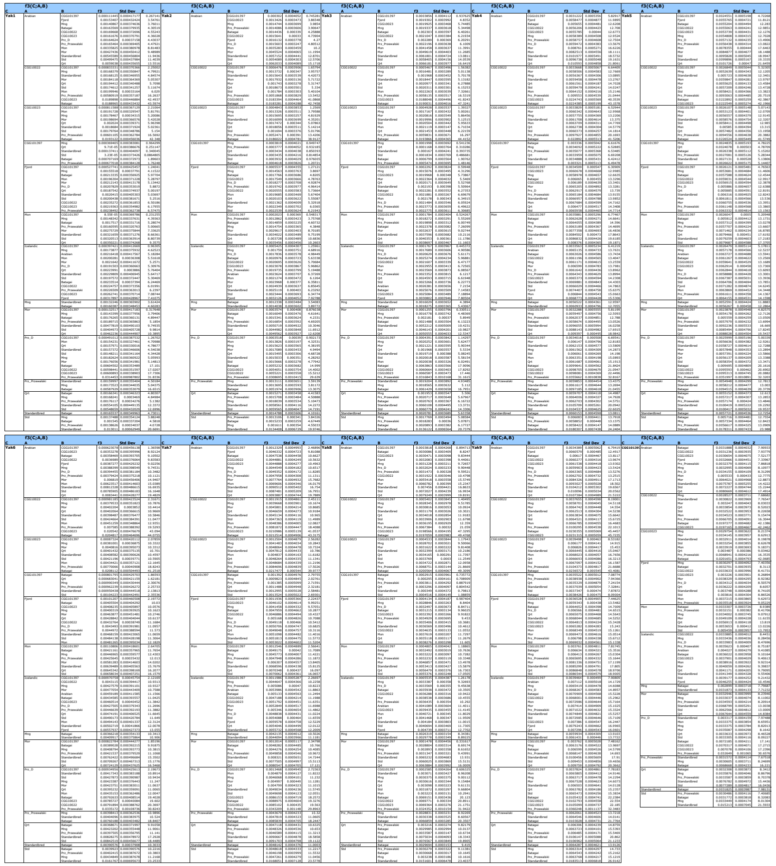
| | refYAK | | refFM | |
|---|---|---|---|---|
| **Test genome(s)** | **MPV** | **sd** | **MPV** | **sd** |
| Yakutian horses | * | * | 0.7724 | 0.0092 |
| Przewalski's horses | 0.6926 | 0.0009 | 0.6762 | 0.0109 |
| Non-Franches-Montagnes | 0.8073 | 0.0246 | 0.8088 | 0.0283 |
| Franches-Montagnes | 0.8038 | 0.0062 | * | * |
| CGG101397 | 0.8719 | n/a | 0.7637 | n/a |
| Batagai | 0.5967 | n/a | 0.5943 | n/a |
| CGG10022 | 0.6067 | n/a | 0.6010 | n/a |
| CGG10023 | 0.5898 | n/a | 0.5913 | n/a |
| Ancient horses (excl. CGG101397) | 0.5977 | 0.0085 | 0.5955 | 0.0050 |

"MPV": Minimum Projection Value; "sd": standard deviation.

## Table S6.6. Samples used for the *f3*-statistics and TreeMix analyses.

| Group | Samples | Horse type |
|---|---|---|
| Batagai | Batagai | Ancient |
| CGG10022 | CGG10022 | |
| CGG10023 | CGG10023 | |
| CGG101397 | CGG101397 | |
| Icelandic | Icelandic | Domesticated |
| Mng | Mng_D2629 | |
| | Mng_D2628 | |
| Mon | Mon_FM1932 | |
| | Mon_FM1951 | |
| | Mon_FM1785 | |
| | Mon_FM1030 | |
| Mor | Mor_EMS595 | |
| Qrt | Qrt_A5964 | |
| | Qrt_A5659 | |
| Std | Std_M977 | |
| | Std_M5256 | |
| Std_Standardbred | Std_Standardbred | |
| Arabian | Arabian | |
| Fjord | Fjord | |
| Prz_D | Prz_D2630 | Przewalski |
| | Prz_D2631 | |
| Prz_Przewalski | Prz_Przewalski | |
| Yak | Yak1 | Yakutian |
| | Yak2 | |
| | Yak3 | |
| | Yak4 | |
| | Yak5 | |
| | Yak6 | |
| | Yak7 | |
| | Yak8 | |
| | Yak9 | |

**Table S6.7. *$f_3$*-statistics for topologies where ancient specimens are considered as population C.** None *$f_3$* statistics are significantly negative.



| | f3(C;A,B) | | | | |
|---|---|---|---|---|---|
| **C** | **A** | **B** | **f3** | **Std Dev** | **Z** |
| **Yak1** | Arabian | CGG101397 | 0.000111445 | 0.000417177 | 0.267141 |
| | | Fjord | 0.00153407 | 0.000432424 | 3.54761 |
| | | Mng | 0.00140867 | 0.000374636 | 3.76011 |
| | | Batagai | 0.00163569 | 0.000379935 | 4.30519 |
| | | CGG10022 | 0.00169668 | 0.000372696 | 4.55243 |
| | | CGG10023 | 0.00161676 | 0.000370791 | 4.36028 |
| | | Prz_D | 0.00164624 | 0.00037158 | 4.43037 |
| | | Prz_Przewalski | 0.00165464 | 0.000369405 | 4.47921 |
| | | Mon | 0.00335825 | 0.000380978 | 8.81483 |
| | | Mor | 0.00427436 | 0.000450412 | 9.48989 |
| | | Standardbred | 0.00545589 | 0.000456804 | 11.9436 |
| | | Std | 0.00499475 | 0.000437984 | 11.4039 |
| | | Qrt | 0.0059038 | 0.000435655 | 13.5516 |
| | CGG10022 | CGG101397 | 0.000833331 | 0.000370366 | 2.25002 |
| | | Mng | 0.00168379 | 0.000350847 | 4.79921 |
| | | Standardbred | 0.00168125 | 0.000346955 | 4.84574 |
| | | Mor | 0.00184118 | 0.000364369 | 5.05307 |
| | | Qrt | 0.00194412 | 0.000340488 | 5.7098 |
| | | Std | 0.00174612 | 0.000341257 | 5.11674 |
| | | Mon | 0.00199946 | 0.00033164 | 6.029 |
| | | Prz_Przewalski | 0.00580919 | 0.000357187 | 16.2637 |
| | | CGG10023 | 0.0189898 | 0.000446884 | 42.4937 |
| | | Batagai | 0.0188965 | 0.000433432 | 43.5974 |
| | CGG10023 | CGG101397 | 0.000811588 | 0.000367129 | 2.21064 |
| | | Mon | 0.00191728 | 0.000329547 | 5.81791 |
| | | Mng | 0.00178467 | 0.00034315 | 5.20086 |
| | | Mor | 0.00198694 | 0.000366576 | 5.42028 |
| | | Qrt | 0.002024 | 0.000339371 | 5.96398 |
| | | Standardbred | 0.0018684 | 0.000371924 | 5.02361 |
| | | Std | 0.00179764 | 0.000348786 | 5.154 |
| | | Prz_Przewalski | 0.00601105 | 0.000362784 | 16.5692 |
| | | Batagai | 0.0193122 | 0.000456036 | 42.348 |
| | CGG101397 | Mng | 0.000369405 | 0.000383081 | 0.964299 |
| | | Qrt | 9.71E-05 | 0.000386676 | 0.251147 |
| | | Standardbred | 0.000137411 | 0.000404097 | 0.340045 |
| | | Std | 3.23E-05 | 0.000373426 | 0.0864805 |
| | | Batagai | 0.000707169 | 0.000372973 | 1.89603 |
| | | Prz_Przewalski | 0.000677018 | 0.000384128 | 1.76248 |
| | Fjord | CGG101397 | 0.000527741 | 0.000425457 | 1.24041 |
| | | Mng | 0.00155518 | 0.00037791 | 4.11522 |
| | | Batagai | 0.00213335 | 0.000356925 | 5.97704 |
| | | Mon | 0.00196304 | 0.000371228 | 5.28797 |
| | | Mor | 0.00212143 | 0.000413178 | 5.13442 |
| | | Prz_D | 0.00207829 | 0.000353019 | 5.8872 |
| | | Qrt | 0.00187542 | 0.000374937 | 5.00197 |
| | | Standardbred | 0.0020415 | 0.000405303 | 5.03697 |
| | | Std | 0.00200438 | 0.000381671 | 5.2516 |
| | | CGG10022 | 0.00235272 | 0.000361853 | 6.50186 |
| | | CGG10023 | 0.00219363 | 0.000354982 | 6.17955 |
| | | Prz_Przewalski | 0.00238492 | 0.000362738 | 6.57475 |
| | Mon | CGG101397 | 8.55E-05 | 0.000369786 | 0.231255 |
| | | Mng | 0.00148342 | 0.000337631 | 4.39363 |
| | | Batagai | 0.0017517 | 0.000331716 | 5.28074 |
| | | Prz_Przewalski | 0.00160595 | 0.000320763 | 5.00665 |
| | | Mor | 0.00275729 | 0.000375844 | 7.33625 |
| | | Standardbred | 0.00321659 | 0.000371276 | 8.66361 |
| | | Std | 0.00320513 | 0.000363914 | 8.80739 |
| | | Qrt | 0.00350221 | 0.000374268 | 9.3575 |
| | Icelandic | CGG101397 | 0.000397421 | 0.000412669 | 0.96305 |
| | | Mng | 0.00173877 | 0.000375932 | 4.62523 |
| | | Arabian | 0.00216418 | 0.000417567 | 5.18284 |
| | | Mon | 0.00200281 | 0.00036308 | 5.51618 |
| | | Mor | 0.0021642 | 0.000411672 | 5.2571 |
| | | Prz_D | 0.00191503 | 0.000360663 | 5.30976 |
| | | Qrt | 0.00223991 | 0.0003886 | 5.76404 |
| | | Standardbred | 0.00229809 | 0.000406945 | 5.64717 |
| | | Std | 0.00197572 | 0.000372447 | 5.30469 |
| | | Batagai | 0.00233514 | 0.00037394 | 6.24469 |
| | | CGG10022 | 0.00224757 | 0.000373356 | 6.01991 |
| | | Prz_Przewalski | 0.00226509 | 0.000363013 | 6.2397 |
| | | CGG10023 | 0.00262741 | 0.000370714 | 7.08744 |
| | | Fjord | 0.00317897 | 0.000428967 | 7.41075 |
| | Mng | Standardbred | 0.00132246 | 0.000365902 | 3.61424 |
| | | Batagai | 0.00160387 | 0.000348453 | 4.60283 |
| | Mor | CGG101397 | 0.000132517 | 0.000417183 | 0.317648 |
| | | Mng | 0.00143399 | 0.000377956 | 3.79406 |
| | | Batagai | 0.00176265 | 0.000360131 | 4.89447 |
| | | Prz_Przewalski | 0.00189715 | 0.000365863 | 5.1854 |
| | | Standardbred | 0.00477819 | 0.000490103 | 9.74935 |
| | | Std | 0.00404973 | 0.000405728 | 9.9814 |
| | | Qrt | 0.00462236 | 0.000444907 | 10.3895 |
| | Prz_D | CGG101397 | 0.000491012 | 0.000387532 | 1.26702 |
| | | Mon | 0.00154231 | 0.000327461 | 4.70988 |
| | | Qrt | 0.00157971 | 0.000330016 | 4.78677 |
| | | Standardbred | 0.00157372 | 0.000346606 | 4.54038 |
| | | Std | 0.00148211 | 0.000341164 | 4.34428 |
| | | Mng | 0.00182624 | 0.000360922 | 5.05993 |
| | | Mor | 0.00176056 | 0.000341981 | 5.14813 |
| | | Batagai | 0.00604502 | 0.000353489 | 17.101 |
| | | CGG10022 | 0.00598441 | 0.000351597 | 17.0207 |
| | | CGG10023 | 0.00600893 | 0.000338903 | 17.7306 |
| | | Prz_Przewalski | 0.0314453 | 0.000987802 | 31.8336 |
| | Prz_Przewalski | Standardbred | 0.00159997 | 0.000355404 | 4.50184 |
| | | Mng | 0.00173523 | 0.000344035 | 5.04375 |
| | | Batagai | 0.00587929 | 0.000353079 | 16.6515 |
| | Qrt | Batagai | 0.00166341 | 0.000346323 | 4.80307 |
| | | Mng | 0.00168241 | 0.0003469 | 4.84984 |
| | | Prz_Przewalski | 0.00176117 | 0.00034276 | 5.1382 |
| | | Standardbred | 0.00554105 | 0.000449135 | 12.3372 |
| | | Std | 0.00548659 | 0.000432029 | 12.6996 |
| | Standardbred | Batagai | 0.00165537 | 0.000349963 | 4.73011 |
| | Std | Mng | 0.00137488 | 0.000354124 | 3.88248 |
| | | Batagai | 0.001545 | 0.000337204 | 4.5818 |
| | | Prz_Przewalski | 0.00138626 | 0.00034037 | 4.0728 |
| | | Standardbred | 0.013011 | 0.00062949 | 20.6691 |
| **Yak2** | Arabian | CGG101397 | 0.000302 | 0.0004052 | 0.745281 |
| | | CGG10023 | 0.0013426 | 0.0003473 | 3.86548 |
| | | Mng | 0.0014744 | 0.0003699 | 3.9854 |
| | | Prz_Przewalski | 0.0014086 | 0.0003606 | 3.90647 |
| | | Batagai | 0.0014436 | 0.000339 | 4.25888 |
| | | CGG10022 | 0.0015641 | 0.00033 | 4.73904 |
| | | Fjord | 0.0016132 | 0.0003778 | 4.27 |
| | | Prz_D | 0.0015603 | 0.0003247 | 4.80512 |
| | | Mon | 0.0035283 | 0.0003459 | 10.2 |
| | | Mor | 0.0045524 | 0.0004065 | 11.1994 |
| | | Standardbred | 0.0057152 | 0.0004441 | 12.8701 |
| | | Std | 0.0054089 | 0.0004303 | 12.5706 |
| | | Qrt | 0.0062033 | 0.0004089 | 15.1718 |
| | CGG10022 | CGG101397 | 0.0006476 | 0.0003906 | 1.65794 |
| | | Mng | 0.0013732 | 0.000347 | 3.95757 |
| | | Standardbred | 0.0015643 | 0.0003539 | 4.42073 |
| | | Mon | 0.0017932 | 0.0003136 | 5.71722 |
| | | Mor | 0.001743 | 0.0003278 | 5.31747 |
| | | Qrt | 0.0018673 | 0.0003501 | 5.334 |
| | | Std | 0.001784 | 0.0003303 | 5.40104 |
| | | Prz_Przewalski | 0.0051868 | 0.0003829 | 13.5452 |
| | | CGG10023 | 0.0183394 | 0.0004464 | 41.0868 |
| | | Batagai | 0.0183281 | 0.0004288 | 42.7458 |
| | CGG10023 | CGG101397 | 0.0004843 | 0.0003853 | 1.2569 |
| | | Mng | 0.0013326 | 0.0003511 | 3.79588 |
| | | Mon | 0.0015695 | 0.0003257 | 4.81929 |
| | | Standardbred | 0.0016099 | 0.0003699 | 4.35201 |
| | | Mor | 0.0017472 | 0.000344 | 5.07863 |
| | | Qrt | 0.0018056 | 0.0003511 | 5.14214 |
| | | Std | 0.001694 | 0.0003376 | 5.01796 |
| | | Prz_Przewalski | 0.0052471 | 0.000391 | 13.4206 |
| | | Batagai | 0.0186022 | 0.0004781 | 38.9127 |
| | CGG101397 | Mng | 0.0003819 | 0.0004021 | 0.949747 |
| | | Prz_Przewalski | 0.0003777 | 0.0004052 | 0.932185 |
| | | Qrt | 0.0003434 | 0.0004038 | 0.850393 |
| | | Standardbred | 0.0003435 | 0.0004216 | 0.814854 |
| | | Std | 0.0003932 | 0.0004029 | 0.976005 |
| | | Batagai | 0.0004618 | 0.0003826 | 1.20721 |
| | Fjord | CGG101397 | 0.0005537 | 0.0004355 | 1.27121 |
| | | Mng | 0.0014563 | 0.0003763 | 3.8697 |
| | | Batagai | 0.0017766 | 0.0003686 | 4.8205 |
| | | CGG10023 | 0.0017549 | 0.0003666 | 4.78763 |
| | | Prz_D | 0.0018277 | 0.0003708 | 4.92874 |
| | | Prz_Przewalski | 0.0019742 | 0.0003977 | 4.96414 |
| | | CGG10022 | 0.0020555 | 0.0003583 | 5.73664 |
| | | Mon | 0.0019685 | 0.0003469 | 5.67404 |
| | | Qrt | 0.0020103 | 0.0003622 | 5.55087 |
| | | Standardbred | 0.0021362 | 0.0004009 | 5.32918 |
| | | Mor | 0.0022349 | 0.0003702 | 6.0365 |
| | | Std | 0.0022539 | 0.0003624 | 6.21943 |
| | Mon | CGG101397 | 0.0002023 | 0.000369 | 0.548317 |
| | | Prz_Przewalski | 0.0012862 | 0.0003423 | 3.75768 |
| | | Batagai | 0.0014859 | 0.0003225 | 4.60732 |
| | | Mng | 0.0014754 | 0.0003365 | 4.3848 |
| | | Mor | 0.0029617 | 0.0003403 | 8.70185 |
| | | Standardbred | 0.0034022 | 0.0003489 | 9.75199 |
| | | Qrt | 0.003728 | 0.0003489 | 10.6836 |
| | | Std | 0.0035456 | 0.0003456 | 10.2603 |
| | Icelandic | CGG101397 | 0.0005425 | 0.0004307 | 1.25961 |
| | | Mng | 0.001759 | 0.0003751 | 4.68916 |
| | | Prz_D | 0.0017836 | 0.0003578 | 4.98449 |
| | | Batagai | 0.0020976 | 0.0003723 | 5.63338 |
| | | CGG10022 | 0.0020695 | 0.0003626 | 5.70684 |
| | | CGG10023 | 0.0023078 | 0.0003867 | 5.96862 |
| | | Prz_Przewalski | 0.0019735 | 0.0003799 | 5.19488 |
| | | Arabian | 0.0023624 | 0.0003707 | 6.37269 |
| | | Mon | 0.0021274 | 0.0003472 | 6.1264 |
| | | Mor | 0.0023968 | 0.000377 | 6.35811 |
| | | Qrt | 0.0024939 | 0.0003637 | 6.85645 |
| | | Standardbred | 0.0025119 | 0.000403 | 6.23292 |
| | | Std | 0.0023444 | 0.0003693 | 6.34774 |
| | | Fjord | 0.0032126 | 0.0004052 | 7.92788 |
| | Mng | Batagai | 0.0012338 | 0.0003484 | 3.54083 |
| | | Standardbred | 0.0014038 | 0.0003602 | 3.89772 |
| | Mor | CGG101397 | 0.0003574 | 0.000408 | 0.876038 |
| | | Batagai | 0.0016049 | 0.0003476 | 4.61641 |
| | | Mng | 0.0015341 | 0.0003624 | 4.2331 |
| | | Prz_Przewalski | 0.0016854 | 0.0003592 | 4.69205 |
| | | Standardbred | 0.0050719 | 0.0004922 | 10.3046 |
| | | Std | 0.0044982 | 0.0003848 | 11.6912 |
| | | Qrt | 0.0049562 | 0.0003927 | 12.6208 |
| | Prz_D | CGG101397 | 0.0003518 | 0.00039 | 0.902173 |
| | | Mon | 0.0013826 | 0.0003197 | 4.32531 |
| | | Mng | 0.0015623 | 0.0003565 | 4.38195 |
| | | Mor | 0.0017089 | 0.0003453 | 4.9494 |
| | | Qrt | 0.0015495 | 0.0003306 | 4.68728 |
| | | Standardbred | 0.0015033 | 0.000351 | 4.28292 |
| | | Std | 0.0015666 | 0.0003278 | 4.77942 |
| | | Batagai | 0.0055232 | 0.0003682 | 14.999 |
| | | CGG10023 | 0.0054051 | 0.0003754 | 14.4002 |
| | | CGG10022 | 0.0055221 | 0.0003558 | 15.5212 |
| | | Prz_Przewalski | 0.0308695 | 0.0010415 | 29.639 |
| | Prz_Przewalski | Mng | 0.0013111 | 0.0003651 | 3.59139 |
| | | Standardbred | 0.0013695 | 0.0003593 | 3.81172 |
| | | Batagai | 0.0051974 | 0.0003906 | 13.3075 |
| | Qrt | Batagai | 0.0015271 | 0.0003458 | 4.41583 |
| | | Prz_Przewalski | 0.0015708 | 0.0003484 | 4.50888 |
| | | Mng | 0.0018039 | 0.0003534 | 5.10473 |
| | | Standardbred | 0.0058561 | 0.0004116 | 14.2272 |
| | | Std | 0.0059565 | 0.0004047 | 14.7191 |
| | Standardbred | Batagai | 0.0014788 | 0.0003606 | 4.10161 |
| | Std | Prz_Przewalski | 0.0013106 | 0.000341 | 3.84326 |
| | | Batagai | 0.0015233 | 0.0003259 | 4.67449 |
| | | Mng | 0.001611 | 0.000354 | 4.55033 |
| | | Standardbred | 0.0134408 | 0.0006729 | 19.9746 |
| **Yak3** | Arabian | CGG101397 | 0.0003856 | 0.0004112 | 0.937625 |
| | | Fjord | 0.0019302 | 0.0003992 | 4.8352 |
| | | CGG10023 | 0.0019925 | 0.0003468 | 5.74485 |
| | | Mng | 0.0019337 | 0.0003618 | 5.34498 |
| | | Batagai | 0.002303 | 0.0003597 | 6.40261 |
| | | CGG10022 | 0.0021047 | 0.0003384 | 6.21934 |
| | | Prz_D | 0.002289 | 0.000369 | 6.20291 |
| | | Prz_Przewalski | 0.0022196 | 0.0003638 | 6.1009 |
| | | Mon | 0.0041458 | 0.0003637 | 11.3991 |
| | | Mor | 0.0049619 | 0.0004403 | 11.2685 |
| | | Standardbred | 0.0061801 | 0.0004724 | 13.0811 |
| | | Std | 0.0058405 | 0.0004156 | 14.0539 |
| | | Qrt | 0.0066181 | 0.0003977 | 16.6419 |
| | CGG10022 | CGG101397 | 0.0005467 | 0.0003496 | 1.56382 |
| | | Mng | 0.0016481 | 0.0003289 | 5.01138 |
| | | Mor | 0.001968 | 0.0003452 | 5.70178 |
| | | Standardbred | 0.0018447 | 0.0003595 | 5.13182 |
| | | Qrt | 0.0020977 | 0.0003341 | 6.27888 |
| | | Std | 0.0020311 | 0.0003301 | 6.15253 |
| | | Mon | 0.0022263 | 0.0003039 | 7.32661 |
| | | Prz_Przewalski | 0.0058135 | 0.0003517 | 16.5319 |
| | | CGG10023 | 0.0188048 | 0.0004496 | 41.8243 |
| | | Batagai | 0.0190031 | 0.0004016 | 47.3241 |
| | CGG10023 | CGG101397 | 0.0004928 | 0.0003537 | 1.39327 |
| | | Mng | 0.0017167 | 0.0003432 | 5.00261 |
| | | Mor | 0.0020816 | 0.0003549 | 5.86456 |
| | | Standardbred | 0.0019996 | 0.0003902 | 5.12515 |
| | | Std | 0.0020504 | 0.0003442 | 5.95633 |
| | | Mon | 0.0021119 | 0.0003147 | 6.71034 |
| | | Qrt | 0.0021453 | 0.0003448 | 6.22159 |
| | | Prz_Przewalski | 0.0059831 | 0.0003671 | 16.297 |
| | | Batagai | 0.0193866 | 0.0004365 | 44.4152 |
| | CGG101397 | Mng | 0.0001998 | 0.0003692 | 0.541236 |
| | | Qrt | 0.0001168 | 0.0003764 | 0.310286 |
| | | Standardbred | 0.000167 | 0.0004241 | 0.393735 |
| | | Std | 0.0001834 | 0.0003769 | 0.486681 |
| | | Batagai | 0.0006799 | 0.0003564 | 1.90762 |
| | | Prz_Przewalski | 0.0005474 | 0.0003695 | 1.48146 |
| | Fjord | CGG101397 | 0.0002292 | 0.0003824 | 0.59948 |
| | | Mng | 0.0015076 | 0.0003495 | 4.31296 |
| | | CGG10023 | 0.0019968 | 0.000348 | 5.73867 |
| | | Mor | 0.0022364 | 0.0004032 | 5.5469 |
| | | Qrt | 0.0020171 | 0.0003649 | 5.52768 |
| | | Standardbred | 0.002193 | 0.000398 | 5.50964 |
| | | Batagai | 0.0022281 | 0.0003552 | 6.27353 |
| | | CGG10022 | 0.0021881 | 0.0003267 | 6.69679 |
| | | Mon | 0.002178 | 0.000343 | 6.34915 |
| | | Prz_D | 0.0021484 | 0.0003546 | 6.05924 |
| | | Prz_Przewalski | 0.0023773 | 0.0003659 | 6.49622 |
| | | Std | 0.0022775 | 0.0003611 | 6.30685 |
| | Mon | CGG101397 | 0.0001784 | 0.0003404 | 0.524267 |
| | | Mng | 0.0018272 | 0.0003291 | 5.55269 |
| | | Prz_Przewalski | 0.0019898 | 0.0003312 | 6.00749 |
| | | Batagai | 0.0022378 | 0.0003082 | 7.26099 |
| | | Mor | 0.0032637 | 0.0003615 | 9.02744 |
| | | Standardbred | 0.0037596 | 0.0003682 | 10.2116 |
| | | Qrt | 0.0040353 | 0.000361 | 11.1786 |
| | | Std | 0.0038697 | 0.0003467 | 11.1603 |
| | Icelandic | CGG101397 | 0.0001767 | 0.0003967 | 0.445372 |
| | | Mng | 0.0017689 | 0.0003606 | 4.90586 |
| | | Prz_D | 0.0020629 | 0.0003676 | 5.61171 |
| | | Standardbred | 0.0025274 | 0.0004234 | 5.96881 |
| | | CGG10022 | 0.0021607 | 0.0003339 | 6.47177 |
| | | Mon | 0.0022955 | 0.0003413 | 6.72565 |
| | | Mor | 0.0023569 | 0.0003873 | 6.08567 |
| | | Prz_Przewalski | 0.0023352 | 0.0003815 | 6.1217 |
| | | Qrt | 0.0024593 | 0.0003715 | 6.61948 |
| | | Std | 0.0023266 | 0.0003736 | 6.22773 |
| | | Arabian | 0.0026381 | 0.0003656 | 7.2154 |
| | | Batagai | 0.0025076 | 0.0003509 | 7.14583 |
| | | CGG10023 | 0.0025084 | 0.0003489 | 7.18894 |
| | | Fjord | 0.0030802 | 0.0003946 | 7.80504 |
| | Mng | Standardbred | 0.0016029 | 0.0003652 | 4.3894 |
| | | Batagai | 0.0018275 | 0.0003365 | 5.43067 |
| | Mor | CGG101397 | 0.0001255 | 0.0003796 | 0.330511 |
| | | Mng | 0.0016778 | 0.0003742 | 4.48369 |
| | | Prz_Przewalski | 0.002181 | 0.00037 | 5.8949 |
| | | Batagai | 0.0021488 | 0.0003504 | 6.13223 |
| | | Standardbred | 0.0052212 | 0.0005009 | 10.4231 |
| | | Std | 0.0046143 | 0.0004201 | 10.9827 |
| | | Qrt | 0.0050555 | 0.0004033 | 12.5347 |
| | Prz_D | CGG101397 | 0.0004391 | 0.0003842 | 1.14297 |
| | | Mng | 0.0020252 | 0.0003601 | 5.62477 |
| | | Mor | 0.0021221 | 0.0003595 | 5.90344 |
| | | Qrt | 0.001968 | 0.0003557 | 5.5334 |
| | | Standardbred | 0.0019719 | 0.000388 | 5.08245 |
| | | Std | 0.0020018 | 0.0003587 | 5.58154 |
| | | Mon | 0.0020038 | 0.0003326 | 6.02491 |
| | | Batagai | 0.0063863 | 0.0003566 | 17.9096 |
| | | CGG10022 | 0.0060664 | 0.0003403 | 17.8292 |
| | | CGG10023 | 0.0060587 | 0.0003473 | 17.446 |
| | | Prz_Przewalski | 0.0316843 | 0.0010168 | 31.1615 |
| | Prz_Przewalski | Standardbred | 0.0019204 | 0.0003892 | 4.93485 |
| | | Mng | 0.0018565 | 0.0003632 | 5.112 |
| | | Batagai | 0.0061429 | 0.0003658 | 16.7951 |
| | Qrt | Mng | 0.001953 | 0.0003547 | 5.506 |
| | | Prz_Przewalski | 0.0020717 | 0.0003648 | 5.67967 |
| | | Batagai | 0.0020763 | 0.0003367 | 6.16722 |
| | | Standardbred | 0.0060108 | 0.0004495 | 13.3731 |
| | | Std | 0.0060779 | 0.0004241 | 14.3306 |
| | Standardbred | Batagai | 0.0020781 | 0.0003689 | 5.63358 |
| | Std | Mng | 0.0017769 | 0.0003494 | 5.08584 |
| | | Prz_Przewalski | 0.0018283 | 0.00036 | 5.07871 |
| | | Batagai | 0.0020893 | 0.0003382 | 6.17727 |
| | | Standardbred | 0.0136122 | 0.0006564 | 20.7379 |
| **Yak4** | Arabian | CGG101397 | 0.0031222 | 0.0005356 | 5.82917 |
| | | Fjord | 0.0058477 | 0.0004877 | 11.9895 |
| | | Batagai | 0.005655 | 0.0004481 | 12.6208 |
| | | CGG10022 | 0.0056206 | 0.0004423 | 12.708 |
| | | CGG10023 | 0.0055785 | 0.00044 | 12.6773 |
| | | Mng | 0.0058389 | 0.0004508 | 12.9524 |
| | | Prz_Przewalski | 0.0058757 | 0.0004608 | 12.7504 |
| | | Prz_D | 0.0059472 | 0.0004383 | 13.5679 |
| | | Mor | 0.008761 | 0.0005271 | 16.6226 |
| | | Mon | 0.0082513 | 0.0004556 | 18.1111 |
| | | Standardbred | 0.0101977 | 0.0005491 | 18.5732 |
| | | Std | 0.0096738 | 0.0005048 | 19.1631 |
| | | Qrt | 0.010593 | 0.0004858 | 21.8061 |
| | CGG10022 | CGG101397 | 0.0033668 | 0.0005067 | 6.64409 |
| | | Mor | 0.0058506 | 0.0004583 | 12.7669 |
| | | Mng | 0.0056367 | 0.0004306 | 13.0895 |
| | | Standardbred | 0.0059458 | 0.0004478 | 13.2767 |
| | | Qrt | 0.0061561 | 0.0004187 | 14.7028 |
| | | Std | 0.0059479 | 0.0004241 | 14.0247 |
| | | Mon | 0.0064152 | 0.0004216 | 15.2146 |
| | | Prz_Przewalski | 0.009553 | 0.0004539 | 21.0444 |
| | | CGG10023 | 0.0224743 | 0.0005569 | 40.3539 |
| | | Batagai | 0.0224385 | 0.0005199 | 43.1578 |
| | CGG10023 | CGG101397 | 0.0033829 | 0.0005181 | 6.52944 |
| | | Mor | 0.0060342 | 0.0004643 | 12.9968 |
| | | Mng | 0.0057755 | 0.0004369 | 13.2206 |
| | | Standardbred | 0.0061708 | 0.0004614 | 13.375 |
| | | Mon | 0.0063709 | 0.0004311 | 14.7769 |
| | | Qrt | 0.0062738 | 0.0004304 | 14.5779 |
| | | Std | 0.0060373 | 0.0004257 | 14.1814 |
| | | Prz_Przewalski | 0.0097927 | 0.0004855 | 20.1693 |
| | | Batagai | 0.022892 | 0.0005513 | 41.5206 |
| | CGG101397 | Batagai | 0.003336 | 0.0005042 | 6.61676 |
| | | Mng | 0.0034092 | 0.0005222 | 6.52885 |
| | | Prz_Przewalski | 0.0035077 | 0.0005368 | 6.53485 |
| | | Qrt | 0.0033959 | 0.0005075 | 6.69208 |
| | | Standardbred | 0.0034888 | 0.0005431 | 6.42412 |
| | | Std | 0.003321 | 0.0005113 | 6.49478 |
| | Fjord | CGG101397 | 0.0034509 | 0.000547 | 6.30909 |
| | | CGG10023 | 0.0060678 | 0.0004668 | 12.9985 |
| | | Mng | 0.0058978 | 0.0004657 | 12.6635 |
| | | Batagai | 0.006065 | 0.000455 | 13.3296 |
| | | CGG10022 | 0.006189 | 0.0004639 | 13.3401 |
| | | Mor | 0.0065205 | 0.0004892 | 13.3281 |
| | | Prz_D | 0.0062917 | 0.0004579 | 13.739 |
| | | Prz_Przewalski | 0.0065184 | 0.0004684 | 13.9155 |
| | | Standardbred | 0.0066957 | 0.0004788 | 13.9852 |
| | | Mon | 0.0067684 | 0.0004599 | 14.7162 |
| | | Qrt | 0.006477 | 0.0004517 | 14.3408 |
| | | Std | 0.0065958 | 0.0004465 | 14.7714 |
| | Mon | CGG101397 | 0.0035881 | 0.0005296 | 6.77467 |
| | | Batagai | 0.0062628 | 0.0004271 | 14.6641 |
| | | Mng | 0.0064055 | 0.0004389 | 14.596 |
| | | Prz_Przewalski | 0.0063189 | 0.0004367 | 14.4699 |
| | | Mor | 0.0077358 | 0.0004693 | 16.4836 |
| | | Standardbred | 0.0084502 | 0.000473 | 17.8643 |
| | | Qrt | 0.0086832 | 0.0004426 | 19.6185 |
| | | Std | 0.008376 | 0.0004365 | 19.1871 |
| | Icelandic | CGG101397 | 0.0033563 | 0.0005234 | 6.41219 |
| | | Arabian | 0.0065135 | 0.0004733 | 13.7631 |
| | | Batagai | 0.0063026 | 0.0004586 | 13.7434 |
| | | CGG10022 | 0.0061196 | 0.0004565 | 13.4047 |
| | | Mng | 0.0061171 | 0.0004615 | 13.2559 |
| | | Mor | 0.006599 | 0.0004783 | 13.7981 |
| | | Prz_D | 0.0061642 | 0.0004436 | 13.8962 |
| | | Prz_Przewalski | 0.0064343 | 0.0004629 | 13.8994 |
| | | CGG10023 | 0.0065373 | 0.0004594 | 14.2308 |
| | | Standardbred | 0.0069881 | 0.0004843 | 14.4303 |
| | | Std | 0.0066029 | 0.0004466 | 14.7863 |
| | | Fjord | 0.0074407 | 0.0004758 | 15.6375 |
| | | Mon | 0.006844 | 0.0004543 | 15.0663 |
| | | Qrt | 0.0068773 | 0.0004428 | 15.5306 |
| | Mng | Batagai | 0.0056522 | 0.0004361 | 12.9597 |
| | | Standardbred | 0.0060933 | 0.0004796 | 12.7055 |
| | Mor | CGG101397 | 0.0032288 | 0.0005555 | 5.81199 |
| | | Mng | 0.0059497 | 0.0004756 | 12.5093 |
| | | Prz_Przewalski | 0.0062037 | 0.0004851 | 12.788 |
| | | Batagai | 0.0058674 | 0.0004495 | 13.0526 |
| | | Standardbred | 0.0096055 | 0.0005994 | 16.0258 |
| | | Std | 0.0088143 | 0.0004982 | 17.6915 |
| | | Qrt | 0.009397 | 0.000495 | 18.9842 |
| | Prz_D | CGG101397 | 0.0034016 | 0.000508 | 6.69541 |
| | | Mor | 0.006147 | 0.0004796 | 12.8183 |
| | | Standardbred | 0.0062153 | 0.0004577 | 13.5809 |
| | | Mng | 0.0061562 | 0.0004309 | 14.2873 |
| | | Std | 0.006061 | 0.0004269 | 14.198 |
| | | Mon | 0.0063351 | 0.0004198 | 15.0893 |
| | | Qrt | 0.0061687 | 0.0004061 | 15.1913 |
| | | Batagai | 0.0099641 | 0.0004562 | 21.8424 |
| | | CGG10023 | 0.0098705 | 0.0004679 | 21.0947 |
| | | CGG10022 | 0.0098081 | 0.0004369 | 22.4496 |
| | | Prz_Przewalski | 0.0355662 | 0.0010838 | 32.8167 |
| | Prz_Przewalski | Mng | 0.0059853 | 0.0004454 | 13.4375 |
| | | Standardbred | 0.0061617 | 0.0004644 | 13.2694 |
| | | Batagai | 0.0097185 | 0.0004618 | 21.0455 |
| | Qrt | Batagai | 0.0059707 | 0.0004118 | 14.4987 |
| | | Mng | 0.0064006 | 0.0004327 | 14.7928 |
| | | Prz_Przewalski | 0.0062703 | 0.0004302 | 14.5751 |
| | | Standardbred | 0.0105709 | 0.0005261 | 20.0942 |
| | | Std | 0.0104537 | 0.0004625 | 22.6025 |
| | Standardbred | Batagai | 0.0060153 | 0.0004598 | 13.0811 |
| | Std | Mng | 0.006083 | 0.0004411 | 13.7918 |
| | | Prz_Przewalski | 0.0058852 | 0.000448 | 13.1375 |
| | | Batagai | 0.0058422 | 0.0004147 | 14.0889 |
| | | Standardbred | 0.0180307 | 0.0007438 | 24.2404 |
| **Yak5** | Arabian | CGG101397 | 0.0024551 | 0.0005198 | 4.72268 |
| | | Fjord | 0.0055765 | 0.0004721 | 11.8121 |
| | | Batagai | 0.0055204 | 0.0004494 | 12.283 |
| | | CGG10022 | 0.0055063 | 0.0004241 | 12.9851 |
| | | CGG10023 | 0.0053739 | 0.0004431 | 12.1276 |
| | | Mng | 0.0054808 | 0.0004315 | 12.7028 |
| | | Prz_D | 0.0057132 | 0.0004235 | 13.4906 |
| | | Prz_Przewalski | 0.0056438 | 0.0004314 | 13.0822 |
| | | Mon | 0.0078355 | 0.000444 | 17.6463 |
| | | Mor | 0.0084877 | 0.0004677 | 18.1488 |
| | | Std | 0.0095828 | 0.0005045 | 18.9947 |
| | | Standardbred | 0.0099896 | 0.0005167 | 19.3335 |
| | | Qrt | 0.0101726 | 0.00047 | 21.6459 |
| | CGG10022 | CGG101397 | 0.0026849 | 0.0005044 | 5.32305 |
| | | Mng | 0.0052639 | 0.0004343 | 12.1205 |
| | | Standardbred | 0.005723 | 0.0004638 | 12.3401 |
| | | Mon | 0.0059847 | 0.0004281 | 13.9797 |
| | | Mor | 0.0055625 | 0.0004133 | 13.4584 |
| | | Qrt | 0.0057209 | 0.0004246 | 13.4728 |
| | | Std | 0.0058421 | 0.0004366 | 13.3823 |
| | | Prz_Przewalski | 0.0093063 | 0.0004485 | 20.7519 |
| | | Batagai | 0.0222892 | 0.0005259 | 42.3867 |
| | | CGG10023 | 0.0222549 | 0.0005274 | 42.1966 |
| | CGG10023 | CGG101397 | 0.0026107 | 0.0005148 | 5.07143 |
| | | Mng | 0.0053123 | 0.0004401 | 12.0709 |
| | | Mor | 0.0056557 | 0.0004379 | 12.9145 |
| | | Standardbred | 0.0058576 | 0.0004754 | 12.3207 |
| | | Std | 0.0058411 | 0.0004498 | 12.985 |
| | | Mon | 0.00585 | 0.0004394 | 13.315 |
| | | Qrt | 0.0057482 | 0.0004356 | 13.1958 |
| | | Prz_Przewalski | 0.0094556 | 0.0004638 | 20.3882 |
| | | Batagai | 0.0226524 | 0.0005449 | 41.5749 |
| | CGG101397 | Mng | 0.0024835 | 0.0005193 | 4.78225 |
| | | Qrt | 0.0024078 | 0.000503 | 4.78709 |
| | | Batagai | 0.0026338 | 0.0005112 | 5.15202 |
| | | Prz_Przewalski | 0.0027081 | 0.0005126 | 5.28344 |
| | | Standardbred | 0.0027131 | 0.000528 | 5.13806 |
| | | Std | 0.0026622 | 0.0005175 | 5.14407 |
| | Fjord | CGG101397 | 0.0026121 | 0.0005481 | 4.76563 |
| | | Mng | 0.0053681 | 0.0004684 | 11.4606 |
| | | Batagai | 0.0057588 | 0.0004624 | 12.4544 |
| | | CGG10022 | 0.0059031 | 0.0004544 | 12.9917 |
| | | CGG10023 | 0.0056915 | 0.0004594 | 12.3878 |
| | | Prz_D | 0.005886 | 0.0004657 | 12.6388 |
| | | Qrt | 0.005885 | 0.0004591 | 12.8191 |
| | | Standardbred | 0.006316 | 0.0004925 | 12.8243 |
| | | Mon | 0.0061811 | 0.0004566 | 13.536 |
| | | Mor | 0.0060755 | 0.0004536 | 13.3951 |
| | | Prz_Przewalski | 0.0061148 | 0.0004568 | 13.3865 |
| | | Std | 0.0063332 | 0.0004814 | 13.1563 |
| | Mon | CGG101397 | 0.0026047 | 0.0005 | 5.20948 |
| | | Batagai | 0.005812 | 0.0004412 | 13.1731 |
| | | Mng | 0.0057312 | 0.0004399 | 13.0278 |
| | | Prz_Przewalski | 0.0057707 | 0.0004224 | 13.6607 |
| | | Mor | 0.0071462 | 0.0004234 | 16.8785 |
| | | Qrt | 0.0079466 | 0.000451 | 17.6192 |
| | | Standardbred | 0.0079259 | 0.0004603 | 17.2183 |
| | | Std | 0.0079687 | 0.0004588 | 17.3705 |
| | Icelandic | CGG101397 | 0.0026479 | 0.0005114 | 5.17811 |
| | | Mng | 0.0057178 | 0.0004566 | 12.5237 |
| | | Arabian | 0.0063727 | 0.0004688 | 13.5941 |
| | | Batagai | 0.0061267 | 0.0004622 | 13.2569 |
| | | CGG10022 | 0.0059641 | 0.0004529 | 13.1689 |
| | | CGG10023 | 0.0062914 | 0.000458 | 13.7369 |
| | | Mor | 0.0062844 | 0.0004618 | 13.6093 |
| | | Prz_D | 0.0058888 | 0.0004428 | 13.3001 |
| | | Standardbred | 0.0067387 | 0.0005031 | 13.3936 |
| | | Std | 0.0064706 | 0.0004694 | 13.7851 |
| | | Fjord | 0.0071282 | 0.0004874 | 14.6245 |
| | | Mon | 0.0063869 | 0.0004452 | 14.3448 |
| | | Prz_Przewalski | 0.0061611 | 0.0004358 | 14.1367 |
| | | Qrt | 0.0064155 | 0.0004531 | 14.1598 |
| | Mng | Batagai | 0.0052591 | 0.0004424 | 11.8883 |
| | | Standardbred | 0.0056267 | 0.000458 | 12.2864 |
| | Mor | CGG101397 | 0.0023878 | 0.0004948 | 4.82575 |
| | | Mng | 0.0054178 | 0.0004262 | 12.7126 |
| | | Batagai | 0.005559 | 0.0004259 | 13.0509 |
| | | Prz_Przewalski | 0.0057979 | 0.0004232 | 13.6994 |
| | | Standardbred | 0.0092236 | 0.0005533 | 16.669 |
| | | Std | 0.0085494 | 0.0004796 | 17.8245 |
| | | Qrt | 0.0088028 | 0.0004504 | 19.5437 |
| | Prz_D | CGG101397 | 0.0025999 | 0.0005196 | 5.00339 |
| | | Mng | 0.0056636 | 0.0004382 | 12.9261 |
| | | Standardbred | 0.0058727 | 0.0004614 | 12.7288 |
| | | Mon | 0.0057848 | 0.0004353 | 13.2894 |
| | | Mor | 0.0057391 | 0.0004224 | 13.5866 |
| | | Qrt | 0.0056137 | 0.0004209 | 13.3388 |
| | | Std | 0.0058354 | 0.0004372 | 13.3471 |
| | | Batagai | 0.009695 | 0.0004809 | 20.1616 |
| | | CGG10022 | 0.0095593 | 0.000462 | 20.6902 |
| | | CGG10023 | 0.0095313 | 0.0004745 | 20.0882 |
| | | Prz_Przewalski | 0.0351997 | 0.0010891 | 32.3203 |
| | Prz_Przewalski | Mng | 0.0054948 | 0.0004299 | 12.7813 |
| | | Standardbred | 0.0058212 | 0.0004477 | 13.003 |
| | | Batagai | 0.0094515 | 0.0004639 | 20.3739 |
| | Qrt | Batagai | 0.0055152 | 0.0004373 | 12.613 |
| | | Mng | 0.0057217 | 0.0004307 | 13.2853 |
| | | Prz_Przewalski | 0.0057174 | 0.000424 | 13.4846 |
| | | Standardbred | 0.0100419 | 0.0005222 | 19.2292 |
| | | Std | 0.0100417 | 0.0005032 | 19.9537 |
| | Standardbred | Batagai | 0.0057721 | 0.0004536 | 12.7254 |
| | Std | Batagai | 0.005716 | 0.0004481 | 12.7554 |
| | | Mng | 0.0057334 | 0.0004464 | 12.8423 |
| | | Prz_Przewalski | 0.0056617 | 0.0004325 | 13.0909 |
| | | Standardbred | 0.0178311 | 0.0007433 | 23.988 |
| **Yak6** | Arabian | CGG101397 | 0.000623078 | 0.000456138 | 1.36598 |
| | | CGG10023 | 0.00353278 | 0.000395996 | 8.92124 |
| | | Batagai | 0.00358469 | 0.000393705 | 9.10502 |
| | | CGG10022 | 0.0036989 | 0.000376064 | 9.83583 |
| | | Fjord | 0.00387277 | 0.000429232 | 9.02256 |
| | | Mng | 0.00388399 | 0.000398549 | 9.74531 |
| | | Prz_D | 0.00394455 | 0.000381184 | 10.3482 |
| | | Prz_Przewalski | 0.00379424 | 0.000375218 | 10.1121 |
| | | Mor | 0.006819 | 0.000456406 | 14.9407 |
| | | Mon | 0.00623517 | 0.000414603 | 15.0389 |
| | | Standardbred | 0.00812328 | 0.000486615 | 16.6935 |
| | | Std | 0.00785905 | 0.000486183 | 16.1648 |
| | | Qrt | 0.0083441 | 0.000428277 | 19.4829 |
| | CGG10022 | CGG101397 | 0.000981185 | 0.000423524 | 2.31671 |
| | | Mng | 0.00379533 | 0.000351823 | 10.7876 |
| | | Mor | 0.00402204 | 0.0003852 | 10.4414 |
| | | Qrt | 0.00402064 | 0.000366615 | 10.9669 |
| | | Standardbred | 0.00398487 | 0.000376477 | 10.5846 |
| | | Std | 0.00424664 | 0.000384391 | 11.0477 |
| | | Mon | 0.00451259 | 0.000348864 | 12.9351 |
| | | Prz_Przewalski | 0.007585 | 0.000388392 | 19.5293 |
| | | CGG10023 | 0.020542 | 0.000476679 | 43.094 |
| | | Batagai | 0.0204817 | 0.000464696 | 44.0755 |
| | CGG10023 | CGG101397 | 0.00087324 | 0.000420112 | 2.07859 |
| | | Mng | 0.00381001 | 0.00036875 | 10.3322 |
| | | Mor | 0.0040816 | 0.000395073 | 10.3312 |
| | | Qrt | 0.00401432 | 0.000375135 | 10.701 |
| | | Standardbred | 0.00408582 | 0.000390626 | 10.4597 |
| | | Std | 0.00421196 | 0.00039771 | 10.5905 |
| | | Mon | 0.00434421 | 0.000357121 | 12.1645 |
| | | Prz_Przewalski | 0.00770066 | 0.00040908 | 18.8243 |
| | | Batagai | 0.0208112 | 0.000504493 | 41.2518 |
| | CGG101397 | Batagai | 0.000801795 | 0.000429703 | 1.86593 |
| | | Qrt | 0.000683041 | 0.000421159 | 1.62181 |
| | | Mng | 0.000990349 | 0.000430444 | 2.30076 |
| | | Prz_Przewalski | 0.000962239 | 0.000426272 | 2.25733 |
| | | Standardbred | 0.000950438 | 0.000444518 | 2.13813 |
| | | Std | 0.00104223 | 0.000442491 | 2.35536 |
| | Fjord | CGG101397 | 0.00101207 | 0.000460508 | 2.19772 |
| | | Batagai | 0.00405504 | 0.000404292 | 10.03 |
| | | CGG10023 | 0.00408235 | 0.000405897 | 10.0576 |
| | | Mng | 0.00400319 | 0.000393925 | 10.1623 |
| | | Mor | 0.00463877 | 0.000451481 | 10.2746 |
| | | Qrt | 0.00428842 | 0.000404044 | 10.6137 |
| | | CGG10022 | 0.00432764 | 0.00038749 | 11.1684 |
| | | Prz_D | 0.0043493 | 0.000391981 | 11.0957 |
| | | Prz_Przewalski | 0.0044972 | 0.000388594 | 11.573 |
| | | Standardbred | 0.00468159 | 0.000423065 | 11.0659 |
| | | Std | 0.00484138 | 0.000428198 | 11.3064 |
| | | Mon | 0.00481265 | 0.000395034 | 12.1829 |
| | Mon | CGG101397 | 0.00110806 | 0.000418601 | 2.64705 |
| | | Batagai | 0.00421161 | 0.000357965 | 11.7654 |
| | | Mng | 0.00446965 | 0.000359577 | 12.4303 |
| | | Prz_Przewalski | 0.00425645 | 0.00035432 | 12.013 |
| | | Mor | 0.00581283 | 0.000414603 | 14.0202 |
| | | Standardbred | 0.00639489 | 0.000405563 | 15.7679 |
| | | Qrt | 0.00645342 | 0.000398169 | 16.2077 |
| | | Std | 0.00658034 | 0.000407251 | 16.1579 |
| | Icelandic | CGG101397 | 0.000970758 | 0.00045754 | 2.12169 |
| | | CGG10022 | 0.0043115 | 0.000394417 | 10.9313 |
| | | Mng | 0.00427579 | 0.000391101 | 10.9327 |
| | | Mor | 0.00477054 | 0.000443409 | 10.7588 |
| | | Arabian | 0.00459189 | 0.000411585 | 11.1566 |
| | | Batagai | 0.00434585 | 0.000393557 | 11.0425 |
| | | CGG10023 | 0.00460514 | 0.00040452 | 11.3842 |
| | | Prz_D | 0.00427505 | 0.000379343 | 11.2696 |
| | | Prz_Przewalski | 0.00446639 | 0.000392193 | 11.3882 |
| | | Qrt | 0.00474191 | 0.000406525 | 11.6645 |
| | | Std | 0.00490173 | 0.000420784 | 11.649 |
| | | Mon | 0.00494143 | 0.000401337 | 12.3124 |
| | | Standardbred | 0.00502719 | 0.00041866 | 12.0078 |
| | | Fjord | 0.00557937 | 0.000427373 | 13.055 |
| | Mng | Batagai | 0.00366218 | 0.000354133 | 10.3413 |
| | | Standardbred | 0.00409917 | 0.000375864 | 10.906 |
| | Mor | CGG101397 | 0.000822784 | 0.000444277 | 1.85196 |
| | | Batagai | 0.00389028 | 0.000392215 | 9.91875 |
| | | Mng | 0.00408794 | 0.000393773 | 10.3815 |
| | | Prz_Przewalski | 0.00421537 | 0.000379529 | 11.1068 |
| | | Standardbred | 0.00762422 | 0.000496684 | 15.3502 |
| | | Std | 0.00709267 | 0.000467313 | 15.1776 |
| | | Qrt | 0.00724129 | 0.000437625 | 16.5468 |
| | Prz_D | CGG101397 | 0.000934956 | 0.000425613 | 2.19673 |
| | | Standardbred | 0.00417343 | 0.000383188 | 10.8914 |
| | | Std | 0.00427873 | 0.000390987 | 10.9434 |
| | | Mng | 0.00423387 | 0.000376443 | 11.247 |
| | | Mor | 0.00423751 | 0.000383011 | 11.0637 |
| | | Qrt | 0.00395232 | 0.000359091 | 11.0065 |
| | | Mon | 0.00435153 | 0.000362486 | 12.0047 |
| | | Batagai | 0.00792633 | 0.000408514 | 19.4028 |
| | | CGG10023 | 0.00785727 | 0.00040084 | 19.602 |
| | | CGG10022 | 0.00791894 | 0.000388762 | 20.3697 |
| | | Prz_Przewalski | 0.0335172 | 0.00108736 | 30.8243 |
| | Prz_Przewalski | Mng | 0.00398413 | 0.000362536 | 10.9896 |
| | | Standardbred | 0.00404096 | 0.000383975 | 10.524 |
| | | Batagai | 0.00760188 | 0.000403482 | 18.8407 |
| | Qrt | Batagai | 0.00368671 | 0.000371997 | 9.91058 |
| | | Mng | 0.00423202 | 0.000355448 | 11.9061 |
| | | Prz_Przewalski | 0.00397505 | 0.000356795 | 11.141 |
| | | Standardbred | 0.00828275 | 0.000478972 | 17.2928 |
| | | Std | 0.00842519 | 0.000456677 | 18.4489 |
| | Standardbred | Batagai | 0.00390576 | 0.00037908 | 10.3033 |
| | Std | Batagai | 0.0039923 | 0.000390576 | 10.2216 |
| | | Prz_Przewalski | 0.00402415 | 0.000387672 | 10.3803 |
| | | Mng | 0.00434849 | 0.000387678 | 11.2168 |
| | | Standardbred | 0.0161767 | 0.000695673 | 23.2533 |
| **Yak7** | Arabian | CGG101397 | 0.0012329 | 0.0004993 | 2.46896 |
| | | Fjord | 0.0046332 | 0.0004723 | 9.81088 |
| | | Batagai | 0.0047538 | 0.0004458 | 10.6627 |
| | | CGG10022 | 0.0044881 | 0.0004249 | 10.5632 |
| | | CGG10023 | 0.0044888 | 0.0004277 | 10.4963 |
| | | Mng | 0.0044549 | 0.0004182 | 10.6517 |
| | | Prz_D | 0.0049352 | 0.0004172 | 11.8285 |
| | | Prz_Przewalski | 0.0047958 | 0.0004308 | 11.1311 |
| | | Mor | 0.0077764 | 0.0004932 | 15.7682 |
| | | Mon | 0.0069606 | 0.0004346 | 16.0179 |
| | | Standardbred | 0.0090512 | 0.0005402 | 16.754 |
| | | Std | 0.0087481 | 0.0005209 | 16.7951 |
| | | Qrt | 0.0093887 | 0.0004744 | 19.7889 |
| | CGG10022 | CGG101397 | 0.0011915 | 0.0004861 | 2.45111 |
| | | Mng | 0.0039668 | 0.0003901 | 10.1674 |
| | | Mor | 0.0045801 | 0.0004214 | 10.8685 |
| | | Qrt | 0.0046659 | 0.0004273 | 10.9184 |
| | | Standardbred | 0.0045134 | 0.0004116 | 10.965 |
| | | Std | 0.0047363 | 0.0004119 | 11.4989 |
| | | Mon | 0.0048386 | 0.0004005 | 12.0827 |
| | | Prz_Przewalski | 0.0081872 | 0.0004447 | 18.4088 |
| | | CGG10023 | 0.0210986 | 0.0005139 | 41.0537 |
| | | Batagai | 0.0212514 | 0.0004906 | 43.3175 |
| | CGG10023 | CGG101397 | 0.0012504 | 0.0004879 | 2.56282 |
| | | Mng | 0.0041483 | 0.0004034 | 10.2843 |
| | | Mor | 0.0048064 | 0.0004478 | 10.7324 |
| | | Standardbred | 0.0047812 | 0.0004433 | 10.786 |
| | | Mon | 0.004837 | 0.0004163 | 11.6182 |
| | | Qrt | 0.0048264 | 0.0004335 | 11.1341 |
| | | Std | 0.0048684 | 0.0004339 | 11.2196 |
| | | Prz_Przewalski | 0.0084696 | 0.0004839 | 17.5026 |
| | | Batagai | 0.0217477 | 0.000544 | 39.9777 |
| | CGG101397 | Batagai | 0.001392 | 0.0004976 | 2.79755 |
| | | Mng | 0.0009823 | 0.0004845 | 2.02761 |
| | | Prz_Przewalski | 0.001385 | 0.0005099 | 2.71591 |
| | | Qrt | 0.0011488 | 0.0004948 | 2.32181 |
| | | Standardbred | 0.0012995 | 0.0005028 | 2.58481 |
| | | Std | 0.0013524 | 0.0005022 | 2.69301 |
| | Fjord | CGG101397 | 0.0011936 | 0.0005366 | 2.22437 |
| | | CGG10023 | 0.0046102 | 0.0004614 | 9.99251 |
| | | Mng | 0.0041458 | 0.0004332 | 9.57051 |
| | | Batagai | 0.0047959 | 0.0004662 | 10.2877 |
| | | CGG10022 | 0.0046886 | 0.0004494 | 10.4327 |
| | | Mor | 0.005168 | 0.0004826 | 10.7088 |
| | | Prz_D | 0.0049118 | 0.000466 | 10.5412 |
| | | Prz_Przewalski | 0.0050706 | 0.0004747 | 10.6825 |
| | | Qrt | 0.0049048 | 0.0004757 | 10.3116 |
| | | Mon | 0.0051098 | 0.0004482 | 11.4016 |
| | | Standardbred | 0.0051813 | 0.0004475 | 11.5772 |
| | | Std | 0.0053022 | 0.0004602 | 11.5204 |
| | Mon | CGG101397 | 0.0012546 | 0.0004889 | 2.56641 |
| | | Batagai | 0.0049175 | 0.00042 | 11.7089 |
| | | Mng | 0.0045773 | 0.0004007 | 11.4221 |
| | | Prz_Przewalski | 0.0047948 | 0.0004286 | 11.1872 |
| | | Mor | 0.006307 | 0.0004557 | 13.8405 |
| | | Standardbred | 0.0068596 | 0.0004338 | 15.8125 |
| | | Qrt | 0.0070348 | 0.000437 | 16.097 |
| | | Std | 0.0070062 | 0.0004361 | 16.0657 |
| | Icelandic | CGG101397 | 0.0011986 | 0.0005287 | 2.26697 |
| | | Mng | 0.0044647 | 0.0004366 | 10.2258 |
| | | Prz_Przewalski | 0.005086 | 0.00047 | 10.8223 |
| | | Arabian | 0.0053986 | 0.0004542 | 11.8863 |
| | | Batagai | 0.005133 | 0.0004563 | 11.2494 |
| | | CGG10022 | 0.0047188 | 0.0004214 | 11.1988 |
| | | CGG10023 | 0.0051792 | 0.0004451 | 11.6351 |
| | | Mon | 0.0052849 | 0.0004517 | 11.6988 |
| | | Mor | 0.005346 | 0.0004654 | 11.4862 |
| | | Prz_D | 0.0048838 | 0.0004431 | 11.0224 |
| | | Std | 0.0054088 | 0.000464 | 11.6559 |
| | | Fjord | 0.0059578 | 0.0004758 | 12.5229 |
| | | Qrt | 0.0054046 | 0.0004499 | 12.0122 |
| | | Standardbred | 0.0055732 | 0.0004568 | 12.2009 |
| | Mng | Batagai | 0.0042135 | 0.0004012 | 10.5026 |
| | | Standardbred | 0.0044094 | 0.0003966 | 11.1181 |
| | Mor | CGG101397 | 0.0012014 | 0.0005117 | 2.34798 |
| | | Batagai | 0.0048282 | 0.0004485 | 10.766 |
| | | Mng | 0.0044276 | 0.0004254 | 10.4085 |
| | | Prz_Przewalski | 0.0049858 | 0.0004546 | 10.9672 |
| | | Standardbred | 0.008321 | 0.0005651 | 14.7244 |
| | | Std | 0.0077505 | 0.0004997 | 15.5113 |
| | | Qrt | 0.0080547 | 0.0004705 | 17.121 |
| | Prz_D | CGG101397 | 0.0013468 | 0.0004945 | 2.72363 |
| | | Mon | 0.004879 | 0.0004127 | 11.8222 |
| | | Mng | 0.0046068 | 0.0004101 | 11.232 |
| | | Mor | 0.004997 | 0.000449 | 11.1281 |
| | | Qrt | 0.004799 | 0.0004323 | 11.1015 |
| | | Standardbred | 0.0049034 | 0.0004236 | 11.5745 |
| | | Std | 0.0049698 | 0.0004123 | 12.0551 |
| | | CGG10023 | 0.0086153 | 0.0004719 | 18.2572 |
| | | Batagai | 0.0088975 | 0.0004604 | 19.3276 |
| | | CGG10022 | 0.0085102 | 0.000435 | 19.563 |
| | | Prz_Przewalski | 0.0343209 | 0.001148 | 29.8963 |
| | Prz_Przewalski | Mng | 0.0043679 | 0.0004233 | 10.32 |
| | | Standardbred | 0.0047819 | 0.0004323 | 11.0605 |
| | | Batagai | 0.0085839 | 0.0004705 | 18.2447 |
| | Qrt | Batagai | 0.0047118 | 0.0004431 | 10.6325 |
| | | Prz_Przewalski | 0.0048326 | 0.0004536 | 10.653 |
| | | Mng | 0.0046589 | 0.0004115 | 11.3213 |
| | | Standardbred | 0.0090667 | 0.0004878 | 18.5858 |
| | | Std | 0.0091703 | 0.0004798 | 19.1122 |
| | Standardbred | Batagai | 0.0048142 | 0.0004376 | 11.0003 |
| | Std | Batagai | 0.0048618 | 0.0004333 | 11.2217 |
| | | Mng | 0.0046198 | 0.0003999 | 11.5532 |
| | | Prz_Przewalski | 0.0047261 | 0.0004279 | 11.0456 |
| | | Standardbred | 0.0168051 | 0.0007128 | 23.5778 |
| **Yak8** | Arabian | CGG101397 | 0.0003818 | 0.0004268 | 0.894713 |
| | | Batagai | 0.0030086 | 0.0003409 | 8.8247 |
| | | CGG10023 | 0.0030471 | 0.0003098 | 9.83444 |
| | | Fjord | 0.0032083 | 0.0003398 | 9.44061 |
| | | Mng | 0.0031228 | 0.0003211 | 9.72657 |
| | | Prz_D | 0.0032024 | 0.0003233 | 9.90448 |
| | | Prz_Przewalski | 0.0031473 | 0.000328 | 9.59521 |
| | | CGG10022 | 0.0031922 | 0.0003046 | 10.4798 |
| | | Mon | 0.0055418 | 0.0003558 | 15.5749 |
| | | Mor | 0.0060782 | 0.000399 | 15.2347 |
| | | Standardbred | 0.007456 | 0.0004421 | 16.8633 |
| | | Std | 0.0072027 | 0.0004239 | 16.9934 |
| | | Qrt | 0.0079249 | 0.0003999 | 19.8191 |
| | CGG10022 | CGG101397 | 0.0005402 | 0.0003845 | 1.40495 |
| | | Mng | 0.0028345 | 0.0002978 | 9.51785 |
| | | Mor | 0.0030816 | 0.0003053 | 10.0924 |
| | | Standardbred | 0.0031179 | 0.0003026 | 10.3021 |
| | | Qrt | 0.0034018 | 0.0002854 | 11.9183 |
| | | Std | 0.0033906 | 0.0002903 | 11.6798 |
| | | Mon | 0.0036195 | 0.0002929 | 12.359 |
| | | Prz_Przewalski | 0.0067384 | 0.00032 | 21.059 |
| | | CGG10023 | 0.0198566 | 0.0004195 | 47.3295 |
| | | Batagai | 0.0197059 | 0.0003983 | 49.4768 |
| | CGG10023 | CGG101397 | 0.0004533 | 0.0003844 | 1.17943 |
| | | Mng | 0.0028702 | 0.0003021 | 9.50066 |
| | | Mor | 0.0031622 | 0.0003222 | 9.81408 |
| | | Standardbred | 0.0032399 | 0.0003171 | 10.2186 |
| | | Qrt | 0.0034165 | 0.000291 | 11.7397 |
| | | Std | 0.0033769 | 0.0003 | 11.2549 |
| | | Mon | 0.0034722 | 0.0002871 | 12.0958 |
| | | Prz_Przewalski | 0.0068751 | 0.0003144 | 21.8669 |
| | | Batagai | 0.0200564 | 0.0004067 | 49.3099 |
| | CGG101397 | Batagai | 0.0002915 | 0.0003961 | 0.735772 |
| | | Mng | 0.000295 | 0.0004161 | 0.708909 |
| | | Prz_Przewalski | 0.0003811 | 0.0004292 | 0.887919 |
| | | Qrt | 0.0003296 | 0.0004097 | 0.804542 |
| | | Standardbred | 0.000349 | 0.0004372 | 0.79813 |
| | | Std | 0.0004516 | 0.0004149 | 1.08859 |
| | Fjord | CGG101397 | 0.0004134 | 0.0004187 | 0.987404 |
| | | Mng | 0.0028846 | 0.0003263 | 8.8404 |
| | | Prz_D | 0.0032497 | 0.0003673 | 8.84711 |
| | | Batagai | 0.0031215 | 0.0003134 | 9.96025 |
| | | CGG10023 | 0.0032392 | 0.0003266 | 9.91822 |
| | | Prz_Przewalski | 0.0034929 | 0.0003695 | 9.453 |
| | | Mor | 0.0035406 | 0.0003415 | 10.3681 |
| | | Standardbred | 0.0036569 | 0.0003496 | 10.4612 |
| | | CGG10022 | 0.0034635 | 0.0003133 | 11.0552 |
| | | Mon | 0.0037619 | 0.0003207 | 11.7297 |
| | | Qrt | 0.0035118 | 0.0003117 | 11.2679 |
| | | Std | 0.0038276 | 0.0003298 | 11.605 |
| | Mon | CGG101397 | 0.0004805 | 0.0004042 | 1.18865 |
| | | Batagai | 0.0032492 | 0.0003019 | 10.7636 |
| | | Mng | 0.0033222 | 0.0003097 | 10.7282 |
| | | Prz_Przewalski | 0.0032232 | 0.0003119 | 10.3348 |
| | | Mor | 0.0046857 | 0.0003471 | 13.4978 |
| | | Standardbred | 0.0053413 | 0.0003427 | 15.5879 |
| | | Qrt | 0.0056479 | 0.0003376 | 16.7303 |
| | | Std | 0.0055376 | 0.0003248 | 17.0505 |
| | Icelandic | CGG101397 | 0.0005535 | 0.0004387 | 1.26178 |
| | | Mng | 0.0033387 | 0.000358 | 9.32693 |
| | | Prz_D | 0.0033569 | 0.000355 | 9.45638 |
| | | Batagai | 0.0035938 | 0.0003472 | 10.3505 |
| | | CGG10022 | 0.0036288 | 0.0003316 | 10.9422 |
| | | Mor | 0.0038538 | 0.0003686 | 10.4555 |
| | | Prz_Przewalski | 0.0036435 | 0.000354 | 10.292 |
| | | Arabian | 0.0041089 | 0.0003604 | 11.4011 |
| | | CGG10023 | 0.0039435 | 0.0003455 | 11.4145 |
| | | Mon | 0.0040721 | 0.000345 | 11.8029 |
| | | Qrt | 0.0041468 | 0.000347 | 11.9509 |
| | | Standardbred | 0.004184 | 0.0003803 | 11.0015 |
| | | Std | 0.0040694 | 0.0003469 | 11.7305 |
| | | Fjord | 0.004739 | 0.0003711 | 12.7702 |
| | Mng | Batagai | 0.0026319 | 0.0003154 | 8.34381 |
| | | Standardbred | 0.0029778 | 0.0003349 | 8.89225 |
| | Mor | CGG101397 | 0.0001478 | 0.0004456 | 0.331617 |
| | | Batagai | 0.0028804 | 0.0003314 | 8.69174 |
| | | Mng | 0.002893 | 0.0003358 | 8.61652 |
| | | Prz_Przewalski | 0.0031347 | 0.0003321 | 9.43991 |
| | | Standardbred | 0.0065232 | 0.0004683 | 13.931 |
| | | Std | 0.0060025 | 0.0003869 | 15.5131 |
| | | Qrt | 0.0063884 | 0.0003992 | 16.0009 |
| | Prz_D | CGG101397 | 0.0002585 | 0.0004264 | 0.606325 |
| | | Standardbred | 0.003071 | 0.0003427 | 8.96208 |
| | | Mng | 0.0030375 | 0.0003375 | 9.00112 |
| | | Mor | 0.0030616 | 0.0003344 | 9.15484 |
| | | Qrt | 0.003098 | 0.0003223 | 9.61131 |
| | | Std | 0.0031872 | 0.0003297 | 9.66804 |
| | | Mon | 0.003223 | 0.0003131 | 10.2941 |
| | | Batagai | 0.0069151 | 0.0003436 | 20.123 |
| | | CGG10022 | 0.0069771 | 0.000334 | 20.8911 |
| | | CGG10023 | 0.0069364 | 0.0003276 | 21.1751 |
| | | Prz_Przewalski | 0.0324352 | 0.0010154 | 31.9422 |
| | Prz_Przewalski | Mng | 0.0028831 | 0.0003231 | 8.92269 |
| | | Standardbred | 0.0030338 | 0.0003525 | 8.60567 |
| | | Batagai | 0.0066859 | 0.0003285 | 20.3507 |
| | Qrt | Prz_Przewalski | 0.003216 | 0.0003274 | 9.82179 |
| | | Batagai | 0.0029985 | 0.0002994 | 10.0137 |
| | | Mng | 0.0033587 | 0.0003147 | 10.6734 |
| | | Standardbred | 0.0075034 | 0.0004093 | 18.3343 |
| | | Std | 0.0076567 | 0.0003849 | 19.8905 |
| | Standardbred | Batagai | 0.0029694 | 0.0003153 | 9.419 |
| | Std | Prz_Przewalski | 0.0030279 | 0.0003322 | 9.11381 |
| | | Batagai | 0.0030668 | 0.0003017 | 10.1645 |
| | | Mng | 0.003238 | 0.0003186 | 10.1616 |
| | | Standardbred | 0.0151601 | 0.0006478 | 23.4037 |
| **Yak9** | Arabian | CGG101397 | 0.0034389 | 0.0005062 | 6.79419 |
| | | Fjord | 0.0060579 | 0.000485 | 12.4917 |
| | | Batagai | 0.00617 | 0.0004467 | 13.8117 |
| | | CGG10022 | 0.0060277 | 0.0004316 | 13.9674 |
| | | CGG10023 | 0.0061538 | 0.0004418 | 13.9302 |
| | | Mng | 0.0059903 | 0.0004423 | 13.5424 |
| | | Prz_D | 0.0063365 | 0.0004684 | 13.5276 |
| | | Prz_Przewalski | 0.0062702 | 0.0004732 | 13.2515 |
| | | Mon | 0.0084326 | 0.0004911 | 17.1713 |
| | | Mor | 0.0093027 | 0.0005028 | 18.502 |
| | | Standardbred | 0.0104212 | 0.0005302 | 19.6567 |
| | | Std | 0.010096 | 0.0005034 | 20.0562 |
| | | Qrt | 0.0107384 | 0.0004989 | 21.5222 |
| | CGG10022 | CGG101397 | 0.0037655 | 0.0004598 | 8.19001 |
| | | Mng | 0.0058701 | 0.0004045 | 14.5118 |
| | | Mor | 0.0064742 | 0.0004448 | 14.554 |
| | | Standardbred | 0.0062513 | 0.0004304 | 14.5238 |
| | | Qrt | 0.0063834 | 0.0004075 | 15.6642 |
| | | Std | 0.0064521 | 0.0004117 | 15.6725 |
| | | Mon | 0.0066785 | 0.0004055 | 16.4683 |
| | | Prz_Przewalski | 0.0100295 | 0.0004538 | 22.1013 |
| | | Batagai | 0.0230355 | 0.0005069 | 45.4457 |
| | | CGG10023 | 0.0231315 | 0.0005059 | 45.7235 |
| | CGG10023 | CGG101397 | 0.0039498 | 0.000463 | 8.53182 |
| | | Mng | 0.006177 | 0.0004141 | 14.915 |
| | | Mor | 0.006826 | 0.0004417 | 15.453 |
| | | Standardbred | 0.0066445 | 0.0004416 | 15.0467 |
| | | Mon | 0.0068023 | 0.0004057 | 16.7656 |
| | | Qrt | 0.0066693 | 0.0004086 | 16.3213 |
| | | Std | 0.0067097 | 0.0004152 | 16.1587 |
| | | Prz_Przewalski | 0.0104373 | 0.0004817 | 21.6688 |
| | | Batagai | 0.0236572 | 0.000545 | 43.4081 |
| | CGG101397 | Mng | 0.0035522 | 0.0005012 | 7.08774 |
| | | Prz_Przewalski | 0.0038938 | 0.0004902 | 7.94366 |
| | | Qrt | 0.0035329 | 0.0004879 | 7.24154 |
| | | Standardbred | 0.0037039 | 0.0005054 | 7.32841 |
| | | Std | 0.0037347 | 0.000474 | 7.87872 |
| | | Batagai | 0.0038426 | 0.000475 | 8.09004 |
| | Fjord | CGG101397 | 0.0036528 | 0.0004905 | 7.44625 |
| | | Mng | 0.0059343 | 0.000438 | 13.5482 |
| | | Batagai | 0.0064652 | 0.0004398 | 14.7009 |
| | | Prz_D | 0.006566 | 0.0004481 | 14.6515 |
| | | Qrt | 0.0065075 | 0.0004468 | 14.5643 |
| | | Standardbred | 0.0068044 | 0.0004685 | 14.5252 |
| | | CGG10022 | 0.0064813 | 0.0004224 | 15.3428 |
| | | CGG10023 | 0.0065282 | 0.0004283 | 15.243 |
| | | Mon | 0.0068349 | 0.0004519 | 15.1261 |
| | | Mor | 0.0069473 | 0.0004616 | 15.0514 |
| | | Prz_Przewalski | 0.006798 | 0.0004338 | 15.6712 |
| | | Std | 0.0069032 | 0.0004561 | 15.1338 |
| | Mon | CGG101397 | 0.003761 | 0.0004811 | 7.81745 |
| | | Batagai | 0.006634 | 0.0004321 | 15.3516 |
| | | Mng | 0.006413 | 0.0004267 | 15.0304 |
| | | Prz_Przewalski | 0.0065695 | 0.0004332 | 15.1644 |
| | | Mor | 0.0081336 | 0.0004751 | 17.1199 |
| | | Standardbred | 0.0085299 | 0.0004791 | 17.805 |
| | | Qrt | 0.0086848 | 0.0004578 | 18.9706 |
| | | Std | 0.0086544 | 0.0004579 | 18.9005 |
| | Icelandic | CGG101397 | 0.0039464 | 0.000499 | 7.90809 |
| | | Arabian | 0.007112 | 0.0005018 | 14.1729 |
| | | Mng | 0.0065419 | 0.0004478 | 14.6088 |
| | | Prz_D | 0.0068267 | 0.0004583 | 14.8957 |
| | | Batagai | 0.0070909 | 0.0004568 | 15.5228 |
| | | CGG10022 | 0.0068001 | 0.0004446 | 15.2962 |
| | | Mon | 0.0072986 | 0.0004644 | 15.7164 |
| | | Mor | 0.007414 | 0.0004909 | 15.1025 |
| | | Prz_Przewalski | 0.0071022 | 0.0004632 | 15.3324 |
| | | Standardbred | 0.0074849 | 0.0004821 | 15.5257 |
| | | Std | 0.0072985 | 0.0004646 | 15.7109 |
| | | CGG10023 | 0.007386 | 0.0004547 | 16.2447 |
| | | Fjord | 0.0079243 | 0.0004662 | 16.9965 |
| | | Qrt | 0.007296 | 0.0004543 | 16.0601 |
| | Mng | Batagai | 0.0059934 | 0.0004309 | 13.9103 |
| | | Standardbred | 0.0061431 | 0.000446 | 13.7722 |
| | Mor | CGG101397 | 0.003762 | 0.0005028 | 7.48162 |
| | | Mng | 0.0063176 | 0.0004522 | 13.9697 |
| | | Batagai | 0.006599 | 0.0004526 | 14.5799 |
| | | Prz_Przewalski | 0.0068147 | 0.0004638 | 14.693 |
| | | Standardbred | 0.0100456 | 0.0005609 | 17.91 |
| | | Std | 0.009453 | 0.0004856 | 19.4656 |
| | | Qrt | 0.009759 | 0.0004792 | 20.3662 |
| | Prz_D | CGG101397 | 0.0037824 | 0.0004861 | 7.78183 |
| | | Mon | 0.0065805 | 0.0004412 | 14.9146 |
| | | Mng | 0.0063717 | 0.0004478 | 14.2294 |
| | | Mor | 0.0067528 | 0.0004623 | 14.6073 |
| | | Standardbred | 0.006503 | 0.0004641 | 14.0128 |
| | | Qrt | 0.0063782 | 0.0004186 | 15.2357 |
| | | Std | 0.0065473 | 0.0004256 | 15.3834 |
| | | Batagai | 0.0105432 | 0.0004741 | 22.2384 |
| | | CGG10022 | 0.0102793 | 0.0004558 | 22.554 |
| | | CGG10023 | 0.0105099 | 0.0004737 | 22.185 |
| | | Prz_Przewalski | 0.0360248 | 0.0011137 | 32.3483 |
| | Prz_Przewalski | Mng | 0.0062061 | 0.0004523 | 13.7218 |
| | | Standardbred | 0.0064546 | 0.0004606 | 14.0141 |
| | | Batagai | 0.0103028 | 0.0004731 | 21.7764 |
| | Qrt | Batagai | 0.006306 | 0.0004239 | 14.8759 |
| | | Mng | 0.0063723 | 0.0004101 | 15.5393 |
| | | Prz_Przewalski | 0.006485 | 0.0004171 | 15.5469 |
| | | Standardbred | 0.0106147 | 0.0005088 | 20.862 |
| | | Std | 0.0106962 | 0.0005064 | 21.1233 |
| | Standardbred | Batagai | 0.0064287 | 0.0004621 | 13.9126 |
| | Std | Mng | 0.0063314 | 0.0004297 | 14.733 |
| | | Batagai | 0.0064543 | 0.0004242 | 15.2162 |
| | | Prz_Przewalski | 0.0063768 | 0.0004217 | 15.1219 |
| | | Standardbred | 0.0183513 | 0.0006818 | 26.9142 |
| **CGG101397** | Arabian | Batagai | 0.0031888 | 0.0004032 | 7.90933 |
| | | CGG10022 | 0.0031236 | 0.0003935 | 7.93779 |
| | | CGG10023 | 0.0030654 | 0.0004075 | 7.52177 |
| | | Fjord | 0.0032666 | 0.0004451 | 7.33967 |
| | | Prz_Przewalski | 0.0032379 | 0.0004115 | 7.86818 |
| | | Mng | 0.0032995 | 0.0004069 | 8.10977 |
| | | Prz_D | 0.0034155 | 0.0004109 | 8.31299 |
| | | Mon | 0.005533 | 0.000433 | 12.7775 |
| | | Mor | 0.0064021 | 0.0004968 | 12.8873 |
| | | Standardbred | 0.0075787 | 0.0005255 | 14.4222 |
| | | Std | 0.0072227 | 0.0004744 | 15.2244 |
| | | Qrt | 0.0080669 | 0.0004612 | 17.4914 |
| | CGG10022 | Mng | 0.0028527 | 0.0003711 | 7.68682 |
| | | Standardbred | 0.0030822 | 0.0003969 | 7.76547 |
| | | Mor | 0.003247 | 0.0004043 | 8.03033 |
| | | Qrt | 0.0033854 | 0.0003973 | 8.52023 |
| | | Std | 0.0032522 | 0.0003953 | 8.22658 |
| | | Mon | 0.0034523 | 0.0003771 | 9.15474 |
| | | Prz_Przewalski | 0.0066705 | 0.0004091 | 16.3044 |
| | | Batagai | 0.0197277 | 0.0004682 | 42.1386 |
| | | CGG10023 | 0.0197165 | 0.0004667 | 42.2462 |
| | CGG10023 | Mng | 0.0029754 | 0.0003622 | 8.21374 |
| | | Mor | 0.0034145 | 0.0003957 | 8.62913 |
| | | Standardbred | 0.0032911 | 0.0004014 | 8.19878 |
| | | Std | 0.0033254 | 0.0003855 | 8.62678 |
| | | Mon | 0.0033919 | 0.0003693 | 9.18371 |
| | | Qrt | 0.003487 | 0.000386 | 9.03462 |
| | | Prz_Przewalski | 0.0068941 | 0.0004216 | 16.3535 |
| | | Batagai | 0.0201651 | 0.0004793 | 42.0685 |
| | Fjord | Mng | 0.0030297 | 0.0004092 | 7.40356 |
| | | Batagai | 0.0032701 | 0.0003935 | 8.3113 |
| | | CGG10022 | 0.0033633 | 0.0003996 | 8.41731 |
| | | CGG10023 | 0.003226 | 0.0003901 | 8.26894 |
| | | Mor | 0.0038329 | 0.0004295 | 8.92382 |
| | | Prz_D | 0.0034312 | 0.0004034 | 8.50579 |
| | | Qrt | 0.0036222 | 0.0004209 | 8.6057 |
| | | Standardbred | 0.003748 | 0.0004288 | 8.74105 |
| | | Std | 0.003816 | 0.0004304 | 8.86526 |
| | | Mon | 0.0037215 | 0.0003865 | 9.62972 |
| | | Prz_Przewalski | 0.0035518 | 0.0003924 | 9.05229 |
| | Mon | Batagai | 0.0033307 | 0.0003726 | 8.9389 |
| | | Prz_Przewalski | 0.0032151 | 0.000382 | 8.41706 |
| | | Mng | 0.0034002 | 0.0003745 | 9.07912 |
| | | Mor | 0.0049109 | 0.0004228 | 11.6155 |
| | | Qrt | 0.0056913 | 0.0004118 | 13.819 |
| | | Standardbred | 0.0053653 | 0.000412 | 13.0226 |
| | | Std | 0.005459 | 0.0003957 | 13.7953 |
| | Icelandic | CGG10022 | 0.0033885 | 0.0004012 | 8.4451 |
| | | Mng | 0.0033436 | 0.0004036 | 8.28456 |
| | | Prz_D | 0.0033983 | 0.0004009 | 8.47696 |
| | | Prz_Przewalski | 0.0035623 | 0.000407 | 8.75187 |
| | | Arabian | 0.004027 | 0.0004279 | 9.41085 |
| | | Batagai | 0.0036022 | 0.0003958 | 9.10164 |
| | | CGG10023 | 0.0037901 | 0.0004029 | 9.40613 |
| | | Mon | 0.0038916 | 0.0003922 | 9.92314 |
| | | Mor | 0.0040059 | 0.0004262 | 9.39837 |
| | | Qrt | 0.0041171 | 0.0004205 | 9.79095 |
| | | Standardbred | 0.0041349 | 0.0004328 | 9.55406 |
| | | Std | 0.0039177 | 0.0004252 | 9.21432 |
| | | Fjord | 0.0046255 | 0.0004424 | 10.4546 |
| | Mng | Batagai | 0.002899 | 0.0003718 | 7.79687 |
| | | Standardbred | 0.0031873 | 0.0004133 | 7.71232 |
| | Mor | Batagai | 0.0032946 | 0.0003999 | 8.23946 |
| | | Mng | 0.0033037 | 0.0004074 | 8.11002 |
| | | Prz_Przewalski | 0.0034593 | 0.0004214 | 8.20914 |
| | | Standardbred | 0.0068799 | 0.0005291 | 13.0038 |
| | | Std | 0.0062566 | 0.0004812 | 13.0009 |
| | | Qrt | 0.0067644 | 0.0004559 | 14.8384 |
| | Prz_D | Standardbred | 0.003317 | 0.0004159 | 7.97488 |
| | | Mon | 0.0033375 | 0.0003854 | 8.65951 |
| | | Mng | 0.0033375 | 0.0003784 | 8.82013 |
| | | Mor | 0.0035087 | 0.0004298 | 8.1642 |
| | | Qrt | 0.0033633 | 0.0003973 | 8.46528 |
| | | Std | 0.0033305 | 0.0004116 | 8.09227 |
| | | Batagai | 0.0072185 | 0.0004211 | 17.141 |
| | | CGG10022 | 0.0070317 | 0.0004071 | 17.2715 |
| | | CGG10023 | 0.007078 | 0.0004106 | 17.2396 |
| | | Prz_Przewalski | 0.032649 | 0.001065 | 30.6576 |
| | Prz_Przewalski | Standardbred | 0.0031572 | 0.0004114 | 7.67518 |
| | | Mng | 0.0030605 | 0.0003711 | 8.24699 |
| | | Batagai | 0.0068668 | 0.0004233 | 16.221 |
| | Qrt | Batagai | 0.0032308 | 0.0003873 | 8.3412 |
| | | Mng | 0.0035876 | 0.0004046 | 8.86796 |
| | | Prz_Przewalski | 0.0033587 | 0.0003859 | 8.70378 |
| | | Standardbred | 0.0076782 | 0.0004972 | 15.4436 |
| | | Std | 0.0077289 | 0.0004817 | 16.0436 |
| | Standardbred | Batagai | 0.0031825 | 0.0003987 | 7.98131 |
| | Std | Prz_Przewalski | 0.0030486 | 0.0004116 | 7.40685 |
| | | Batagai | 0.0031772 | 0.0003818 | 8.32083 |
| | | Mng | 0.0033449 | 0.0004174 | 8.01304 |
| | | Standardbred | 0.015213 | 0.0007045 | 21.5933 |

# 7 Section 7: Genetic determinants of the Yakutian horse adaptations

Our demographic analyses (see **section S6**) show that contemporary Yakutian horses derive from a recent founder event, likely associated with the arrival of Yakut people in the area a few centuries ago (1, 2). It follows that Yakutian horses have evolved striking physiological and morphological adaptations to extreme climate conditions in a small number of generations. Here, we analyzed the genetic basis and evolutionary mechanisms underlying these adaptations.

## *7.1 Selection scans*

### 7.1.1 Genetically differentiated regions: $F_{ST}$-outlier approach

We estimated the $F_{ST}$ index by comparing the genomes of modern Yakutian horses against two different domestic populations, including one comprising 12 Franches-Montagnes horses (hereafter referred to as YAK-FM), and another one encompassing all 27 individuals from nine different domestic breeds (YAK-DOM). The genome of the 19th century specimen CGG101397 was included within the set of modern Yakutian horse genomes, as it is genetically undifferentiated from modern individuals (see **sections S5.3** and **S6**). A schematic description of the compared populations is summarized in **Table S7.1**.

The $F_{ST}$ values were estimated in 50 Kb sliding windows (step 10 Kb), using the ngsFst program of the ngsPopGen package (113). Briefly, for each compared population pair (i.e. YAK-DOM and YAK-FM), priors for the allele frequencies of each position (saf files) were estimated with ANGSD v0.615, which implements a statistical framework to integrate over genotype uncertainties (46). To generate saf files, we assumed Hardy-Weinberg equilibrium, and only considered autosomal sites with a minimum mapping quality of 25 (-minMapQ), a minimum base quality of 20 (-minBQ), and for which data were available in at least half of the individuals (-minInd). The saf files were then used as input information for the realSFS 2dsfs command in ANGSD, which generated the 2D Site Frequency Spectrum between each pair of populations considered. The saf files and the 2D SFS were then used as input for the ngsFST command from the ngsPopGen package (113).

Outlier $F_{ST}$ regions were detected following two different approaches. In the first approach (hereafter referred to as "Gene-max"), we simply extracted those genes located in regions ranking in the top-1% (or 5%) $F_{ST}$ values. This provided a total number of 251 (top-1%) or 1,255 (top-5%) gene candidates underpinning significant population differentiation for the two data sets considered (YAK-DOM and YAK-FM). In the second approach ("Region-peaks"), we applied the smoothing procedure described in the Supplementary section S7.2 of (30). This provided 312 (or 1,489) and 318 (1,350) outlier genes for the YAK-DOM and YAK-FM data sets, respectively. Significant outcomes "Region-peaks" smooting method are shown for five different chromosomes in **Figures S7.1** and **S7.2**, which are representative of the profiles observed across all chromosomes. **Table S7.2** summarizes the number of significant genes for each approach. **Table S7.3** shows the complete list of genes within the top-1% $F_{ST}$-outlier windows, and **Tables S7.4-S7.7** the corresponding enrichment tests.

Outlier $F_{ST}$ windows were found to be depleted of protein-coding genes **(Figure S7.3)**, pointing to a long-term effect of negative selection in reducing their genetic differentiation, but also to ongoing adaptive pressures in non-coding loci, and thus in potentially regulatory regions.

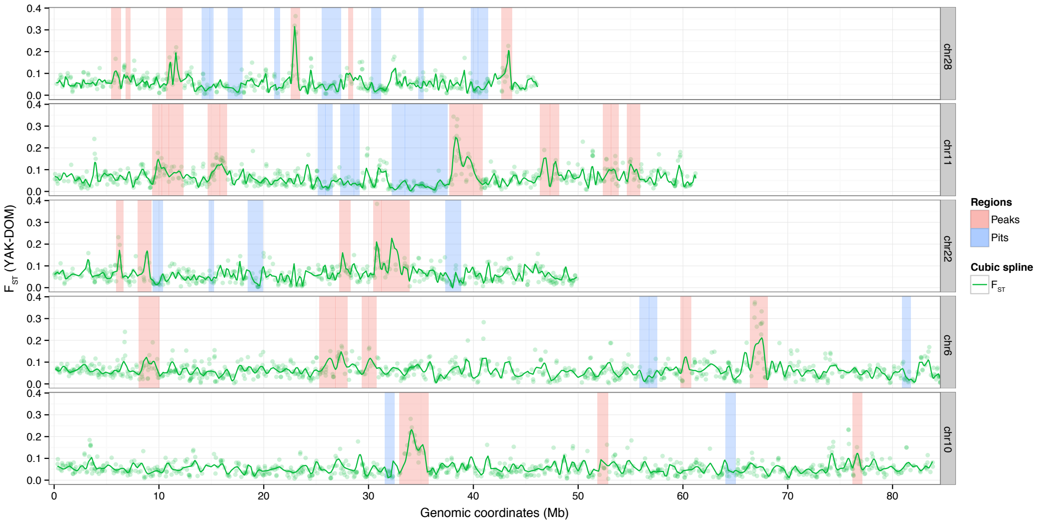


**Figure S7.1. Illustrative examples of $F_{ST}$ outlier regions identified by the "Region peaks" smoothing method (YAK-DOM).**

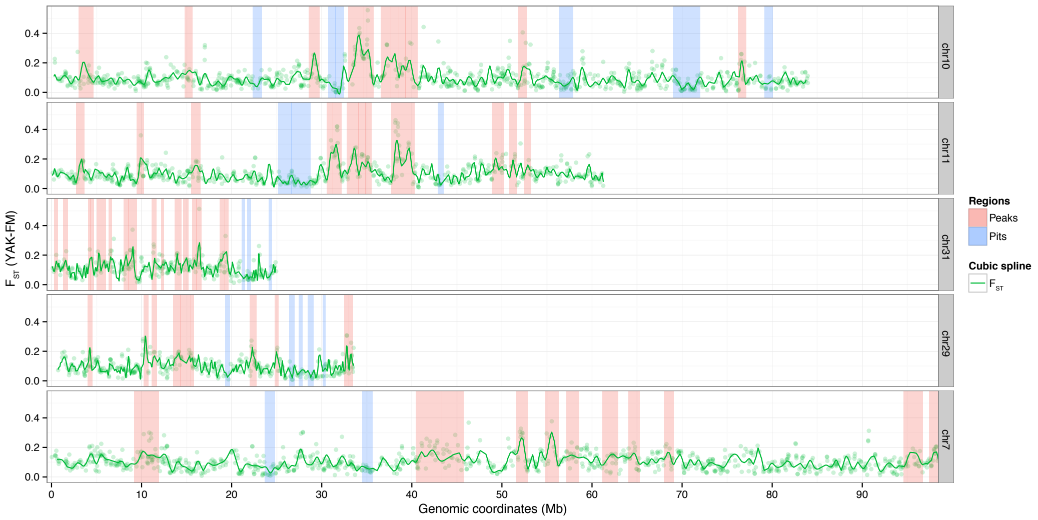


**Figure S7.2. Illustrative examples of $F_{ST}$ outlier regions identified by the "Region-peaks" smoothing method (YAK-FM).**

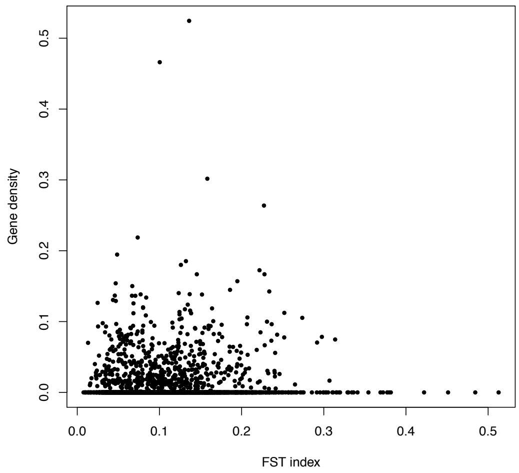


**Figure S7.3. Relationship between protein-coding gene density and $F_{ST}$ in 50Kb sliding windows.**

### *7.2 Coding vs. non-coding contribution to adaptation*

It is well known that mutations affecting the function of *cis*-regulatory elements (CREs) are a major force driving the evolution of transcriptional regulation, significantly contributing to the phenotypic variation within and between species (120). Nevertheless, the relative contribution of the CRE and non-synonymous mutations to adaptation remains one of the most controversial issues in evolutionary genetics (121–125).

At one end of the spectrum, adaptive changes are believed to mainly involve CRE mutations, because they can alter gene expression in very specific spatio-temporal frames, thereby fine-tuning transcriptional expression to optimize individual fitness (122). Unlike CRE mutations, protein-coding variants have a larger potential for high pleiotropic effects, as they can introduce phenotypic changes in all the developmental stages and/or cell types where the underlying gene is expressed. This may reduce the "evolvability" of the protein-coding regions, often yielding low evolutionary rates, especially if they have high levels of expression breadth, such as those housekeeping genes.

The critics of such extreme view argue that there are other mechanisms decreasing the functional constraints imposed by pleiotropy, such as increasing redundancy through gene duplication. Indeed, it is generally accepted that gene duplications are a major evolutionary mechanism for generating functional innovation, either through neofunctionalization or subfunctionalization, which may also play an important in adaptation (124, 126).

### 7.2.1 Genetic distances across functional site categories

We used current gene model annotations (Ensembl version 2.78) to classify the horse genome positions into functional categories. In particular, we split protein-coding regions into zero-fold (0d-fold), two-fold (2d-fold) and four-fold (4d-fold) positions, depending on their degeneracy. Mutations at zero-fold degenerate positions always result in amino acid replacement, implying non-synonymous changes. In contrast, mutations at four-fold positions, which are synonymous, provide proxies for sites putatively evolving under neutrality. Changes at two-fold positions involve both synonymous and non-synonymous changes. Additionally, we partitioned the 10Kb located upstream of Translation Start Sites (TSS) into non-overlapping 1Kb bins.

We then calculated the genetic distance (average net number of pairwise nucleotide differences, or $d_A$) (127) between the Yakutian horse population (including the sequence information from contemporary horses and the historical horse CGG1101397) and two panels of domesticated horses **(Table S7.1)**. The first encompasses all the 27 domesticated horses considered in this study (YAK-DOM, see **section S7.1.1**), while the second is restricted to the 12 horses within the Franches-Montagnes breed (YAK-FM, see **section S7.1.1**). The inclusion of different breeds in the YAK-DOM panel alleviates possible breed-specific selective processes, while YAK-FM reduces the bias caused by the inherent breed structure. We used mstatspop (http://bioinformatics.cragenomica.es/numgenomics/people/sebas/software/files/page3_4.zip) to calculate $d_A$ on those positions with up to 10% of missing alleles.

The neutral expectation of nucleotide divergence was approximated by the $d_A$ values at four-fold degenerate positions, using the top-0.001% quantile as the minimal threshold for detecting sites putatively evolving under positive selection (**Figure 4A** for the YAK-FM panel, and **Figure S7.4** for YAK-DOM). Comparing the proportion of adaptive mutations across sites is not straightforward, because the sample size is unbalanced across site categories, i.e. the more sites analyzed in a particular category, the higher the probability of observing positively selected mutations. We therefore applied a bootstrapping approach to circumvent this issue, re-sampling with replacement the same number of sites within each category (e.g. zero-fold, first 1Kb located upstream of TSS) as those analyzed for the neutral reference. For each of the 10,000 bootstrap replicates, we estimated the proportion of adaptive mutations within each site category, obtaining full data distributions that are shown in **Figure 4B** (YAK-FM panel) and **Figure S7.5** (YAK-DOM panel). The complete list of genes harbouring significant instances of adaptive alleles at their upstream regions is available upon request.

We also verified that the identification of dA-outliers was not biased by underlying segmental duplications, as the presence of unidentified paralogs could lead to spurious read alignments and, hence the calculation of erroneous dA values. To achieve this, we recovered any site under selection (as identified by the dA-based selection scans using DOM and FM as reference panels) that was also found in regions identified as segmental duplications (see **section S4.2**) extended by 1 Kb upstream and downstream of the duplicated region. The overlap was found to be negligible when considering both the FM (2/184 = 1.1%) and the DOM (14/1,486 = 0.9%) panels, suggesting a limited impact of segmental duplications on the detection of sites under selection. Similarly, only ~1% of the genes with dA outlier sites at their upstream regions were embedded within CNVs. Consequently, the functional enrichment analyses reported almost identical results (methods for functional enrichment analyses described in **section S7.3**), corroborating the biological significance of our findings.

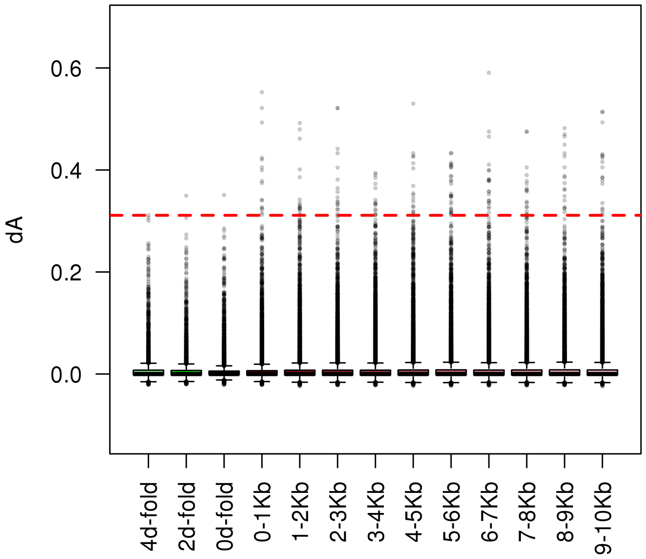


**Figure S7.4. Boxplot of the $d_A$ estimates across different site categories, for the YAK-DOM population pair.**

The red line delimits the threshold of neutrality (i.e. the top 0.001% of the 4d-fold degenerate sites). Points above this red line represent nucleotide positions genetically more differentiated than this threshold, and therefore sites positively selected since the ancestral population split.

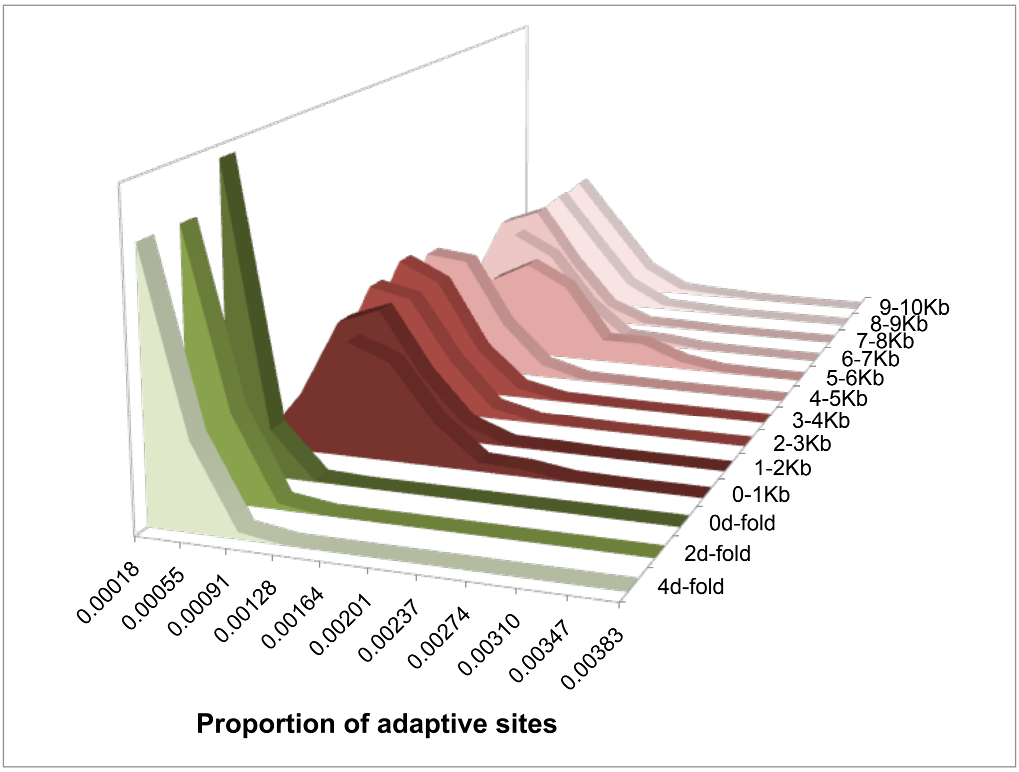


**Figure S7.5. Proportion of adaptive sites for each site category in the YAK-DOM population pair.**

### *7.3 Functional enrichment analyses*

We tested for functional enrichment among the genes targeted by positive selection, either at their protein-coding ($F_{ST}$) or upstream regions ($d_A$-based analysis). Functional enrichment was carried out on the online platform WebGestalt (WEB-based Gene SeT AnaLysis Toolkit; (128)), separately using the functional annotations of either the humans or mice ortholog genes as query (orthologs were downloaded from Ensembl, using the BioMart query tool), and the following databases:

- Pathways: KEGG (03/21/2011), WikiPathways (11/11/2012),
- Phenotypes: Mammalian Phenotype Ontology (for mouse gene ID, 04/10/2013) and Human Phenotype Ontology (for human gene ID 04/10/2013),
- Phenome-Wide Association Study: PheWAS catalog (05/20/2014; for human gene ID only),
- Disease associated genes: GLAD4U (1/26/2013, for human gene ID only).

We used default parameters and Benjamini-Hochberg correction for multiple tests (129), and retained the top-10 significant enrichment clusters. Enrichment results are in **Tables S4.4-S4.8** (for the ancestral informative markers)**, Tables S4.13-S4.15** (for the Copy Number Variants)**, Tables S7.4-S7.7** (for the $F_{ST}$-outlier windows)**,** and **S7.8-S7.13** (for $d_A$-outlier positions).

## *7.4 Supplementary Tables for Section 7*

### Table S7.1. Specimens used in $F_{ST}$ selection scan.

| Horse ID | Horse type | Reference | Reference panel(s) |
|---|---|---|---|
| Yak1 | Yakutian | this study | refYAK |
| Yak2 | Yakutian | this study | refYAK |
| Yak3 | Yakutian | this study | refYAK |
| Yak4 | Yakutian | this study | refYAK |
| Yak5 | Yakutian | this study | refYAK |
| Yak6 | Yakutian | this study | refYAK |
| Yak7 | Yakutian | this study | refYAK |
| Yak8 | Yakutian | this study | refYAK |
| Yak9 | Yakutian | this study | refYAK |
| CGG101397 | Yakutian 19th century | this study | refYAK |
| Arabian | Arabian | Orlando et al., 2013 (13) | refDOM |
| Fjord | Norwegian Fjord | Orlando et al., 2013 (13) | refDOM |
| Icelandic | Icelandic | Orlando et al., 2013 (13) | refDOM |
| Mng_D2628 | Mongolian | Do et al., 2014 (19) | refDOM |
| Mng_D2629 | Mongolian | Do et al., 2014 (19) | refDOM |
| Mor_EMS595 | Morgan | this study | refDOM |
| Qrt_A5964 | Quarter | this study | refDOM |
| Qrt_A5659 | Quarter | this study | refDOM |
| Qrt_A1543 | Quarter | this study | refDOM |
| Qrt_A2085 | Quarter | this study | refDOM |
| Std_M977 | Standardbred | this study | refDOM |
| Std_M5256 | Standardbred | this study | refDOM |
| Std_M1009 | Standardbred | this study | refDOM |
| Std_Standardbred | Standardbred | Orlando et al., 2013 (13) | refDOM |
| Thoroughbred | Thoroughbred | Wade et al., 2009 (29) | refDOM |
| Mon_FM1798 | Franches-Montagnes | this study | refDOM, refFM |
| Mon_FM1932 | Franches-Montagnes | this study | refDOM, refFM |
| Mon_FM1948 | Franches-Montagnes | this study | refDOM, refFM |
| Mon_FM2218 | Franches-Montagnes | this study | refDOM, refFM |
| Mon_FM1190 | Franches-Montagnes | this study | refDOM, refFM |
| Mon_FM1041 | Franches-Montagnes | this study | refDOM, refFM |
| Mon_FM1951 | Franches-Montagnes | this study | refDOM, refFM |
| Mon_FM0467 | Franches-Montagnes | this study | refDOM, refFM |
| Mon_FM_431 | Franches-Montagnes | this study | refDOM, refFM |
| Mon_FM1785 | Franches-Montagnes | this study | refDOM, refFM |
| Mon_FM1030 | Franches-Montagnes | this study | refDOM, refFM |
| Mon_FM0450 | Franches-Montagnes | this study | refDOM, refFM |

We compared Yakutian horses (including the CGG101397 specimen) against the YAK-DOM and YAK-FM panels.

**Table S7.2. Number of outlier genes identified through $F_{ST}$ selection scans.**

| Population 1 | Population 2 | Method | Significance threshold | Outlier genes |
|---|---|---|---|---|
| YAK | DOM | Region-peaks | 0.01 | 312 |
| | | | 0.05 | 1,489 |
| | | Gene-max | 0.01 | 251 |
| | | | 0.05 | 1,255 |
| YAK | FM | Region-peaks | 0.01 | 318 |
| | | | 0.05 | 1,350 |
| | | Gene-max | 0.01 | 251 |
| | | | 0.05 | 1,255 |

We used two different detection methods ("Gene-max" and "Region-peaks") and two significance thresholds ($\alpha = 0.01$ and 0.05).

## Table S7.3. Genes within the top 1% $F_{ST}$-outlier windows.

| GeneID | TranscriptID | QValue(Min) | QValue(Max) | Overlap | Biotype | Name | Description |
|---|---|---:|---:|---:|---|---|---|
| ENSECAG00000000384 | ENSECAT00000000362 | 6.58E-05 | 6.58E-05 | 1,108 | protein_coding | *ERP29* | endoplasmic reticulum protein 29 |
| ENSECAG00000000395 | ENSECAT00000000572 | 1.09E-02 | 3.67E-02 | 516 | protein_coding | *RAPGEF5* | Rap guanine nucleotide exchange factor (GEF) 5 |
| ENSECAG00000000420 | ENSECAT00000000470 | 6.58E-05 | 1.62E-04 | 1,526 | protein_coding | *BANP* | BTG3 associated nuclear protein |
| ENSECAG00000000658 | ENSECAT00000001905 | 1.97E-02 | 4.04E-02 | 1,190 | protein_coding | *GPNMB* | glycoprotein (transmembrane) nmb |
| ENSECAG00000001022 | ENSECAT00000000843 | 4.07E-03 | 4.07E-03 | 339 | protein_coding | *UCN2* | urocortin 2 |
| ENSECAG00000001103 | ENSECAT00000000912 | 6.58E-05 | 6.58E-05 | 825 | protein_coding | | Uncharacterized protein |
| ENSECAG00000001493 | ENSECAT00000001322 | 2.60E-04 | 2.60E-04 | 891 | protein_coding | | Uncharacterized protein |
| ENSECAG00000001843 | ENSECAT00000001659 | 1.04E-04 | 1.04E-04 | 681 | protein_coding | | Uncharacterized protein |
| ENSECAG00000001866 | ENSECAT00000001683 | 1.46E-04 | 1.46E-04 | 597 | protein_coding | | Uncharacterized protein |
| ENSECAG00000001999 | ENSECAT00000001963 | 2.61E-02 | 2.86E-02 | 258 | protein_coding | *GMDS* | GDP-mannose 4,6-dehydratase |
| ENSECAG00000002212 | ENSECAT00000002055 | 6.58E-05 | 6.58E-05 | 1,437 | protein_coding | *FOXO1* | forkhead box O1 |
| ENSECAG00000002894 | ENSECAT00000002734 | 1.62E-04 | 1.62E-04 | 294 | protein_coding | *SPATA45* | spermatogenesis associated 45 |
| ENSECAG00000002985 | ENSECAT00000002828 | 6.58E-05 | 6.58E-05 | 540 | protein_coding | | Uncharacterized protein |
| ENSECAG00000003561 | ENSECAT00000003425 | 1.32E-03 | 1.32E-03 | 519 | protein_coding | *TMEM11* | transmembrane protein 11 |
| ENSECAG00000003600 | ENSECAT00000003692 | 6.58E-05 | 6.58E-05 | 1,097 | protein_coding | *MRPS31* | mitochondrial ribosomal protein S31 |
| ENSECAG00000003664 | ENSECAT00000003504 | 1.62E-04 | 1.62E-04 | 1,290 | protein_coding | | Uncharacterized protein |
| ENSECAG00000003992 | ENSECAT00000003872 | 1.62E-04 | 1.62E-04 | 954 | protein_coding | | Uncharacterized protein |
| ENSECAG00000004727 | ENSECAT00000004599 | 3.38E-02 | 4.29E-02 | 1,059 | protein_coding | *S1PR2* | sphingosine-1-phosphate receptor 2 |
| ENSECAG00000004819 | ENSECAT00000008274 | 3.77E-03 | 3.12E-02 | 3,164 | protein_coding | *DNAH11* | dynein, axonemal, heavy chain 11 |
| ENSECAG00000005550 | ENSECAT00000005515 | 1.62E-04 | 1.62E-04 | 747 | protein_coding | *SFN* | stratifin |
| ENSECAG00000006237 | ENSECAT00000006194 | 1.27E-03 | 1.48E-03 | 1,526 | protein_coding | | Uncharacterized protein |
| ENSECAG00000006490 | ENSECAT00000006548 | 2.49E-03 | 2.49E-03 | 906 | protein_coding | *OR4K14* | olfactory receptor, family 4, subfamily K, member 14 |
| ENSECAG00000006608 | ENSECAT00000006575 | 8.96E-04 | 1.41E-03 | 872 | protein_coding | *OR4K13* | olfactory receptor, family 4, subfamily K, member 13 |
| ENSECAG00000006635 | ENSECAT00000006635 | 5.60E-04 | 5.60E-04 | 927 | protein_coding | *OR4N5* | olfactory receptor, family 4, subfamily N, member 5 |
| ENSECAG00000006720 | ENSECAT00000006683 | 1.62E-04 | 1.62E-04 | 939 | protein_coding | | Uncharacterized protein |
| ENSECAG00000007163 | ENSECAT00000007206 | 1.62E-04 | 1.62E-04 | 1,162 | protein_coding | *NR0B2* | nuclear receptor subfamily 0, group B, member 2 |
| ENSECAG00000007531 | ENSECAT00000007542 | 1.62E-04 | 1.62E-04 | 900 | protein_coding | *OR4C16* | olfactory receptor, family 4, subfamily C, member 16 (gene/pseudogene) |
| ENSECAG00000007653 | ENSECAT00000007936 | 1.62E-04 | 1.62E-04 | 1,591 | protein_coding | *GPATCH3* | G patch domain containing 3 |
| ENSECAG00000007790 | ENSECAT00000007819 | 1.73E-03 | 1.73E-03 | 903 | protein_coding | *OR4P4* | olfactory receptor, family 4, subfamily P, member 4 |
| ENSECAG00000008082 | ENSECAT00000008235 | 7.83E-03 | 2.83E-02 | 2,225 | protein_coding | *PJA2* | praja ring finger 2, E3 ubiquitin protein ligase |
| ENSECAG00000008109 | ENSECAT00000029102 | 3.20E-02 | 3.61E-02 | 694 | protein_coding | *BANK1* | B-cell scaffold protein with ankyrin repeats 1 |
| ENSECAG00000008110 | ENSECAT00000008250 | 6.58E-05 | 6.58E-05 | 1,368 | protein_coding | *TMEM116* | transmembrane protein 116 |
| ENSECAG00000008220 | ENSECAT00000008357 | 2.07E-02 | 2.07E-02 | 565 | protein_coding | *FDX1L* | ferredoxin 1-like |
| ENSECAG00000008513 | ENSECAT00000008623 | 1.62E-04 | 1.62E-04 | 828 | protein_coding | *TATDN3* | TatD DNase domain containing 3 |
| ENSECAG00000008638 | ENSECAT00000009209 | 6.95E-03 | 8.77E-03 | 522 | protein_coding | *RERE* | arginine-glutamic acid dipeptide (RE) repeats |
| ENSECAG00000008659 | ENSECAT00000008879 | 2.36E-04 | 3.31E-04 | 792 | protein_coding | *SHISA5* | shisa family member 5 |
| ENSECAG00000008754 | ENSECAT00000009442 | 6.58E-05 | 6.58E-05 | 2,913 | protein_coding | *NAA25* | N(alpha)-acetyltransferase 25, NatB auxiliary subunit |
| ENSECAG00000008821 | ENSECAT00000009135 | 6.58E-05 | 6.58E-05 | 1,086 | protein_coding | *FCAR* | Equus caballus Fc fragment of IgA, receptor for (FCAR), mRNA. |
| ENSECAG00000009092 | ENSECAT00000009245 | 1.50E-03 | 1.50E-03 | 930 | protein_coding | | Uncharacterized protein |
| ENSECAG00000009299 | ENSECAT00000009838 | 1.94E-02 | 1.94E-02 | 2,755 | protein_coding | *ICAM5* | intercellular adhesion molecule 5, telencephalin |
| ENSECAG00000009341 | ENSECAT00000009505 | 1.62E-04 | 1.62E-04 | 831 | protein_coding | | Uncharacterized protein |
| ENSECAG00000009355 | ENSECAT00000009653 | 1.62E-04 | 1.62E-04 | 944 | protein_coding | *NSL1* | NSL1, MIS12 kinetochore complex component |
| ENSECAG00000009384 | ENSECAT00000009603 | 1.62E-04 | 9.51E-03 | 1,122 | protein_coding | *NECAB1* | N-terminal EF-hand calcium binding protein 1 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| ENSECAG00000009599 | ENSECAT00000009793 | 6.58E-05 | 6.58E-05 | 3,736 | protein_coding | | Uncharacterized protein |
| ENSECAG00000009723 | ENSECAT00000010381 | 3.03E-02 | 3.03E-02 | 739 | protein_coding | *IGF2BP1* | insulin-like growth factor 2 mRNA binding protein 1 |
| ENSECAG00000009737 | ENSECAT00000009944 | 6.58E-05 | 6.58E-05 | 622 | protein_coding | *IL17C* | interleukin 17C |
| ENSECAG00000009769 | ENSECAT00000010081 | 1.11E-03 | 3.29E-02 | 1,303 | protein_coding | *CNTN4* | contactin 4 |
| ENSECAG00000009796 | ENSECAT00000010439 | 5.43E-04 | 9.58E-03 | 862 | protein_coding | *FER* | fer (fps/fes related) tyrosine kinase |
| ENSECAG00000010327 | ENSECAT00000011267 | 4.81E-03 | 1.33E-02 | 590 | protein_coding | *INPP4B* | inositol polyphosphate-4-phosphatase, type II, 105kDa |
| ENSECAG00000010723 | ENSECAT00000011174 | 1.62E-04 | 1.62E-04 | 1,656 | protein_coding | *ANGEL2* | angel homolog 2 (Drosophila) |
| ENSECAG00000010761 | ENSECAT00000011071 | 2.36E-04 | 2.36E-04 | 1,021 | protein_coding | *TREX1* | three prime repair exonuclease 1 |
| ENSECAG00000010784 | ENSECAT00000011297 | 2.36E-04 | 3.11E-04 | 2,200 | protein_coding | *ATRIP* | ATR interacting protein |
| ENSECAG00000010790 | ENSECAT00000011790 | 4.81E-03 | 3.81E-02 | 1,864 | protein_coding | *COG6* | component of oligomeric golgi complex 6 |
| ENSECAG00000010806 | ENSECAT00000011160 | 6.58E-05 | 6.58E-05 | 1,095 | protein_coding | | Uncharacterized protein |
| ENSECAG00000010965 | ENSECAT00000011277 | 6.58E-05 | 6.58E-05 | 645 | protein_coding | *CYBA* | cytochrome b-245, alpha polypeptide |
| ENSECAG00000011042 | ENSECAT00000011367 | 2.36E-04 | 3.31E-04 | 297 | protein_coding | *BATF3* | basic leucine zipper transcription factor, ATF-like 3 |
| ENSECAG00000011206 | ENSECAT00000011581 | 6.58E-05 | 1.46E-02 | 1,308 | protein_coding | *MVD* | mevalonate (diphospho) decarboxylase |
| ENSECAG00000011234 | ENSECAT00000012935 | 4.07E-03 | 1.43E-02 | 6,098 | protein_coding | *COL7A1* | collagen, type VII, alpha 1 |
| ENSECAG00000011296 | ENSECAT00000011993 | 1.62E-04 | 7.05E-04 | 2,991 | protein_coding | *SLC9A1* | solute carrier family 9, subfamily A (NHE1, cation proton antiporter 1), member 1 |
| ENSECAG00000011483 | ENSECAT00000011960 | 2.28E-02 | 2.60E-02 | 381 | protein_coding | *MORN1* | MORN repeat containing 1 |
| ENSECAG00000011486 | ENSECAT00000011931 | 3.15E-03 | 4.07E-03 | 1,198 | protein_coding | *ATF3* | activating transcription factor 3 |
| ENSECAG00000011508 | ENSECAT00000012117 | 6.58E-05 | 6.58E-05 | 1,276 | protein_coding | *TRA2A* | transformer 2 alpha homolog (Drosophila) |
| ENSECAG00000011599 | ENSECAT00000012512 | 6.58E-05 | 6.58E-05 | 2,846 | protein_coding | *ZC3H18* | zinc finger CCCH-type containing 18 |
| ENSECAG00000012131 | ENSECAT00000012555 | 3.84E-02 | 3.84E-02 | 846 | protein_coding | *SNAI3* | snail family zinc finger 3 |
| ENSECAG00000012403 | ENSECAT00000012989 | 1.62E-04 | 1.62E-04 | 1,458 | protein_coding | *GPN2* | GPN-loop GTPase 2 |
| ENSECAG00000012543 | ENSECAT00000013268 | 2.56E-02 | 2.56E-02 | 826 | protein_coding | *C-SKI* | Equus caballus SKI (C-SKI), mRNA. |
| ENSECAG00000012669 | ENSECAT00000013543 | 1.71E-02 | 4.31E-02 | 726 | protein_coding | *TANC2* | tetratricopeptide repeat, ankyrin repeat and coiled-coil containing 2 |
| ENSECAG00000012768 | ENSECAT00000013265 | 6.58E-05 | 6.58E-05 | 1,258 | protein_coding | | Uncharacterized protein |
| ENSECAG00000012805 | ENSECAT00000013279 | 1.94E-02 | 2.01E-02 | 818 | protein_coding | *ICAM4* | intercellular adhesion molecule 4 (Landsteiner-Wiener blood group) |
| ENSECAG00000013354 | ENSECAT00000014009 | 4.00E-03 | 5.20E-03 | 752 | protein_coding | *B4GALNT2* | beta-1,4-N-acetyl-galactosaminyl transferase 2 |
| ENSECAG00000013436 | ENSECAT00000014189 | 3.17E-02 | 3.95E-02 | 848 | protein_coding | *TCEA1* | transcription elongation factor A (SII), 1 |
| ENSECAG00000013539 | ENSECAT00000014564 | 1.25E-03 | 1.87E-02 | 1,321 | protein_coding | *ITFG1* | integrin alpha FG-GAP repeat containing 1 |
| ENSECAG00000013677 | ENSECAT00000016045 | 1.62E-04 | 5.79E-04 | 8,088 | protein_coding | *FRY* | furry homolog (Drosophila) |
| ENSECAG00000013762 | ENSECAT00000014568 | 3.21E-02 | 3.21E-02 | 924 | protein_coding | *NCR1* | Natural cytotoxicity triggering receptor 1; Uncharacterized protein |
| ENSECAG00000013883 | ENSECAT00000014969 | 2.01E-02 | 2.01E-02 | 1,584 | protein_coding | *ICAM1* | intercellular adhesion molecule 1 |
| ENSECAG00000014089 | ENSECAT00000014815 | 5.06E-03 | 4.60E-02 | 3,639 | protein_coding | *NFATC3* | nuclear factor of activated T-cells, cytoplasmic, calcineurin-dependent 3 |
| ENSECAG00000014169 | ENSECAT00000014974 | 1.91E-03 | 3.34E-02 | 1,908 | protein_coding | *CCSER2* | coiled-coil serine-rich protein 2 |
| ENSECAG00000014285 | ENSECAT00000014882 | 1.27E-02 | 1.33E-02 | 958 | protein_coding | *INHBB* | inhibin, beta B |
| ENSECAG00000014543 | ENSECAT00000015156 | 3.31E-04 | 5.43E-04 | 1,476 | protein_coding | *CCDC51* | coiled-coil domain containing 51 |
| ENSECAG00000014555 | ENSECAT00000015455 | 6.60E-03 | 6.60E-03 | 158 | protein_coding | *PLA2G15* | phospholipase A2, group XV |
| ENSECAG00000014577 | ENSECAT00000015968 | 1.62E-04 | 1.62E-04 | 2,054 | protein_coding | *ADAM28* | ADAM metallopeptidase domain 28 |
| ENSECAG00000014767 | ENSECAT00000015985 | 5.43E-04 | 5.43E-04 | 287 | protein_coding | *PLXNB1* | plexin B1 |
| ENSECAG00000014888 | ENSECAT00000015647 | 5.60E-04 | 2.81E-03 | 2,087 | protein_coding | *NEK5* | NIMA-related kinase 5 |
| ENSECAG00000015180 | ENSECAT00000016039 | 2.70E-02 | 3.37E-02 | 1,299 | protein_coding | *NEXN* | nexilin (F actin binding protein) |
| ENSECAG00000015839 | ENSECAT00000016675 | 1.62E-04 | 1.62E-04 | 831 | protein_coding | *ZDHHC18* | zinc finger, DHHC-type containing 18 |
| ENSECAG00000015846 | ENSECAT00000016670 | 1.62E-04 | 1.62E-04 | 1,063 | protein_coding | *VASH2* | vasohibin 2 |
| ENSECAG00000015967 | ENSECAT00000016728 | 5.65E-03 | 2.43E-02 | 644 | protein_coding | *ASF1B* | anti-silencing function 1B histone chaperone |
| ENSECAG00000016105 | ENSECAT00000016879 | 6.53E-03 | 8.53E-03 | 540 | protein_coding | *MALSU1* | mitochondrial assembly of ribosomal large subunit 1 |
| ENSECAG00000016156 | ENSECAT00000016906 | 6.58E-05 | 6.58E-05 | 1,080 | protein_coding | | Uncharacterized protein |
| ENSECAG00000016313 | ENSECAT00000017650 | 6.58E-05 | 6.58E-05 | 2,966 | protein_coding | *PTPN4* | protein tyrosine phosphatase, non-receptor type 4 (megakaryocyte) |

|  |  |  |  |  |  |  |  |
|---|---|---|---|---|---|---|---|
| ENSECAG00000016336 | ENSECAT00000017427 | 1.22E-02 | 1.90E-02 | 1,745 | protein_coding | *CDC25A* | cell division cycle 25A |
| ENSECAG00000016418 | ENSECAT00000018285 | 6.58E-05 | 6.58E-05 | 1,487 | protein_coding | *KIR3DL* | Equus caballus killer cell immunoglobulin-like receptor with three domains and long cytoplasmic tail (KIR3DL), mRNA. |
| ENSECAG00000016637 | ENSECAT00000018108 | 6.58E-05 | 1.04E-04 | 2,421 | protein_coding | *EPB41L5* | erythrocyte membrane protein band 4.1 like 5 |
| ENSECAG00000016792 | ENSECAT00000017817 | 3.07E-02 | 3.09E-02 | 2,520 | protein_coding |  | Uncharacterized protein |
| ENSECAG00000016797 | ENSECAT00000018011 | 6.58E-05 | 6.53E-03 | 2,129 | protein_coding | *IGF2BP3* | insulin-like growth factor 2 mRNA binding protein 3 |
| ENSECAG00000017013 | ENSECAT00000018069 | 2.30E-02 | 2.69E-02 | 771 | protein_coding | *MRPL4* | mitochondrial ribosomal protein L4 |
| ENSECAG00000017119 | ENSECAT00000017998 | 1.62E-04 | 1.62E-04 | 1,671 | protein_coding | *PIGV* | phosphatidylinositol glycan anchor biosynthesis, class V |
| ENSECAG00000017338 | ENSECAT00000019133 | 5.24E-04 | 1.59E-03 | 1,449 | protein_coding | *PFKFB4* | 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 4 |
| ENSECAG00000017375 | ENSECAT00000018683 | 5.36E-03 | 5.36E-03 | 2,126 | protein_coding | *ESRP2* | epithelial splicing regulatory protein 2 |
| ENSECAG00000017458 | ENSECAT00000018629 | 4.35E-03 | 4.35E-03 | 297 | protein_coding | *BNIP3* | BCL2/adenovirus E1B 19kDa interacting protein 3 |
| ENSECAG00000017514 | ENSECAT00000018710 | 1.62E-04 | 1.62E-04 | 812 | protein_coding |  | 60S ribosomal protein L6 |
| ENSECAG00000017527 | ENSECAT00000018479 | 1.86E-02 | 2.14E-02 | 131 | protein_coding | *NDUFAF2* | NADH dehydrogenase (ubiquinone) complex I, assembly factor 2 |
| ENSECAG00000017787 | ENSECAT00000019051 | 1.62E-04 | 1.62E-04 | 1,765 | protein_coding | *FLVCR1* | feline leukemia virus subgroup C cellular receptor 1 |
| ENSECAG00000017894 | ENSECAT00000018991 | 1.62E-04 | 1.62E-04 | 2,203 | protein_coding | *CKAP2* | cytoskeleton associated protein 2 |
| ENSECAG00000017967 | ENSECAT00000018959 | 1.85E-02 | 1.85E-02 | 226 | protein_coding | *METAP1* | methionyl aminopeptidase 1 |
| ENSECAG00000017991 | ENSECAT00000019703 | 1.62E-04 | 3.31E-04 | 5,032 | protein_coding | *ARID1A* | AT rich interactive domain 1A (SWI-like) |
| ENSECAG00000018146 | ENSECAT00000019570 | 1.04E-04 | 1.62E-04 | 434 | protein_coding | *VPS36* | vacuolar protein sorting 36 homolog (S. cerevisiae) |
| ENSECAG00000018216 | ENSECAT00000019228 | 1.04E-04 | 1.04E-04 | 216 | protein_coding |  |  |
| ENSECAG00000018240 | ENSECAT00000019411 | 1.47E-03 | 4.08E-03 | 617 | protein_coding | *PHKB* | phosphorylase b kinase regulatory subunit beta |
| ENSECAG00000018281 | ENSECAT00000019301 | 1.04E-04 | 1.04E-04 | 1,425 | protein_coding | *LENG9* | leukocyte receptor cluster (LRC) member 9 |
| ENSECAG00000018321 | ENSECAT00000019431 | 6.31E-03 | 1.43E-02 | 1,563 | protein_coding | *ALG11* | ALG11, alpha-1,2-mannosyltransferase |
| ENSECAG00000018343 | ENSECAT00000019558 | 1.04E-04 | 1.04E-04 | 2,867 | protein_coding | *LENG8* | leukocyte receptor cluster (LRC) member 8 |
| ENSECAG00000018372 | ENSECAT00000019403 | 1.04E-04 | 1.04E-04 | 1,401 | protein_coding |  | Uncharacterized protein |
| ENSECAG00000018510 | ENSECAT00000019719 | 3.34E-02 | 3.67E-02 | 806 | protein_coding | *RGS20* | regulator of G-protein signaling 20 |
| ENSECAG00000018676 | ENSECAT00000019971 | 1.62E-04 | 1.62E-04 | 1,281 | protein_coding | *RAD52* | RAD52 homolog (S. cerevisiae) |
| ENSECAG00000018785 | ENSECAT00000019859 | 1.69E-03 | 1.69E-03 | 393 | protein_coding |  | Uncharacterized protein |
| ENSECAG00000018835 | ENSECAT00000020002 | 1.62E-04 | 1.62E-04 | 1,601 | protein_coding | *FAM46B* | family with sequence similarity 46, member B |
| ENSECAG00000019161 | ENSECAT00000021056 | 4.37E-02 | 4.37E-02 | 788 | protein_coding | *DIAPH3* | diaphanous-related formin 3 |
| ENSECAG00000019742 | ENSECAT00000020901 | 6.58E-05 | 6.58E-05 | 576 | protein_coding | *CCDC126* | coiled-coil domain containing 126 |
| ENSECAG00000019817 | ENSECAT00000021018 | 6.46E-03 | 6.46E-03 | 347 | protein_coding | *SNRPF* | small nuclear ribonucleoprotein polypeptide F |
| ENSECAG00000019922 | ENSECAT00000021251 | 1.62E-04 | 1.62E-04 | 1,383 | protein_coding | *ADAMDEC1* | ADAM-like, decysin 1 |
| ENSECAG00000019943 | ENSECAT00000021835 | 1.43E-02 | 1.43E-02 | 140 | protein_coding | *ATP7B* | ATPase, Cu++ transporting, beta polypeptide |
| ENSECAG00000020102 | ENSECAT00000021747 | 6.58E-05 | 1.62E-04 | 3,368 | protein_coding | *ZEB1* | zinc finger E-box binding homeobox 1 |
| ENSECAG00000020537 | ENSECAT00000022031 | 1.62E-04 | 1.18E-03 | 1,370 | protein_coding | *ADAM7* | ADAM metallopeptidase domain 7 |
| ENSECAG00000020561 | ENSECAT00000021824 | 1.62E-04 | 1.62E-04 | 1,444 | protein_coding | *KDF1* | keratinocyte differentiation factor 1 |
| ENSECAG00000020725 | ENSECAT00000021995 | 2.75E-02 | 2.75E-02 | 1,342 | protein_coding |  | Uncharacterized protein |
| ENSECAG00000021107 | ENSECAT00000022541 | 6.58E-05 | 6.58E-05 | 2,079 | protein_coding | *TRAFD1* | TRAF-type zinc finger domain containing 1 |
| ENSECAG00000021311 | ENSECAT00000022834 | 1.62E-04 | 1.62E-04 | 1,139 | protein_coding | *NUDC* | nudC nuclear distribution protein |
| ENSECAG00000021316 | ENSECAT00000022692 | 6.58E-05 | 6.58E-05 | 881 | protein_coding | *FAM221A* | family with sequence similarity 221, member A |
| ENSECAG00000021555 | ENSECAT00000023706 | 8.59E-04 | 1.54E-03 | 3,350 | protein_coding | *LPHN1* | latrophilin 1 |
| ENSECAG00000021688 | ENSECAT00000023337 | 2.36E-04 | 5.60E-04 | 1,579 | protein_coding | *NEK3* | NIMA-related kinase 3 |
| ENSECAG00000021706 | ENSECAT00000023249 | 1.62E-04 | 3.72E-02 | 1,727 | protein_coding | *RPS6KC1* | ribosomal protein S6 kinase, 52kDa, polypeptide 1 |
| ENSECAG00000021735 | ENSECAT00000023334 | 1.85E-03 | 2.46E-03 | 492 | protein_coding | *EPHA6* | EPH receptor A6 |
| ENSECAG00000021801 | ENSECAT00000023283 | 9.33E-03 | 9.33E-03 | 414 | protein_coding | *CCDC38* | coiled-coil domain containing 38 |
| ENSECAG00000021842 | ENSECAT00000023428 | 6.58E-05 | 6.58E-05 | 3,165 | protein_coding | *STK31* | Equus caballus serine/threonine kinase 31 (STK31), mRNA. |
| ENSECAG00000021879 | ENSECAT00000023271 | 2.02E-02 | 2.02E-02 | 399 | protein_coding |  | Uncharacterized protein |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| ENSECAG00000022031 | ENSECAT00000023479 | 4.93E-02 | 4.93E-02 | 196 | protein_coding | *GRID2* | glutamate receptor, ionotropic, delta 2 |
| ENSECAG00000022100 | ENSECAT00000023675 | 5.60E-04 | 2.82E-03 | 1,001 | protein_coding | *DHRS7B* | dehydrogenase/reductase (SDR family) member 7B |
| ENSECAG00000022220 | ENSECAT00000023894 | 1.86E-02 | 1.99E-02 | 276 | protein_coding | | Uncharacterized protein |
| ENSECAG00000022238 | ENSECAT00000023679 | 4.52E-04 | 5.60E-04 | 234 | protein_coding | *NATD1* | N-acetyltransferase domain containing 1 |
| ENSECAG00000022242 | ENSECAT00000024114 | 1.62E-04 | 1.62E-04 | 1,031 | protein_coding | *MAP2K3* | mitogen-activated protein kinase kinase 3 |
| ENSECAG00000022846 | ENSECAT00000024378 | 1.56E-02 | 1.56E-02 | 945 | protein_coding | | Uncharacterized protein |
| ENSECAG00000022854 | ENSECAT00000024382 | 1.56E-02 | 1.56E-02 | 945 | protein_coding | | Uncharacterized protein |
| ENSECAG00000022858 | ENSECAT00000024390 | 4.92E-03 | 4.92E-03 | 945 | protein_coding | | Uncharacterized protein |
| ENSECAG00000022863 | ENSECAT00000024401 | 3.67E-03 | 3.67E-03 | 669 | protein_coding | | Uncharacterized protein |
| ENSECAG00000022876 | ENSECAT00000024415 | 1.50E-03 | 1.50E-03 | 960 | protein_coding | | Uncharacterized protein |
| ENSECAG00000022882 | ENSECAT00000024420 | 1.50E-03 | 1.50E-03 | 945 | protein_coding | | Uncharacterized protein |
| ENSECAG00000023324 | ENSECAT00000025006 | 7.82E-04 | 1.04E-03 | 1,719 | protein_coding | *SP4* | Sp4 transcription factor |
| ENSECAG00000023403 | ENSECAT00000025178 | 1.25E-03 | 4.35E-03 | 3,277 | protein_coding | *OTUD4* | OTU deubiquitinase 4 |
| ENSECAG00000023501 | ENSECAT00000025447 | 1.62E-04 | 4.52E-04 | 5,396 | protein_coding | *WNK1* | WNK lysine deficient protein kinase 1 |
| ENSECAG00000023602 | ENSECAT00000025656 | 6.58E-05 | 1.04E-04 | 11,976 | protein_coding | *HECTD4* | HECT domain containing E3 ubiquitin protein ligase 4 |
| ENSECAG00000023648 | ENSECAT00000025599 | 6.15E-04 | 1.04E-03 | 1,909 | protein_coding | *ABCE1* | ATP-binding cassette, sub-family E (OABP), member 1 |
| ENSECAG00000023701 | ENSECAT00000025441 | 2.09E-03 | 1.23E-02 | 474 | protein_coding | *CA5A* | carbonic anhydrase VA, mitochondrial |
| ENSECAG00000024340 | ENSECAT00000026285 | 1.04E-04 | 1.04E-04 | 1,230 | protein_coding | *TTYH1* | tweety family member 1 |
| ENSECAG00000024416 | ENSECAT00000026243 | 6.15E-04 | 1.48E-03 | 682 | protein_coding | *ANAPC10* | anaphase promoting complex subunit 10 |
| ENSECAG00000024476 | ENSECAT00000026328 | 6.15E-04 | 6.36E-04 | 3,822 | protein_coding | | Uncharacterized protein |
| ENSECAG00000024560 | ENSECAT00000026567 | 1.09E-03 | 3.45E-03 | 753 | protein_coding | *PTPN11* | protein tyrosine phosphatase, non-receptor type 11 |
| ENSECAG00000024908 | ENSECAT00000026919 | 1.04E-04 | 1.62E-04 | 2,836 | protein_coding | *ERC1* | ELKS/RAB6-interacting/CAST family member 1 |
| ENSECAG00000024966 | ENSECAT00000026953 | 2.56E-02 | 2.56E-02 | 1,856 | protein_coding | *RAVER1* | ribonucleoprotein, PTB-binding 1 |
| ENSECAG00000026276 | ENSECAT00000028288 | 1.43E-02 | 1.43E-02 | 61 | miRNA | eca-mir-711 | eca-mir-711 |
| ENSECAG00000027193 | ENSECAT00000029359 | 3.12E-02 | 3.12E-02 | 103 | miRNA | | |
| ENSECAG00000027275 | ENSECAT00000029441 | 1.46E-04 | 1.46E-04 | 75 | miRNA | | |
| ENSECAG00000026627 | ENSECAT00000028639 | 1.62E-04 | 1.62E-04 | 128 | rRNA | 5S_rRNA | 5S ribosomal RNA |
| ENSECAG00000026538 | ENSECAT00000028550 | 8.95E-03 | 8.95E-03 | 146 | snoRNA | SNORA12 | Small nucleolar RNA SNORA12 |
| ENSECAG00000027212 | ENSECAT00000029378 | 1.62E-04 | 1.62E-04 | 72 | snoRNA | SNORD112 | Small nucleolar RNA SNORD112 |
| ENSECAG00000026565 | ENSECAT00000028577 | 6.58E-05 | 6.58E-05 | 108 | snRNA | U6 | U6 spliceosomal RNA |

## Table S7.4: GO-term enrichment for genes showing 1% top $F_{ST}$ values.

| Panel | Model Organism | GO terms | Adjusted p-value | Number of Genes |
|---|---|---|---|---|
| DOM | human | fluid transport GO:0042044 | 1.60E-03 | 6 |
| | | water transport GO:0006833 | 1.60E-03 | 6 |
| | | positive regulation of response to nutrient levels GO:0032109 | 5.62E-02 | 3 |
| | | positive regulation of response to extracellular stimulus GO:0032106 | 5.62E-02 | 3 |
| | | cellular component organization GO:0016043 | 8.99E-02 | 55 |
| | | cellular component organization or biogenesis GO:0071840 | 9.37E-02 | 56 |
| | | neuron apoptotic process GO:0051402 | 9.63E-02 | 7 |
| | | neuron death GO:0070997 | 9.84E-02 | 7 |
| | | natural killer cell chemotaxis GO:0035747 | 1.17E-01 | 2 |
| | | eosinophil chemotaxis GO:0048245 | 1.17E-01 | 2 |
| | | water transmembrane transporter activity GO:0005372 | 3.00E-04 | 4 |
| | | water channel activity GO:0015250 | 3.00E-04 | 4 |
| | | anion binding GO:0043168 | 9.00E-04 | 41 |
| | | purine ribonucleotide binding GO:0032555 | 1.20E-03 | 33 |
| | | purine ribonucleoside triphosphate binding GO:0035639 | 1.20E-03 | 33 |
| | | purine ribonucleoside binding GO:0032550 | 1.20E-03 | 33 |
| | | Rab GTPase binding GO:0017137 | 1.20E-03 | 5 |
| | | ribonucleoside binding GO:0032549 | 1.20E-03 | 33 |
| | | ribonucleotide binding GO:0032553 | 1.20E-03 | 33 |
| | | purine nucleoside binding GO:0001883 | 1.20E-03 | 33 |
| | | neuronal cell body GO:0043025 | 2.14E-01 | 8 |
| | | mitotic spindle GO:0072686 | 2.14E-01 | 2 |
| | | cell body GO:0044297 | 2.14E-01 | 8 |
| | | cytoplasmic part GO:0044444 | 2.14E-01 | 73 |
| | | dendritic spine GO:0043197 | 2.14E-01 | 5 |
| | | dendrite GO:0030425 | 2.14E-01 | 8 |
| | | endoplasmic reticulum GO:0005783 | 2.14E-01 | 20 |
| | | neuron spine GO:0044309 | 2.14E-01 | 5 |
| | | extracellular region part GO:0044421 | 2.14E-01 | 18 |
| | | neuron projection GO:0043005 | 2.14E-01 | 13 |
| | mouse | water transport GO:0006833 | 4.60E-03 | 4 |
| | | fluid transport GO:0042044 | 5.70E-03 | 4 |
| | | protein splicing GO:0030908 | 5.90E-03 | 2 |
| | | phosphorus metabolic process GO:0006793 | 5.90E-03 | 35 |
| | | phosphate-containing compound metabolic process GO:0006796 | 5.90E-03 | 34 |
| | | regulation of apoptotic process GO:0042981 | 5.90E-03 | 20 |
| | | cellular process GO:0009987 | 5.90E-03 | 114 |
| | | organophosphate catabolic process GO:0046434 | 5.90E-03 | 13 |
| | | regulation of programmed cell death GO:0043067 | 5.90E-03 | 20 |
| | | apoptotic process GO:0006915 | 5.90E-03 | 23 |
| | | protein binding GO:0005515 | 5.15E-06 | 76 |
| | | anion binding GO:0043168 | 7.38E-06 | 39 |
| | | binding GO:0005488 | 1.76E-05 | 108 |
| | | water channel activity GO:0015250 | 2.31E-05 | 4 |
| | | water transmembrane transporter activity GO:0005372 | 2.31E-05 | 4 |
| | | ribonucleoside binding GO:0032549 | 2.31E-05 | 31 |
| | | purine ribonucleoside triphosphate binding GO:0035639 | 2.31E-05 | 31 |
| | | purine nucleoside binding GO:0001883 | 2.31E-05 | 31 |
| | | purine ribonucleoside binding GO:0032550 | 2.31E-05 | 31 |

| | | | | |
|---|---|---|---|---|
| | | nucleoside binding GO:0001882 | 2.34E-05 | 31 |
| | | cell part GO:0044464 | 1.00E-04 | 120 |
| | | cell GO:0005623 | 1.00E-04 | 120 |
| | | intracellular part GO:0044424 | 6.00E-04 | 104 |
| | | intracellular GO:0005622 | 1.30E-03 | 104 |
| | | cytoplasmic part GO:0044444 | 1.50E-03 | 61 |
| | | cytoplasm GO:0005737 | 1.70E-03 | 82 |
| | | organelle GO:0043226 | 1.70E-03 | 91 |
| | | intracellular organelle GO:0043229 | 1.70E-03 | 91 |
| | | dendritic spine GO:0043197 | 2.80E-03 | 7 |
| | | neuron spine GO:0044309 | 2.80E-03 | 7 |
| FM | human | positive regulation of response to food GO:0032097 | 6.99E-02 | 2 |
| | | response to food GO:0032094 | 6.99E-02 | 4 |
| | | positive regulation of appetite GO:0032100 | 6.99E-02 | 2 |
| | | cellular trivalent inorganic anion homeostasis GO:0072502 | 1.80E-01 | 2 |
| | | heme transport GO:0015886 | 1.80E-01 | 2 |
| | | neurological system process GO:0050877 | 1.80E-01 | 23 |
| | | cellular phosphate ion homeostasis GO:0030643 | 1.80E-01 | 2 |
| | | positive regulation of multicellular organism growth GO:0040018 | 1.85E-01 | 3 |
| | | phosphate ion homeostasis GO:0055062 | 1.85E-01 | 2 |
| | | cellular anion homeostasis GO:0030002 | 1.85E-01 | 2 |
| | | adenosine receptor binding GO:0031685 | 4.50E-02 | 2 |
| | | microfilament motor activity GO:0000146 | 4.50E-02 | 3 |
| | | carbon-carbon lyase activity GO:0016830 | 4.50E-02 | 4 |
| | | malate dehydrogenase activity GO:0016615 | 5.31E-02 | 2 |
| | | calmodulin binding GO:0005516 | 6.37E-02 | 6 |
| | | carboxy-lyase activity GO:0016831 | 6.37E-02 | 3 |
| | | alkali metal ion binding GO:0031420 | 6.37E-02 | 2 |
| | | actin-dependent ATPase activity GO:0030898 | 6.37E-02 | 2 |
| | | ATPase activity, coupled GO:0042623 | 6.37E-02 | 8 |
| | | heme transporter activity GO:0015232 | 6.37E-02 | 2 |
| | | cell body GO:0044297 | 3.14E-02 | 10 |
| | | neuronal cell body GO:0043025 | 3.14E-02 | 10 |
| | | perikaryon GO:0043204 | 3.66E-02 | 4 |
| | | neuron projection GO:0043005 | 4.71E-02 | 14 |
| | | cell projection part GO:0044463 | 4.71E-02 | 14 |
| | | presynaptic membrane GO:0042734 | 4.71E-02 | 4 |
| | | axolemma GO:0030673 | 7.40E-02 | 2 |
| | | anchored to membrane GO:0031225 | 1.08E-01 | 5 |
| | | terminal button GO:0043195 | 1.08E-01 | 3 |
| | | extracellular region part GO:0044421 | 1.08E-01 | 18 |
| | mouse | cellular process GO:0009987 | 3.10E-03 | 116 |
| | | cellular trivalent inorganic anion homeostasis GO:0072502 | 6.82E-02 | 2 |
| | | positive regulation of response to food GO:0032097 | 6.82E-02 | 2 |
| | | positive regulation of appetite GO:0032100 | 6.82E-02 | 2 |
| | | cellular ketone body metabolic process GO:0046950 | 6.82E-02 | 2 |
| | | cellular metabolic process GO:0044237 | 6.82E-02 | 70 |
| | | response to food GO:0032094 | 6.82E-02 | 3 |
| | | cellular phosphate ion homeostasis GO:0030643 | 6.82E-02 | 2 |
| | | fat cell differentiation GO:0045444 | 7.88E-02 | 5 |
| | | regulation of systemic arterial blood pressure by baroreceptor feedback GO:0003025 | 7.88E-02 | 2 |
| | | binding GO:0005488 | 8.00E-04 | 99 |
| | | ion binding GO:0043167 | 8.00E-04 | 58 |
| | | catalytic activity GO:0003824 | 8.00E-04 | 57 |

| | | | | |
|---|---|---|---|---|
| | | anion binding GO:0043168 | 1.20E-03 | 32 |
| | | small molecule binding GO:0036094 | 3.70E-03 | 32 |
| | | nucleotide binding GO:0000166 | 9.20E-03 | 29 |
| | | ATP binding GO:0005524 | 9.20E-03 | 21 |
| | | nucleoside phosphate binding GO:1901265 | 9.20E-03 | 29 |
| | | adenyl nucleotide binding GO:0030554 | 1.00E-02 | 21 |
| | | adenyl ribonucleotide binding GO:0032559 | 1.00E-02 | 21 |
| | | neuron spine GO:0044309 | 6.10E-03 | 7 |
| | | cell part GO:0044464 | 6.10E-03 | 114 |
| | | clathrin-coated endocytic vesicle GO:0045334 | 6.10E-03 | 3 |
| | | cell GO:0005623 | 6.10E-03 | 114 |
| | | dendritic spine GO:0043197 | 6.10E-03 | 7 |
| | | AP-2 adaptor complex GO:0030122 | 1.09E-02 | 2 |
| | | neuronal cell body GO:0043025 | 1.09E-02 | 10 |
| | | clathrin-coated endocytic vesicle membrane GO:0030669 | 1.30E-02 | 2 |
| | | clathrin coat of endocytic vesicle GO:0030128 | 1.30E-02 | 2 |
| | | cell projection part GO:0044463 | 1.30E-02 | 12 |

## Table S7.5: Pathway enrichment for genes showing 1% top $F_{ST}$ values.

| Database | Panel | Model Organism | Pathway | Gene Name | Adjusted p-value |
|---|---|---|---|---|---|
| KEGG | DOM | human | Chemokine signaling pathway | *CCL13, CCL7, GRK4, PIK3CD, RAP1A, ADCY5, CCL1* | 6.00E-04 |
| | | | Vasopressin-regulated water reabsorption | *AQP2, NSF, DYNC1LI2, AQP4* | 6.00E-04 |
| | | | Valine, leucine and isoleucine degradation | *OXCT1, BCKDHB, HADHA, ALDH9A1* | 6.00E-04 |
| | | | Riboflavin metabolism | *ENPP1, ACP6* | 1.32E-02 |
| | | | Amyotrophic lateral sclerosis (ALS) | *APAF1, GPX1, PRPH* | 1.43E-02 |
| | | | Glutathione metabolism | *RRM2B, GPX1, MGST3* | 1.43E-02 |
| | | | Adipocytokine signaling pathway | *PRKAG1, MAPK10, NPY* | 2.45E-02 |
| | | | Phagosome | *TUBA1C, COLEC12, TUBA1A, DYNC1LI2* | 2.57E-02 |
| | | | beta-Alanine metabolism | *HADHA, ALDH9A1* | 2.57E-02 |
| | | | Purine metabolism | *ENPP1, RRM2B, ADCY5, PDE4B* | 2.71E-02 |
| | | mouse | Chemokine signaling pathway | *Pik3cd, Adcy5, Rap1a, Grk4, Ccl7, Ccl11, Ccl1* | 9.02E-05 |
| | | | Valine, leucine and isoleucine degradation | *Oxct1, Aldh9a1, Hadha, Bckdhb* | 3.00E-04 |
| | | | Vasopressin-regulated water reabsorption | *Aqp2, Dync1li2, Aqp4, Nsf* | 3.00E-04 |
| | | | Amyotrophic lateral sclerosis (ALS) | *Prph, Apaf1, Gpx1* | 6.40E-03 |
| | | | Glutathione metabolism | *Gpx1, Mgst3, Rrm2b* | 6.40E-03 |
| | | | Riboflavin metabolism | *Acp6, Enpp1* | 6.40E-03 |
| | | | Metabolic pathways | *Dgat1, Pigb, Inpp4b, B4galt7, Aldh9a1, Hadha, Bckdhb, Uck2, Enpp1, Smpd1, Rrm2b* | 7.30E-03 |
| | | | Adipocytokine signaling pathway | *Mapk10, Prkag1, Npy* | 8.80E-03 |
| | | | Phagosome | *Dync1li2, Tuba1a, Colec12, Tuba1c* | 1.13E-02 |
| | | | Gap junction | *Adcy5, Tuba1a, Tuba1c* | 1.13E-02 |
| | FM | human | Neuroactive ligand-receptor interaction | *CTSG, HTR1E, OPRK1, CGA, GRIK2, ADRA1D, GHSR, GPR35* | 5.00E-04 |
| | | | Glutathione metabolism | *RRM2B, ODC1, GPX1* | 1.80E-02 |
| | | | Metabolic pathways | *GAD2, ENPP1, RRM2B, BCKDHB, MDH2, SMPD1, UCK2, PIGB, ME1, RPN2, NDUFA7, ODC1* | 2.28E-02 |
| | | | Gastric acid secretion | *SST, ADCY5, MYLK2* | 2.88E-02 |
| | | | Butanoate metabolism | *GAD2, OXCT1* | 4.37E-02 |
| | | | GnRH signaling pathway | *MAPK10, CGA, ADCY5* | 4.37E-02 |
| | | | Primary immunodeficiency | *ICOS, CIITA* | 4.37E-02 |
| | | | Vascular smooth muscle contraction | *ADRA1D, ADCY5, MYLK2* | 4.48E-02 |
| | | | Pyruvate metabolism | *ME1, MDH2* | 4.48E-02 |
| | | | Valine, leucine and isoleucine degradation | *OXCT1, BCKDHB* | 4.79E-02 |
| | | mouse | Neuroactive ligand-receptor interaction | *Ctsg, Grik2, Cga, Oprk1, Adra1d, Gpr35, Ghsr* | 6.00E-04 |
| | | | Metabolic pathways | *Rpn2, Mdh2, Odc1, Gad2, Ndufa7, Pigb, Bckdhb, Uck2, Enpp1, Smpd1, Rrm2b, Me1* | 3.60E-03 |
| | | | Glutathione metabolism | *Odc1, Gpx1, Rrm2b* | 6.00E-03 |
| | | | Gastric acid secretion | *Adcy5, Mylk2, Sst* | 1.08E-02 |
| | | | Butanoate metabolism | *Oxct1, Gad2* | 2.04E-02 |
| | | | GnRH signaling pathway | *Adcy5, Mapk10, Cga* | 2.04E-02 |
| | | | Primary immunodeficiency | *Ciita, Icos* | 2.21E-02 |
| | | | Vascular smooth muscle contraction | *Adcy5, Mylk2, Adra1d* | 2.40E-02 |
| | | | Toxoplasmosis | *Lama3, Ciita, Mapk10* | 2.40E-02 |
| | | | Pyruvate metabolism | *Mdh2, Me1* | 2.45E-02 |
| WikiPathway | DOM | human | Oxidative Stress | *MAPK10, GPX1, SOD2* | 8.60E-03 |
| | | | Integrated Pancreatic Cancer Pathway | *APAF1, RAP1A, ESR2, TOP1, PDE4B* | 1.72E-02 |
| | | | Insulin Signaling | *ENPP1, PIK3CD, MAPK10, STXBP1* | 3.75E-02 |
| | | | Wnt Signaling Pathway and Pluripotency | *RACGAP1, MAPK10, PPM1J* | 3.75E-02 |
| | | | Apoptosis Modulation by HSP70 | *APAF1, MAPK10* | 3.75E-02 |
| | | | G13 Signaling Pathway | *PIK3CD, MAPK10* | 3.75E-02 |
| | | | Matrix Metalloproteinases | *TIMP2, MMP17* | 3.75E-02 |
| | | | Ovarian Infertility Genes | *ESR2, SMPD1* | 3.75E-02 |
| | | | Hedgehog Signaling Pathway | *DHH, IHH* | 3.75E-02 |

| | | | | | |
|---|---|---|---|---|---|
| | | | Integrin-mediated cell adhesion | *RAP1A, MAPK10, ITGA4* | 3.75E-02 |
| | | mouse | Chemokine signaling pathway | *Pik3cd, Adcy5, Rap1a, Grk4, Ccl7, Ccl11, Ccl1* | 4.96E-05 |
| | | | Oxidative Stress | *Sod2, Mapk10, Gpx1* | 1.30E-03 |
| | | | Insulin Signaling | *Pik3cd, Mapk10, Stxbp1, Enpp1* | 1.11E-02 |
| | | | Oxidative Damage | *Mapk10, Apaf1* | 1.11E-02 |
| | | | Apoptosis Modulation by HSP70 | *Mapk10, Apaf1* | 1.40E-02 |
| | | | Matrix Metalloproteinases | *Timp2, Mmp17* | 1.40E-02 |
| | | | Integrin-mediated cell adhesion | *Rap1a, Mapk10, Itga4* | 1.40E-02 |
| | | | Wnt Signaling Pathway and Pluripotency | *Racgap1, Mapk10, Ppm1j* | 1.40E-02 |
| | | | Ovarian Infertility Genes | *Esr2, Smpd1* | 1.70E-02 |
| | | | G13 Signaling Pathway | *Pik3cd, Mapk10* | 1.90E-02 |
| | FM | human | GPCRs, Class A Rhodopsin-like | *HTR1E, OPRK1, OR10J1, OR2B6, ADRA1D, GHSR, GPR35* | 1.80E-03 |
| | | | Lymphocyte TarBase | *MSI2, HDAC4, MYO10, BCKDHB, TTK, FNDC3B, NCEH1, DHX40, RHOB* | 3.20E-03 |
| | | | miRNA regulation of DNA Damage Response | *RFC1, RRM2B, SESN1* | 1.12E-02 |
| | | | Endochondral Ossification | *ENPP1, HDAC4, IHH* | 1.12E-02 |
| | | | DNA damage response | *RFC1, RRM2B, SESN1* | 1.12E-02 |
| | | | Epithelium TarBase | *FNDC3B, HDAC4, MYO10, BCKDHB, DHX40, RHOB* | 1.12E-02 |
| | | | Monoamine GPCRs | *HTR1E, ADRA1D* | 1.77E-02 |
| | | | Muscle cell TarBase | *FNDC3B, HDAC4, NCEH1, MYO10, DHX40, TTK* | 1.77E-02 |
| | | | Oxidative Stress | *MAPK10, GPX1* | 1.77E-02 |
| | | | mRNA processing | *DDX20, RNGTT, SRPK1* | 3.11E-02 |
| | | mouse | miRNA regulation of DNA Damage Response | *Rfc1, Rrm2b, Sesn1* | 1.15E-02 |
| | | | Endochondral Ossification | *Hdac4, Ihh, Enpp1* | 1.15E-02 |
| | | | mRNA processing | *Ddx20, Rngtt, Msi2, Rps28, Srpk1, Zfml* | 1.52E-02 |
| | | | Oxidative Stress | *Mapk10, Gpx1* | 1.52E-02 |
| | | | GPCRs, Class A Rhodopsin-like | *Olfr716, Oprk1, Adra1d, Gpr35* | 1.76E-02 |
| | | | GPCRs, Other | *Oprk1, Olfr11, Ghsr* | 2.99E-02 |
| | | | Selenium metabolism-Selenoproteins | *Sepsecs, Gpx1* | 2.99E-02 |
| | | | One carbon metabolism and related pathways | *Gad2, Gpx1* | 2.99E-02 |
| | | | Peptide GPCRs | *Oprk1, Ghsr* | 3.65E-02 |
| | | | Focal Adhesion | *Lama3, Mylk2, Rhob* | 3.65E-02 |

## Table S7.6: Phenotype-term enrichment for genes showing 1% top $F_{ST}$ values.

| Database | Panel | Model Organism | Phenotype | Gene Name | Adjusted p-value |
|---|---|---|---|---|---|
| Phenotype | DOM | human | Aplasia/Hypoplasia of the fallopian tube | *DHH, DCAF17,* | 4.33E-02 |
| | | | Hypoplasia of the fallopian tube | *DHH, DCAF17,* | 4.33E-02 |
| | | | Abnormality of the fallopian tube | *DHH, DCAF17,* | 4.33E-02 |
| | | | Spastic tetraplegia | *SEPSECS, TUBA1A, ELOVL4, STXBP1,* | 4.33E-02 |
| | | | Thin corpus callosum | *SEPSECS, TUBA1A, STXBP1, MAPK10,* | 5.77E-02 |
| | | | Abnormal thickness of corpus callosum | *SEPSECS, TUBA1A, STXBP1, MAPK10,* | 5.77E-02 |
| | | | Protracted diarrhea | *MYO5B, CIITA,* | 1.18E-01 |
| | | | Generalized myoclonic seizures | *ELOVL4, STXBP1, MAPK10,* | 1.89E-01 |
| | | | Autoimmune thrombocytopenia | *MLL2, NDRG1,* | 2.17E-01 |
| | | | Hypoplasia of the uterus | *DHH, DCAF17,* | 2.24E-01 |
| | | mouse | nervous system phenotype | *Cntn5, Stxbp1, Tuba1a, Tmbim6, Ubr5, Gpx1, Smpd1, Slc24a4, Mapk10, Slc6a15, Apaf1, Dhh, Faim2, Accn1, Sod2, Lrpap1, Ihh, Timp2, Esr2, Prph, Grid2, Ndrg1, Pask, Atoh1, Rrm2b, Oxct1, Mterfd2, Trim3, Tg, Elovl4, Enpp1, Sh2d3c, Hsf1, Aqp4, Npy,* | 4.90E-02 |
| | | | abnormal trigeminal motor nucleus morphology | *Grid2, Sod2,* | 4.90E-02 |
| | | | abnormal nervous system physiology | *Accn1, Sod2, Cntn5, Stxbp1, Tuba1a, Timp2, Esr2, Grid2, Tmbim6, Ndrg1, Pask, Gpx1, Atoh1, Smpd1, Oxct1, Slc24a4, Mapk10, Trim3, Apaf1, Tg, Faim2, Elovl4, Sh2d3c, Aqp4, Npy,* | 4.90E-02 |
| | | | failure of zygotic cell division | *Racgap1, Ddx20, Hsf1,* | 4.90E-02 |
| | | | abnormal brainstem morphology | *Accn1, Grid2, Sod2, Atoh1, Apaf1, Smpd1, Esr2,* | 4.90E-02 |
| | | | abnormal erythrocyte morphology | *Aqp2, Sod2, Ikzf1, Tg, Esr2, Ctsg, Prkag1, Elovl4, Rrm2b, Tmpo, Ulk1,* | 5.10E-02 |
| | | | abnormal hematocrit | *Aqp2, Sod2, Ikzf1, Prkag1, Tg, Rrm2b, Tmpo,* | 5.10E-02 |
| | | | decreased susceptibility to pharmacologically induced seizures | *Mapk10, Trim3, Npy,* | 5.10E-02 |
| | | | increased circulating pituitary hormone level | *Grid2, Lhx9, Tg, Dhh, Esr2,* | 5.44E-02 |
| | | | abnormal erythropoiesis | *Aqp2, Sod2, Ikzf1, Tg, Esr2, Ctsg, Prkag1, Elovl4, Rrm2b, Tmpo, Ulk1,* | 6.44E-02 |
| | FM | human | Subvalvular aortic stenosis | *HDAC4, MYLK2,* | 4.99E-01 |
| | | | Protracted diarrhea | *MYO5B, CIITA,* | 4.99E-01 |
| | | | Recurrent lower respiratory tract infections | *ICOS, WDR19, CIITA,* | 4.99E-01 |
| | | | Anxiety | *TMCO1, RRM2B, ADCY5,* | 4.99E-01 |
| | | | Abnormality of the left ventricular outflow tract | *HDAC4, MYLK2,* | 4.99E-01 |
| | | | Reduced tendon reflexes | *DYSF, TMCO1, HDAC4, RRM2B, FLVCR1, BCKDHB, SMPD1,* | 4.99E-01 |
| | | | Abnormality of the urinary system | *TMCO1, LAMA3, HDAC4, OXCT1, RRM2B, BCKDHB, WDR19, TNFRSF11B, COL14A1, FLVCR1, AP2S1, CIITA,* | 5.83E-01 |
| | | | Gingival overgrowth | *TMCO1, MAPK10,* | 5.83E-01 |
| | | | Abnormality of immune system physiology | *LAMA3, ELOVL4, WDR19, SMPD1, FLVCR1, ICOS, MAPK10, CIITA,* | 5.83E-01 |
| | | | Abnormal respiratory system morphology | *TMCO1, LAMA3, HDAC4, ELOVL4, BCKDHB, WDR19, SMPD1, MAPK10, ICOS, CIITA, SLC34A2,* | 5.83E-01 |
| | | mouse | decreased susceptibility to pharmacologically induced seizures | *Mapk10, Grik2, Trim3, Npy,* | 5.50E-03 |
| | | | abnormal seizure response to inducing agent | *Hdac4, Mapk10, Grik2, Cntnap2, Cntn5, Trim3, Npy,* | 2.86E-02 |
| | | | behavior/neurological phenotype | *Foxs1, Trib2, Cntnap2, Adra1d, Gad2, Lmo7, Cntn5, Cntn6, Apbb1, Ghsr, Adcy5, Lhfpl5, Oprk1, Clps, Smpd1, Me1, Mapk10, Trim3, Dysf, Prox1, Sst, Hdac4, Add2, Grik2, Elovl4, Anxa4, Enpp1, Npy,* | 5.93E-02 |
| | | | abnormal behavior | *Foxs1, Trib2, Cntnap2, Adra1d, Gad2, Lmo7, Cntn5, Cntn6, Apbb1, Ghsr, Adcy5, Lhfpl5, Oprk1, Clps, Smpd1, Me1, Mapk10, Trim3, Dysf, Prox1, Sst, Hdac4, Add2, Grik2, Elovl4, Anxa4, Enpp1, Npy,* | 5.93E-02 |
| | | | decreased body size | *Trib2, Cntnap2, Gad2, Lmo7, Ihh, Mfsd7b, Ghsr, Clps, Ghrh, Gpx1, Rrm2b, Smpd1, Me1, Lama3, Cga, Rhob, Hdac4, Msi2, Grik2, Elovl4, Tnfrsf11b, Enpp1, Npy,* | 5.93E-02 |
| | | | pancreas inflammation | *Ctla4, Ciita, Gad2, Icos,* | 6.59E-02 |
| | | | seizures | *Hdac4, Mapk10, Grik2, Cntnap2, Gad2, Cntn5, Trim3, Npy,* | 6.59E-02 |
| | | | abnormal body size | *Foxs1, Trib2, Cntnap2, Gad2, Lmo7, Ihh, Mfsd7b, Ghsr, Clps, Ghrh, Gpx1, Rrm2b, Smpd1, Me1, Lama3, Cga, Rhob, Prox1, Sst, Hdac4, Msi2, Grik2, Elovl4, Tnfrsf11b, Enpp1, Npy,* | 6.59E-02 |
| | | | environmentally induced seizures | *Hdac4, Cntnap2, Cntn5,* | 6.59E-02 |
| | | | decreased body weight | *Trib2, Gad2, Lmo7, Ghsr, Clps, Gpx1, Ghrh, Rrm2b, Smpd1, Me1, Lama3, Cga, Hdac4, Msi2, Grik2, Elovl4, Enpp1, Npy* | 8.21E-02 |
| PheWAS | DOM | human | Urinary incontinence PheWAS:599.4 | *DTNB, PLXDC2,* | 2.78E-01 |
| | | | Back pain PheWAS:760 | *TMCO1, PALM2-AKAP2,* | 2.78E-01 |
| | | | Abnormal coagulation profile PheWAS:286.9 | *IKZF1, CLEC16A,* | 2.78E-01 |
| | | | Appendicitis PheWAS:540.1 | *TMCO1, MGST3,* | 2.78E-01 |

| | | | | | |
|---|---|---|---|---|---|
| | | | Other nonspecific findings on examination of urine PheWAS:598.9 | *TIMP2, NSF, PALM2-AKAP2, PLXDC2,* | 2.78E-01 |
| | | | Other benign neoplasm of connective and other soft tissue PheWAS:215 | *IKZF1, ADCY5,* | 2.78E-01 |
| | | | Urinary tract infection PheWAS:591 | *DTNB, CIITA,* | 2.78E-01 |
| | | | Chronic glomerulonephritis PheWAS:580.14 | *NSF, PALM2-AKAP2,* | 2.78E-01 |
| | | | Generalized anxiety disorder PheWAS:300.11 | *MYO1D, SEL1L3, PRPH,* | 2.78E-01 |
| | | | Dermatophytosis of the body PheWAS:110.13 | *TMCO1, CNTN5, SLC24A4,* | 2.78E-01 |
| | FM | | Other disorders of thyroid PheWAS:246 | *PROX1, GPC5, CLEC16A,* | 3.00E-01 |
| | | | Cholangitis PheWAS:575.1 | *SNTB1, ICOS,* | 3.00E-01 |
| | | | Other infectious diseases PheWAS:136 | *CNTN4, GRIK2,* | 3.00E-01 |
| | | | Functional disorders of bladder PheWAS:596.5 | *CTLA4, ADCY5,* | 3.00E-01 |
| | | | Nervous system congenital anomalies PheWAS:752 | *CNTN4, GHSR,* | 3.00E-01 |
| | | | Psoriatic arthropathy PheWAS:696.42 | *SEL1L3, CIITA,* | 3.00E-01 |
| | | | Other abnormal blood chemistry PheWAS:790.6 | *PROX1, CNTN4,* | 3.00E-01 |
| | | | Oliguria and anuria PheWAS:599.6 | *DYSF, SH3BP5,* | 3.00E-01 |
| | | | Pulmonary embolism and infarction PheWAS:415.1 | *PROX1, FAM13A,* | 3.00E-01 |
| | | | Stricture and stenosis of esophagus PheWAS:530.3 | *CLEC16A, MICAL3,* | 3.00E-01 |

## Table S7.7: Disease-term enrichment for genes showing 1% top $F_{ST}$ values (human model).

| Panel | Disease | Gene Name | Adjusted p-value |
|---|---|---|---|
| DOM | Disease Susceptibility DB_ID:PA443919 | *ENPP1, BANK1, CNTN5, FAIM2, TIMP2, NOD1, ADCY5, CLEC16A, NPY, SOD2, TG, ESR2, GPX1, CIITA, CCL1* | 4.00E-04 |
| | Genetic Predisposition to Disease DB_ID:PA446882 | *ENPP1, BANK1, FAIM2, TIMP2, NOD1, ADCY5, CLEC16A, NPY, PDE4B, SOD2, ESR2, GPX1, CIITA, CCL1* | 6.00E-04 |
| | Metabolic Diseases DB_ID:PA444938 | *ENPP1, OXCT1, BCKDHB, ADCY5, CLEC16A, HADHA, SMPD1, NPY, DGAT1, DCAF17, PRPH, SETX* | 6.00E-04 |
| | Immune System Diseases DB_ID:PA444602 | *CCL7, BANK1, NOD1, ITGA4, CLEC16A, AQP4, WIPF1, TG, SEPSECS, PIK3CD, IKZF1, CIITA, CCL1* | 6.00E-04 |
| | diabetes mellitus type 2 and obesity DB_ID:PA447306 | *ENPP1, FAIM2, RASAL2, ADCY5, NPY, BCDIN3D* | 1.20E-03 |
| | Stress DB_ID:PA445752 | *RRM2B, NDRG1, NPY, SOD2, HSF1, PRKAG1, TMBIM6, MAPK10, GPX1, STK4* | 1.20E-03 |
| | Meniere Disease DB_ID:PA444911 | *AQP6, AQP2, AQP4* | 1.30E-03 |
| | Immunologic Deficiency Syndromes DB_ID:PA444601 | *CCL7, WIPF1, TUBA1C, PIK3CD, PDCD6IP, TUBA1A, MAPK10, ITGA4, CIITA* | 1.50E-03 |
| | HIV DB_ID:PA447230 | *CCL7, PDCD6IP, TUBA1A, ADCY5, ITGA4, PRKAG1, TUBA1C, PIK3CD, MOV10, MAPK10, CCL1, TOP1* | 2.10E-03 |
| | Diabetes Mellitus DB_ID:PA443886 | *ENPP1, ADCY5, CLEC16A, NPY, SOD2, DCAF17, GPX1, BCDIN3D* | 2.20E-03 |
| FM | Disease Susceptibility DB_ID:PA443919 | *CTLA4, CNTN5, CNTNAP2, FAM13A, ICOS, CIITA, ENPP1, GAD2, CNTN4, CLEC16A, ADCY5, NPY, RFC1, GPC5, AFF3, GPX1, CCL1* | 1.14E-05 |
| | Genetic Predisposition to Disease DB_ID:PA446882 | *GAD2, ENPP1, CTLA4, ADCY5, CLEC16A, NPY, RFC1, CNTNAP2, GPC5, FAM13A, AFF3, ICOS, GPX1, CIITA, CCL1* | 1.00E-04 |
| | Diabetes Mellitus DB_ID:PA443886 | *GAD2, ENPP1, CTLA4, ADCY5, CLEC16A, NPY, PROX1, TNFRSF11B, DCAF17, GPX1* | 1.00E-04 |
| | Anxiety Disorders DB_ID:PA447196 | *CNTN6, GAD2, CNTN5, CNTNAP2, OPRK1, GRIK2, NPY* | 2.00E-04 |
| | Mental Disorders DB_ID:PA447208 | *GAD2, CNTN5, OPRK1, ASIC2, CNTN4, GRIK2, APBB1, NPY, SST, CNTNAP2, DCAF17* | 1.00E-03 |
| | Immune System Diseases DB_ID:PA444602 | *GAD2, CTLA4, CLEC16A, TNFSF10, TNFRSF11B, SEPSECS, RFC1, GPC5, AFF3, ICOS, CIITA, CCL1* | 1.00E-03 |
| | Endocrine disorder NOS DB_ID:PA165109147 | *GAD2, ENPP1, CTLA4, ADCY5, CLEC16A, TNFRSF11B, DCAF17, GHRH, SLC34A2* | 1.60E-03 |
| | Autoimmune Diseases DB_ID:PA443464 | *GAD2, CTLA4, CLEC16A, TNFRSF11B, SEPSECS, GPC5, AFF3, ICOS, CIITA* | 1.60E-03 |
| | Endocrine System Diseases DB_ID:PA444037 | *GAD2, ENPP1, CTLA4, ADCY5, CLEC16A, TNFRSF11B, DCAF17, GHRH, SLC34A2* | 1.60E-03 |
| | Endocrine disturbance NOS DB_ID:PA165108435 | *GAD2, ENPP1, CTLA4, ADCY5, CLEC16A, TNFRSF11B, DCAF17, GHRH, SLC34A2* | 1.60E-03 |

# Table S7.8: Genes harboring top-1% $d_A$ sites.

| Panel | Site category | Genes |
|---|---|---|
| DOM | 0d-fold | ENSECAG00000021537 (FNDC7) |
| | 2d-fold | ENSECAG00000017796 (NCAPD2) |
| | 0-1Kb | ENSECAG00000000628, ENSECAG00000002575, ENSECAG00000002739, ENSECAG00000003284, ENSECAG00000003576, ENSECAG00000003780 (OR6A2), ENSECAG00000008324 (LECT2), ENSECAG00000008749 (VTCN1), ENSECAG00000011862 (C11orf1), ENSECAG00000015087 (PFKM), ENSECAG00000015444 (SLC37A1), ENSECAG00000017048 (CD101), ENSECAG00000024694 (KIF14) |
| | 1-2Kb | ENSECAG00000008247, ENSECAG00000008324 (LECT2), ENSECAG00000010259 (TRIM45), ENSECAG00000011995 (FTSJ3), ENSECAG00000013724 (FAS), ENSECAG00000014017 (PRKG1), ENSECAG00000015087 (PFKM), ENSECAG00000016463 (TMEM72), ENSECAG00000020193 (TGFBI), ENSECAG00000022355, ENSECAG00000022357 (EIF4E1B), ENSECAG00000023727 (IL1A), ENSECAG00000023803 (TM9SF2), ENSECAG00000024187 (SMARCA2), ENSECAG00000024824 (LPHN2) |
| | 2-3Kb | ENSECAG00000002070, ENSECAG00000002739, ENSECAG00000003576, ENSECAG00000004219 (ZNF831), ENSECAG00000004271 (OR10A4), ENSECAG00000005669, ENSECAG00000009983, ENSECAG00000010259 (TRIM45), ENSECAG00000012980 (TRNAU1AP), ENSECAG00000018025 (SIGLEC15), ENSECAG00000019408 (CAT), ENSECAG00000020193 (TGFBI), ENSECAG00000022313 (COL2A1) |
| | 3-4Kb | ENSECAG00000005530 (ACAD9), ENSECAG00000005652, ENSECAG00000005669, ENSECAG00000008046 (EIF4GI), ENSECAG00000008324 (LECT2), ENSECAG00000008754 (NAA25), ENSECAG00000011330, ENSECAG00000013741 (MORC1), ENSECAG00000016069 (CLYBL), ENSECAG00000016256 (HPCA), ENSECAG00000018825 (SIK2), ENSECAG00000019542 (SLC35D1), ENSECAG00000019681 (MAN1A2), ENSECAG00000022357 (EIF4E1B) |
| | 4-5Kb | ENSECAG00000000598 (C19orf18), ENSECAG00000000765, ENSECAG00000001160, ENSECAG00000002084, ENSECAG00000004271 (OR10A4), ENSECAG00000005669, ENSECAG00000010259 (TRIM45), ENSECAG00000012474 (TTF2), ENSECAG00000013724 (FAS), ENSECAG00000016463 (TMEM72), ENSECAG00000017078 (CYLC2), ENSECAG00000020193 (TGFBI), ENSECAG00000021563 (CLEC2D), ENSECAG00000022497 (AFM), ENSECAG00000023832 (GAK) |
| | 5-6Kb | ENSECAG00000002070, ENSECAG00000005585, ENSECAG00000005669, ENSECAG00000008494, ENSECAG00000011154 (CHRD), ENSECAG00000011444 (THOC5), ENSECAG00000013938 (ZDHHC13), ENSECAG00000016463 (TMEM72), ENSECAG00000017707 (HCK), ENSECAG00000018025 (SIGLEC15), ENSECAG00000020193 (TGFBI), ENSECAG00000021490, ENSECAG00000022438 (CTSS), ENSECAG00000024981 (HFE), ENSECAG00000026829 |
| | 6-7Kb | ENSECAG00000002070, ENSECAG00000003148, ENSECAG00000003256 (KRTAP8-1), ENSECAG00000004271 (OR10A4), ENSECAG00000005669, ENSECAG00000007401, ENSECAG00000008247, ENSECAG00000012584 (GNG2), ENSECAG00000012715, ENSECAG00000015637 (NREP), ENSECAG00000016463 (TMEM72), ENSECAG00000019471 (RASSF6), ENSECAG00000021706 (RPS6KC1), ENSECAG00000022588, ENSECAG00000023832 (GAK) |
| | 7-8Kb | ENSECAG00000002084, ENSECAG00000003256 (KRTAP8-1), ENSECAG00000008247, ENSECAG00000009430 (C16orf72), ENSECAG00000009983, ENSECAG00000015200, ENSECAG00000015637 (NREP), ENSECAG00000016463 (TMEM72), ENSECAG00000017260 (DPEP2), ENSECAG00000017796 (NCAPD2), ENSECAG00000018725 (OGFOD3), ENSECAG00000019542 (SLC35D1) |
| | 8-9Kb | ENSECAG00000000765, ENSECAG00000002084, ENSECAG00000005669, ENSECAG00000008980 (FBXL21), ENSECAG00000009066 (GABBR2), ENSECAG00000009983, ENSECAG00000010461 (LSM1), ENSECAG00000013392 (TMC5), ENSECAG00000015255 (ARMC4), ENSECAG00000016463 (TMEM72), ENSECAG00000018662 (CCDC66), ENSECAG00000018825 (SIK2), ENSECAG00000019471 (RASSF6), ENSECAG00000019634 (DMC1) |
| | 9-10Kb | ENSECAG00000003383, ENSECAG00000005132 (OR2D3), ENSECAG00000008940 (TMEM40), ENSECAG00000009430 (C16orf72), ENSECAG00000011222 (NGEF), ENSECAG00000012507 (CRYAB), ENSECAG00000014389 (MTRF1), ENSECAG00000015637 (NREP), ENSECAG00000015821 (ZER1), ENSECAG00000016463 (TMEM72), ENSECAG00000018401 (CNDP2), ENSECAG00000019137 (MAML3), ENSECAG00000019408 (CAT), ENSECAG00000022355, ENSECAG00000023616 (DDX4) |
| FM | 0d-fold | ENSECAG00000000557 (RPUSD4), ENSECAG00000000955, ENSECAG00000002300 (OR51S1), ENSECAG00000003064, ENSECAG00000003576, ENSECAG00000003780 (OR6A2), ENSECAG00000004548, ENSECAG00000005898, ENSECAG00000006970 (DSG4), ENSECAG00000008040 (ADGRB3), ENSECAG00000008864 (CD1A1), ENSECAG00000009003 (APC), ENSECAG00000009483 (PRLR), ENSECAG00000009561 (FDXACB1), ENSECAG00000010259 (TRIM45), ENSECAG00000011688 (TRIOBP), ENSECAG00000011862 (C11orf1), ENSECAG00000012312 (ARHGAP15), ENSECAG00000012653 (DYRK4), ENSECAG00000012702 (PPBP), ENSECAG00000013471 (ZNF599), ENSECAG00000013505, ENSECAG00000014894 (FASTKD3), ENSECAG00000015622 (CDHR4), ENSECAG00000016457 (E2F8), ENSECAG00000017249 (TRPS1), ENSECAG00000017257, ENSECAG00000017335 (CDON), ENSECAG00000017400 (SMURF1), ENSECAG00000019282 (HDC), ENSECAG00000019670 (UBA7), ENSECAG00000019940 (AKAP6), ENSECAG00000019993 (KALRN), ENSECAG00000020652 (ECHDC1), ENSECAG00000020841 (MRPS25), ENSECAG00000021419, ENSECAG00000021537 (FNDC7), ENSECAG00000021647, ENSECAG00000022699 (PKHD1), ENSECAG00000023231, ENSECAG00000024047 (IL6R) |
| | 2d-fold | ENSECAG00000000701 (FN1), ENSECAG00000003240 (CCR9), ENSECAG00000005293, ENSECAG00000005669, ENSECAG00000008864 (CD1A1), ENSECAG00000008931 (C1orf112), ENSECAG00000009483 (PRLR), ENSECAG00000010684 (PHIP), ENSECAG00000011081, ENSECAG00000011478 (NLN), ENSECAG00000012060 (MYO3A), ENSECAG00000012881 (NME6), ENSECAG00000013742 (MAEA), ENSECAG00000014187 (RSPH3), ENSECAG00000014289 (ARHGAP12), ENSECAG00000014404 (WDR62), ENSECAG00000014894 (FASTKD3), ENSECAG00000017249 (TRPS1), ENSECAG00000017707 (HCK), ENSECAG00000017796 (NCAPD2), ENSECAG00000019159 (NUP85), ENSECAG00000021706 (RPS6KC1), ENSECAG00000021797 (LRRC39), ENSECAG00000022011 (AKAP7), ENSECAG00000022924 (EEF2K), ENSECAG00000023344 (USP28), ENSECAG00000025133 (SYDE2) |
| | 0-1Kb | ENSECAG00000000159 (C16orf78), ENSECAG00000000628, ENSECAG00000000937, ENSECAG00000002155 (NUP93), ENSECAG00000002739, ENSECAG00000003284, ENSECAG00000003576, ENSECAG00000003780 (OR6A2), ENSECAG00000004562, ENSECAG00000005289 (SPTSSB), ENSECAG00000005585, ENSECAG00000005669, ENSECAG00000005914 (PRR29), ENSECAG00000006876 (UTP20), ENSECAG00000007954, ENSECAG00000008049 (EPOR), ENSECAG00000008289 (C1orf27), ENSECAG00000008324 (LECT2), ENSECAG00000008404, ENSECAG00000008749 (VTCN1), ENSECAG00000008980 (FBXL21), ENSECAG00000009482 (PPRC1), ENSECAG00000010221, ENSECAG00000010421 (SLC25A19), ENSECAG00000010723 (ANGEL2), ENSECAG00000011143 (FBXO2), ENSECAG00000011212 (TTC23L), ENSECAG00000011330, ENSECAG00000011689 (SLBP), ENSECAG00000011801 (FHIT), ENSECAG00000011862 (C11orf1), ENSECAG00000012189 (PDE8B), ENSECAG00000013335, ENSECAG00000013471 (ZNF599), ENSECAG00000013489 (GPATCH11), ENSECAG00000013623 (PIK3AP1), ENSECAG00000013682 (VSNL1), ENSECAG00000013712 (ECATH-3), ENSECAG00000014004 (SAMD7), ENSECAG00000014103 (RXFP2), ENSECAG00000014397 (C1orf146), ENSECAG00000014518, ENSECAG00000014680 (MLLT4), ENSECAG00000014825 (LRPAP1), ENSECAG00000015182 (OSBPL1A), ENSECAG00000015444 (SLC37A1), ENSECAG00000016216 (ITCH), ENSECAG00000016256 (HPCA), ENSECAG00000016950 (KCNJ8), ENSECAG00000017048 (CD101), ENSECAG00000017484 (C20orf194), ENSECAG00000017907 (RNFT2), ENSECAG00000017924 (KCNG1), ENSECAG00000018022, ENSECAG00000018025 (SIGLEC15), ENSECAG00000018133 (C10orf54), ENSECAG00000018356, ENSECAG00000018847, ENSECAG00000018864 (CPNE2), ENSECAG00000019757, ENSECAG00000019825 (TRDC), ENSECAG00000019842 (NAA15), ENSECAG00000020278 (RAP1GAP2), ENSECAG00000020594 (ELAVL3), ENSECAG00000020638 (NEB), ENSECAG00000020724 (TPR), ENSECAG00000021095, ENSECAG00000021647, ENSECAG00000022355, ENSECAG00000022430 (TOX), ENSECAG00000022924 (EEF2K), ENSECAG00000022986 (CNDP1), ENSECAG00000023176, ENSECAG00000023321 (ELF5), ENSECAG00000023475, ENSECAG00000023607 (DOCK5), ENSECAG00000024187 (SMARCA2) |
| | 1-2Kb | ENSECAG00000000243, ENSECAG00000000377 (BRIX1), ENSECAG00000000867, ENSECAG00000000937, ENSECAG00000000955, ENSECAG00000000987 (PSMB3), ENSECAG00000001022 (UCN2), ENSECAG00000002070, ENSECAG00000002127, ENSECAG00000002739, ENSECAG00000002922, ENSECAG00000003765, ENSECAG00000004023 (SSBP1), ENSECAG00000005006, ENSECAG00000005111 (RFX4), ENSECAG00000005530 (ACAD9), ENSECAG00000005585, ENSECAG00000005781 (SCPEP1), ENSECAG00000006451 (DDX28), ENSECAG00000006779 (USP12), ENSECAG00000006876 (UTP20), ENSECAG00000007032 (KLHL29), ENSECAG00000007192 (PTPRC), ENSECAG00000007458 (SPRR4), ENSECAG00000007806 (PTPN3), ENSECAG00000007841 (CUX2), ENSECAG00000008247, ENSECAG00000008289 (C1orf27), ENSECAG00000008324 (LECT2), ENSECAG00000008846 (LGALS2), ENSECAG00000009315 (DHTKD1), ENSECAG00000010075 (COG2), ENSECAG00000010259 (TRIM45), ENSECAG00000010422 (DNAJC16), ENSECAG00000011143 (FBXO2), ENSECAG00000011330, ENSECAG00000011794 (C6orf141), ENSECAG00000011995 (FTSJ3), ENSECAG00000012027, ENSECAG00000012312 (ARHGAP15), ENSECAG00000012326 (CD1D), ENSECAG00000012560, ENSECAG00000012628 (EPB41L4A), ENSECAG00000012881 (NME6), ENSECAG00000013122 (ZNF544), ENSECAG00000013623 (PIK3AP1), ENSECAG00000013682 (VSNL1), ENSECAG00000013724 (FAS), ENSECAG00000013881 (DHX35), ENSECAG00000014004 (SAMD7), ENSECAG00000014425 (PIGK), ENSECAG00000014560 (KPNB1), ENSECAG00000014737 (HEATR5B), ENSECAG00000014865 (ADGRA3), ENSECAG00000014918 (ZNF154), ENSECAG00000014926 (MRPL16), ENSECAG00000015043 (CLNS1A), ENSECAG00000015465 (NCOA5), ENSECAG00000015782, ENSECAG00000016069 (CLYBL), ENSECAG00000016546 (GALNT12), ENSECAG00000017074 (FSD2), ENSECAG00000017087 (ACSBG2), ENSECAG00000017400 (SMURF1), ENSECAG00000017513 (DTHD1), ENSECAG00000017789 (SIX2), ENSECAG00000017924 (KCNG1), ENSECAG00000018025 (SIGLEC15), |

| | | |
|---|---|---|
| | | ENSECAG00000018068 (MROH2B), ENSECAG00000018160 (TBX21), ENSECAG00000019245 (ACSF2), ENSECAG00000019757, ENSECAG00000019825 (TRDC), ENSECAG00000019927, ENSECAG00000020341, ENSECAG00000020724 (TPR), ENSECAG00000020955, ENSECAG00000021003 (ARMC7), ENSECAG00000021355 (STAM2), ENSECAG00000021447 (ASPH), ENSECAG00000021621 (MOB3B), ENSECAG00000022192 (ZNF77), ENSECAG00000022296 (SGSM3), ENSECAG00000022510 (CD1E2), ENSECAG00000022924 (EEF2K), ENSECAG00000022986 (CNDP1), ENSECAG00000023424 (GABRA4), ENSECAG00000023475, ENSECAG00000024187 (SMARCA2), ENSECAG00000024647 (MBLAC2), ENSECAG00000024700 (NOLC1), ENSECAG00000024824 (LPHN2), ENSECAG00000024875, ENSECAG00000025178 (PMP22) |
| | 2-3Kb | ENSECAG00000000324 (RASEF), ENSECAG00000000869 (RPL26L1), ENSECAG00000000937, ENSECAG00000001326 (FZD1), ENSECAG00000001775 (RAB38), ENSECAG00000001934, ENSECAG00000002070, ENSECAG00000002084, ENSECAG00000002491 (WTIP), ENSECAG00000002581, ENSECAG00000002739, ENSECAG00000003367 (RAD51C), ENSECAG00000003576, ENSECAG00000003765, ENSECAG00000003791 (NR2C2), ENSECAG00000003916, ENSECAG00000004023 (SSBP1), ENSECAG00000004024 (GGA3), ENSECAG00000004087, ENSECAG00000004219 (ZNF831), ENSECAG00000004242 (AQP4), ENSECAG00000004271 (OR10A4), ENSECAG00000004709 (C2), ENSECAG00000005530 (ACAD9), ENSECAG00000005585, ENSECAG00000005669, ENSECAG00000005998 (TAAR6), ENSECAG00000006170, ENSECAG00000006228, ENSECAG00000006451 (DDX28), ENSECAG00000007241 (PDYN), ENSECAG00000007790 (OR4P4), ENSECAG00000007841 (CUX2), ENSECAG00000008178 (CPNE5), ENSECAG00000008324 (LECT2), ENSECAG00000008846 (LGALS2), ENSECAG00000008980 (FBXL21), ENSECAG00000009159 (PLXND1), ENSECAG00000009665 (DAG1), ENSECAG00000009720 (RIMBP2), ENSECAG00000009759 (TNRC6B), ENSECAG00000009920 (POLR3F), ENSECAG00000009983, ENSECAG00000010259 (TRIM45), ENSECAG00000010684 (PHIP), ENSECAG00000010989 (SH3YL1), ENSECAG00000010990, ENSECAG00000011081, ENSECAG00000011435 (MSI2), ENSECAG00000011794 (C6orf141), ENSECAG00000012326 (CD1D), ENSECAG00000012957 (RHAG), ENSECAG00000013623 (PIK3AP1), ENSECAG00000013671 (EIF3B), ENSECAG00000014004 (SAMD7), ENSECAG00000014173 (ZRANB2), ENSECAG00000014224, ENSECAG00000014425 (PIGK), ENSECAG00000014706 (RAD51AP2), ENSECAG00000014825 (LRPAP1), ENSECAG00000014927 (AZGP1), ENSECAG00000015152 (SERPINA6), ENSECAG00000015444 (SLC37A1), ENSECAG00000016143 (ANKS6), ENSECAG00000016404 (SLC22A1), ENSECAG00000016989 (OTOL1), ENSECAG00000017081 (DZANK1), ENSECAG00000017087 (ACSBG2), ENSECAG00000017484 (C20orf194), ENSECAG00000018025 (SIGLEC15), ENSECAG00000018068 (MROH2B), ENSECAG00000019067 (OSBPL7), ENSECAG00000019408 (CAT), ENSECAG00000019617 (ATP8B4), ENSECAG00000019853 (AMIGO2), ENSECAG00000019882 (CALHM3), ENSECAG00000021431, ENSECAG00000021621 (MOB3B), ENSECAG00000021706 (RPS6KC1), ENSECAG00000021851 (DGKE), ENSECAG00000021875 (USP42), ENSECAG00000022290 (GCA), ENSECAG00000022313 (COL2A1), ENSECAG00000022510 (CD1E2), ENSECAG00000022815 (RNF123), ENSECAG00000022881, ENSECAG00000022943, ENSECAG00000022986 (CNDP1), ENSECAG00000023239 (SSR3), ENSECAG00000023441 (E2F3), ENSECAG00000023475, ENSECAG00000023641 (SNX8), ENSECAG00000023805 (ZSWIM6), ENSECAG00000023818 (UVRAG), ENSECAG00000023949 (CCL2), ENSECAG00000024176 (SIGLECL1), ENSECAG00000024301 (DLG4), ENSECAG00000024345 (SUCNR1), ENSECAG00000024875 |
| | 3-4Kb | ENSECAG00000000159 (C16orf78), ENSECAG00000000324 (RASEF), ENSECAG00000000377 (BRIX1), ENSECAG00000000879 (CAB39L), ENSECAG00000000937, ENSECAG00000001808, ENSECAG00000002021, ENSECAG00000002070, ENSECAG00000002408 (ZNF329), ENSECAG00000002494, ENSECAG00000002548 (RNF182), ENSECAG00000003099, ENSECAG00000004085 (FAM78B), ENSECAG00000004735, ENSECAG00000004808, ENSECAG00000005111 (RFX4), ENSECAG00000005153, ENSECAG00000005193, ENSECAG00000005530 (ACAD9), ENSECAG00000005585, ENSECAG00000005669, ENSECAG00000006451 (DDX28), ENSECAG00000006802 (KCNAB1), ENSECAG00000007740 (HESX1), ENSECAG00000008135 (TRPM3), ENSECAG00000008324 (LECT2), ENSECAG00000008404, ENSECAG00000008426 (PAPSS2), ENSECAG00000008494, ENSECAG00000008754 (NAA25), ENSECAG00000008843 (PODXL2), ENSECAG00000008846 (LGALS2), ENSECAG00000008864 (CD1A1), ENSECAG00000008980 (FBXL21), ENSECAG00000009103 (MLANA), ENSECAG00000009759 (TNRC6B), ENSECAG00000010102, ENSECAG00000010259 (TRIM45), ENSECAG00000010990, ENSECAG00000011212 (TTC23L), ENSECAG00000011330, ENSECAG00000011783 (PDS5A), ENSECAG00000012859 (SP2), ENSECAG00000012864 (ZNF530), ENSECAG00000012957 (RHAG), ENSECAG00000013057 (MREG), ENSECAG00000013270 (ZNF134), ENSECAG00000013638 (DCDC2C), ENSECAG00000013671 (EIF3B), ENSECAG00000013893 (SCARA5), ENSECAG00000014242 (MAD1L1), ENSECAG00000015331 (FARP1), ENSECAG00000016069 (CLYBL), ENSECAG00000016256 (HPCA), ENSECAG00000016682 (RIT2), ENSECAG00000016934 (IL31RA), ENSECAG00000016973 (FILIP1L), ENSECAG00000016984 (RANBP3L), ENSECAG00000016989 (OTOL1), ENSECAG00000017087 (ACSBG2), ENSECAG00000017249 (TRPS1), ENSECAG00000017338 (PFKFB4), ENSECAG00000017413 (TELO2), ENSECAG00000017484 (C20orf194), ENSECAG00000018025 (SIGLEC15), ENSECAG00000018068 (MROH2B), ENSECAG00000018069 (ARAP2), ENSECAG00000018094 (TBC1D16), ENSECAG00000018120, ENSECAG00000018966 (FBXO48), ENSECAG00000019067 (OSBPL7), ENSECAG00000019401 (TRIM77), ENSECAG00000019408 (CAT), ENSECAG00000019659, ENSECAG00000019681 (MAN1A2), ENSECAG00000019757, ENSECAG00000019803 (GALK1), ENSECAG00000019927, ENSECAG00000020113 (SEMA3C), ENSECAG00000020495 (CALHM2), ENSECAG00000020684 (C12orf49), ENSECAG00000020722, ENSECAG00000020904 (ZNF438), ENSECAG00000020955, ENSECAG00000021262 (MMP7), ENSECAG00000022357 (EIF4E1B), ENSECAG00000023063 (PTX4), ENSECAG00000023441 (E2F3), ENSECAG00000023641 (SNX8), ENSECAG00000024133 (FGF14), ENSECAG00000024176 (SIGLECL1), ENSECAG00000024181, ENSECAG00000024722 (SCARB2), ENSECAG00000024996 |
| | 4-5Kb | ENSECAG00000000765, ENSECAG00000000937, ENSECAG00000001160, ENSECAG00000001479, ENSECAG00000001775 (RAB38), ENSECAG00000001796, ENSECAG00000001934, ENSECAG00000002070, ENSECAG00000002084, ENSECAG00000002127, ENSECAG00000002408 (ZNF329), ENSECAG00000003765, ENSECAG00000004023 (SSBP1), ENSECAG00000004130, ENSECAG00000004271 (OR10A4), ENSECAG00000004674, ENSECAG00000005111 (RFX4), ENSECAG00000005153, ENSECAG00000005585, ENSECAG00000005669, ENSECAG00000005981, ENSECAG00000007007 (STT3B), ENSECAG00000007214 (CACNG4), ENSECAG00000007543 (LRRC46), ENSECAG00000007740 (HESX1), ENSECAG00000008107 (FERMT3), ENSECAG00000008289 (C1orf27), ENSECAG00000008334 (SMARCA4), ENSECAG00000008494, ENSECAG00000008545 (ECM1), ENSECAG00000008980 (FBXL21), ENSECAG00000009476, ENSECAG00000009558 (SLC16A12), ENSECAG00000009642 (ADAMTSL5), ENSECAG00000009720 (RIMBP2), ENSECAG00000009759 (TNRC6B), ENSECAG00000010259 (TRIM45), ENSECAG00000010794 (SPACA1), ENSECAG00000010973 (IL7R), ENSECAG00000010990, ENSECAG00000011061 (ZNF789), ENSECAG00000011435 (MSI2), ENSECAG00000011602, ENSECAG00000011801 (FHIT), ENSECAG00000012474 (TTF2), ENSECAG00000012607 (ABTB1), ENSECAG00000012653 (DYRK4), ENSECAG00000012853 (SP6), ENSECAG00000012864 (ZNF530), ENSECAG00000012881 (NME6), ENSECAG00000012980 (TRNAU1AP), ENSECAG00000013070 (DAGLB), ENSECAG00000013238 (FAM83D), ENSECAG00000013328 (TMOD2), ENSECAG00000013370, ENSECAG00000013616 (DSG2), ENSECAG00000013671 (EIF3B), ENSECAG00000013724 (FAS), ENSECAG00000013938 (ZDHHC13), ENSECAG00000014153 (CLK4), ENSECAG00000014304 (ATP11A), ENSECAG00000014425 (PIGK), ENSECAG00000015053 (SDCBP), ENSECAG00000015443 (OSBP2), ENSECAG00000016024, ENSECAG00000016463 (TMEM72), ENSECAG00000016682 (RIT2), ENSECAG00000016984 (RANBP3L), ENSECAG00000017078 (CYLC2), ENSECAG00000017081 (DZANK1), ENSECAG00000017087 (ACSBG2), ENSECAG00000017484 (C20orf194), ENSECAG00000017658 (CRNN), ENSECAG00000018025 (SIGLEC15), ENSECAG00000018068 (MROH2B), ENSECAG00000019067 (OSBPL7), ENSECAG00000019401 (TRIM77), ENSECAG00000019621, ENSECAG00000019626 (FAM46C), ENSECAG00000019757, ENSECAG00000019825 (TRDC), ENSECAG00000019888 (ARFIP2), ENSECAG00000020170 (RBP2), ENSECAG00000020479 (GEMIN5), ENSECAG00000020955, ENSECAG00000021355 (STAM2), ENSECAG00000021609 (SLC40A1), ENSECAG00000022128, ENSECAG00000022141 (TSLP), ENSECAG00000022359 (FOXRED1), ENSECAG00000022584 (DOCK6), ENSECAG00000023244 (TAF1B), ENSECAG00000023475, ENSECAG00000023616 (DDX4), ENSECAG00000023724 (ZNF235), ENSECAG00000023731 (ABCD3), ENSECAG00000023832 (GAK), ENSECAG00000024176 (SIGLECL1), ENSECAG00000024187 (SMARCA2), ENSECAG00000024198 (SELI) |
| | 5-6Kb | ENSECAG00000000188 (NMD3), ENSECAG00000000765, ENSECAG00000001096, ENSECAG00000001302, ENSECAG00000002070, ENSECAG00000002183, ENSECAG00000002548 (RNF182), ENSECAG00000002803, ENSECAG00000003137 (P4HTM), ENSECAG00000003256 (KRTAP8-1), ENSECAG00000003791 (NR2C2), ENSECAG00000003992, ENSECAG00000004271 (OR10A4), ENSECAG00000005289 (SPTSSB), ENSECAG00000005585, ENSECAG00000005669, ENSECAG00000005676, ENSECAG00000005765 (ADI1), ENSECAG00000005889 (TIMMDC1), ENSECAG00000006371, ENSECAG00000006851, ENSECAG00000007176 (ELOVL7), ENSECAG00000007214 (CACNG4), ENSECAG00000007525 (CPSF4), ENSECAG00000008046 (EIF4GI), ENSECAG00000008426 (PAPSS2), ENSECAG00000008494, ENSECAG00000008660 (YAP1), ENSECAG00000008846 (LGALS2), ENSECAG00000008980 (FBXL21), ENSECAG00000009197 (LANCL1), ENSECAG00000009430 (C16orf72), ENSECAG00000009521 (WNT8B), ENSECAG00000010117, ENSECAG00000010331 (PCGF3), ENSECAG00000010407 (BAG4), ENSECAG00000010684 (PHIP), ENSECAG00000011015, ENSECAG00000011154 (CHRD), ENSECAG00000011444 (THOC5), ENSECAG00000011801 (FHIT), ENSECAG00000011824, ENSECAG00000012653 (DYRK4), ENSECAG00000012859 (SP2), ENSECAG00000012864 (ZNF530), ENSECAG00000012870 (NEURL1), ENSECAG00000012927 (KIAA0319L), ENSECAG00000012957 (RHAG), ENSECAG00000012980 (TRNAU1AP), ENSECAG00000013270 (ZNF134), ENSECAG00000013418 (ICT1), ENSECAG00000013471 (ZNF599), ENSECAG00000013553 (ZNF8), ENSECAG00000013582 (DHRS3), ENSECAG00000013671 (EIF3B), ENSECAG00000013938 (ZDHHC13), ENSECAG00000014153 (CLK4), ENSECAG00000014304 (ATP11A), ENSECAG00000014737 (HEATR5B), ENSECAG00000015063 (ADGRG6), ENSECAG00000015545 (ARHGEF15), ENSECAG00000015612 (FAM212A), ENSECAG00000016256 (HPCA), ENSECAG00000016463 (TMEM72), ENSECAG00000016790, ENSECAG00000016814 (PPM1A), ENSECAG00000016984 (RANBP3L), ENSECAG00000016989 (OTOL1), ENSECAG00000017087 (ACSBG2), ENSECAG00000017155, ENSECAG00000017249 (TRPS1), ENSECAG00000017337 (PDE6B), ENSECAG00000018022, ENSECAG00000018025 (SIGLEC15), ENSECAG00000018068 (MROH2B), ENSECAG00000018160 (TBX21), ENSECAG00000018688 (CDK5RAP3), ENSECAG00000019401 (TRIM77), ENSECAG00000019659, ENSECAG00000019757, ENSECAG00000019825 (TRDC), ENSECAG00000019853 (AMIGO2), ENSECAG00000019882 (CALHM3), ENSECAG00000020684 (C12orf49), ENSECAG00000021001 (SENP6), ENSECAG00000021077 (FGF23), ENSECAG00000021355 (STAM2), ENSECAG00000021361 (ANXA13), ENSECAG00000021490, ENSECAG00000022082 (C12orf4), ENSECAG00000022290 (GCA), ENSECAG00000023756 (SMYD4), ENSECAG00000024761 (INOS), ENSECAG00000024875, ENSECAG00000024981 (HFE), ENSECAG00000026829 |
| | 6-7Kb | ENSECAG00000000011 (ADGRA2), ENSECAG00000000869 (RPL26L1), ENSECAG00000000971 (SUCO), ENSECAG00000000999, ENSECAG00000001219 (OR51B4), ENSECAG00000001479, ENSECAG00000002070, ENSECAG00000002084, |

| | | |
|---|---|---|
| | | ENSECAG00000002155 (NUP93), ENSECAG00000002373, ENSECAG00000002408 (ZNF329), ENSECAG00000003148, ENSECAG00000003256 (KRTAP8-1), ENSECAG00000004271 (OR10A4), ENSECAG00000005111 (RFX4), ENSECAG00000005132 (OR2D3), ENSECAG00000005669, ENSECAG00000005765 (ADI1), ENSECAG00000007434 (PRR9), ENSECAG00000007740 (HESX1), ENSECAG00000008090 (TAS2R41), ENSECAG00000008247, ENSECAG00000008324 (LECT2), ENSECAG00000008426 (PAPSS2), ENSECAG00000008494, ENSECAG00000008959 (SAMHD1), ENSECAG00000009759 (TNRC6B), ENSECAG00000010011 (MAP1LC3A), ENSECAG00000010117, ENSECAG00000010367 (RAD51AP1), ENSECAG00000010461 (LSM1), ENSECAG00000011353 (MFSD7), ENSECAG00000011435 (MSI2), ENSECAG00000011602, ENSECAG00000011754 (RPIA), ENSECAG00000011794 (C6orf141), ENSECAG00000011824, ENSECAG00000012312 (ARHGAP15), ENSECAG00000012584 (GNG2), ENSECAG00000012607 (ABTB1), ENSECAG00000012715, ENSECAG00000012853 (SP6), ENSECAG00000012870 (NEURL1), ENSECAG00000012957 (RHAG), ENSECAG00000013070 (DAGLB), ENSECAG00000013553 (ZNF8), ENSECAG00000013671 (EIF3B), ENSECAG00000013706 (SMAD6), ENSECAG00000013859, ENSECAG00000014425 (PIGK), ENSECAG00000014491 (NLRC5), ENSECAG00000014501 (TMEM54), ENSECAG00000014891 (NRCAM), ENSECAG00000015012 (MARCH1), ENSECAG00000015200, ENSECAG00000015444 (SLC37A1), ENSECAG00000015637 (NREP), ENSECAG00000016336 (CDC25A), ENSECAG00000016463 (TMEM72), ENSECAG00000016984 (RANBP3L), ENSECAG00000017151 (LYPD5), ENSECAG00000017337 (PDE6B), ENSECAG00000017400 (SMURF1), ENSECAG00000017484 (C20orf194), ENSECAG00000018022, ENSECAG00000018068 (MROH2B), ENSECAG00000018617 (RPRD2), ENSECAG00000018688 (CDK5RAP3), ENSECAG00000018721, ENSECAG00000018901 (FAM120B), ENSECAG00000019182 (GPRC6A), ENSECAG00000019210 (FNDC5), ENSECAG00000019282 (HDC), ENSECAG00000019393 (ALS2CR11), ENSECAG00000019401 (TRIM77), ENSECAG00000019471 (RASSF6), ENSECAG00000019661 (RPF1), ENSECAG00000019757, ENSECAG00000019825 (TRDC), ENSECAG00000020113 (SEMA3C), ENSECAG00000020278 (RAP1GAP2), ENSECAG00000020341, ENSECAG00000020638 (NEB), ENSECAG00000021077 (FGF23), ENSECAG00000021239 (CCDC92), ENSECAG00000021355 (STAM2), ENSECAG00000021492 (ADSL), ENSECAG00000021706 (RPS6KC1), ENSECAG00000021776, ENSECAG00000021797 (LRRC39), ENSECAG00000021923 (GTF2F2), ENSECAG00000021984 (WBSCR16), ENSECAG00000022202 (ADGRD1), ENSECAG00000022402 (GFOD1), ENSECAG00000022588, ENSECAG00000023141 (LAMC1), ENSECAG00000023326 (TMEM175), ENSECAG00000023832 (GAK), ENSECAG00000024181, ENSECAG00000024345 (SUCNR1), ENSECAG00000024700 (NOLC1), ENSECAG00000026877 (C17orf98) |
| FM | 7-8Kb | ENSECAG00000000327 (RNF181), ENSECAG00000000540 (CCDC174), ENSECAG00000000937, ENSECAG00000001022 (UCN2), ENSECAG00000001808, ENSECAG00000002084, ENSECAG00000002155 (NUP93), ENSECAG00000002599, ENSECAG00000003256 (KRTAP8-1), ENSECAG00000003791 (NR2C2), ENSECAG00000003846 (ANKAR), ENSECAG00000004055 (F2R), ENSECAG00000004085 (FAM78B), ENSECAG00000004526, ENSECAG00000004649, ENSECAG00000004742, ENSECAG00000005291, ENSECAG00000005585, ENSECAG00000005676, ENSECAG00000006818, ENSECAG00000007032 (KLHL29), ENSECAG00000007147 (CAPRIN1), ENSECAG00000007950, ENSECAG00000007979 (PLG), ENSECAG00000008311 (CHMP4B), ENSECAG00000008494, ENSECAG00000008576 (CD300E), ENSECAG00000008738 (ID2), ENSECAG00000008921 (TRIM42), ENSECAG00000008980 (FBXL21), ENSECAG00000009066 (GABBR2), ENSECAG00000009430 (C16orf72), ENSECAG00000009846 (RAD1), ENSECAG00000009897 (CIPC), ENSECAG00000009900, ENSECAG00000009983, ENSECAG00000010117, ENSECAG00000010461 (LSM1), ENSECAG00000010476 (L3MBTL2), ENSECAG00000010669 (CRB1), ENSECAG00000010973 (IL7R), ENSECAG00000010990, ENSECAG00000011669 (TNN), ENSECAG00000011783 (PDS5A), ENSECAG00000011793, ENSECAG00000011801 (FHIT), ENSECAG00000011822 (SPTLC3), ENSECAG00000011824, ENSECAG00000012607 (ABTB1), ENSECAG00000012702 (PPBP), ENSECAG00000012918 (ZNF704), ENSECAG00000012952 (SUV420H1), ENSECAG00000013489 (GPATCH11), ENSECAG00000013553 (ZNF8), ENSECAG00000013671 (EIF3B), ENSECAG00000015182 (OSBPL1A), ENSECAG00000015200, ENSECAG00000015567 (FREM1), ENSECAG00000015637 (NREP), ENSECAG00000016312 (CD274), ENSECAG00000016322 (CYP24A1), ENSECAG00000016463 (TMEM72), ENSECAG00000016506 (PIWIL1), ENSECAG00000016657 (ZNF211), ENSECAG00000017074 (FSD2), ENSECAG00000017428 (TCAF1), ENSECAG00000017484 (C20orf194), ENSECAG00000017796 (NCAPD2), ENSECAG00000017907 (RNFT2), ENSECAG00000017924 (KCNG1), ENSECAG00000018068 (MROH2B), ENSECAG00000018094 (TBC1D16), ENSECAG00000018416 (PRAME), ENSECAG00000018511 (SORBS1), ENSECAG00000018721, ENSECAG00000019847 (NADK2), ENSECAG00000019853 (AMIGO2), ENSECAG00000019882 (CALHM3), ENSECAG00000020025 (HMGN3), ENSECAG00000020278 (RAP1GAP2), ENSECAG00000020356 (MEF2D), ENSECAG00000020722, ENSECAG00000021001 (SENP6), ENSECAG00000021233 (MSANTD3), ENSECAG00000021462 (SFMBT2), ENSECAG00000021556, ENSECAG00000021615 (PSMB2), ENSECAG00000022017 (FGF6), ENSECAG00000022068 (RCN1), ENSECAG00000022073 (KLHL3), ENSECAG00000022252 (SP140), ENSECAG00000022345 (ROR2), ENSECAG00000022352 (CLUL1), ENSECAG00000022474 (CALU), ENSECAG00000023424 (GABRA4), ENSECAG00000024055 (JUN), ENSECAG00000024187 (SMARCA2), ENSECAG00000024990 (SLC17A6), ENSECAG00000024996, ENSECAG00000026817 (ALDH1B1) |
| FM | 8-9Kb | ENSECAG00000000159 (C16orf78), ENSECAG00000000537, ENSECAG00000000540 (CCDC174), ENSECAG00000000616 (BARX2), ENSECAG00000000647, ENSECAG00000000721 (MYCBPAP), ENSECAG00000000765, ENSECAG00000001096, ENSECAG00000001273, ENSECAG00000001326 (FZD1), ENSECAG00000002084, ENSECAG00000002155 (NUP93), ENSECAG00000002739, ENSECAG00000003056 (TM6SF1), ENSECAG00000003256 (KRTAP8-1), ENSECAG00000003791 (NR2C2), ENSECAG00000004055 (F2R), ENSECAG00000004085 (FAM78B), ENSECAG00000004630, ENSECAG00000005132 (OR2D3), ENSECAG00000005153, ENSECAG00000005289 (SPTSSB), ENSECAG00000005291, ENSECAG00000005669, ENSECAG00000006442 (ADAM9), ENSECAG00000006875 (IP6K2), ENSECAG00000007463, ENSECAG00000007670 (BMP10), ENSECAG00000007974 (GALNT3), ENSECAG00000008404, ENSECAG00000008763 (LYPD3), ENSECAG00000008846 (LGALS2), ENSECAG00000008939 (RBM17), ENSECAG00000008980 (FBXL21), ENSECAG00000009066 (GABBR2), ENSECAG00000009200, ENSECAG00000009430 (C16orf72), ENSECAG00000009476, ENSECAG00000009506, ENSECAG00000009809 (ENDOG), ENSECAG00000009841 (AKAP1), ENSECAG00000009983, ENSECAG00000010123 (USP44), ENSECAG00000010190 (FEZ2), ENSECAG00000010242 (MCTP1), ENSECAG00000010259 (TRIM45), ENSECAG00000010461 (LSM1), ENSECAG00000010642, ENSECAG00000010807 (PXYLP1), ENSECAG00000011222 (NGEF), ENSECAG00000011688 (TRIOBP), ENSECAG00000011754 (RPIA), ENSECAG00000011801 (FHIT), ENSECAG00000011824, ENSECAG00000012333 (MRPS23), ENSECAG00000012607 (ABTB1), ENSECAG00000012834 (RRNAD1), ENSECAG00000012859 (SP2), ENSECAG00000013070 (DAGLB), ENSECAG00000013238 (FAM83D), ENSECAG00000013392 (TMC5), ENSECAG00000013682 (VSNL1), ENSECAG00000013706 (SMAD6), ENSECAG00000014336, ENSECAG00000014397 (C1orf146), ENSECAG00000014604 (PNPLA3), ENSECAG00000015255 (ARMC4), ENSECAG00000015465 (NCOA5), ENSECAG00000015567 (FREM1), ENSECAG00000015637 (NREP), ENSECAG00000015790 (ANKRD32), ENSECAG00000015852 (GOT1L1), ENSECAG00000015998 (ABRA), ENSECAG00000016024, ENSECAG00000016069 (CLYBL), ENSECAG00000016322 (CYP24A1), ENSECAG00000016454 (NTSR2), ENSECAG00000016463 (TMEM72), ENSECAG00000016710 (UPK3B), ENSECAG00000017040 (BTBD1), ENSECAG00000017337 (PDE6B), ENSECAG00000017484 (C20orf194), ENSECAG00000017658 (CRNN), ENSECAG00000018068 (MROH2B), ENSECAG00000018069 (ARAP2), ENSECAG00000018341 (CA11), ENSECAG00000018662 (CCDC66), ENSECAG00000018825 (SIK2), ENSECAG00000019282 (HDC), ENSECAG00000019401 (TRIM77), ENSECAG00000019471 (RASSF6), ENSECAG00000019570 (GRAMD4), ENSECAG00000019634 (DMC1), ENSECAG00000019708 (PIK3CG), ENSECAG00000019739 (PXDN), ENSECAG00000019757, ENSECAG00000019853 (AMIGO2), ENSECAG00000020638 (NEB), ENSECAG00000020684 (C12orf49), ENSECAG00000020955, ENSECAG00000021355 (STAM2), ENSECAG00000021875 (USP42), ENSECAG00000021993 (GCLM), ENSECAG00000022077 (ELF2), ENSECAG00000022162 (TBC1D1), ENSECAG00000023607 (DOCK5), ENSECAG00000023860 (TGM3), ENSECAG00000024374 (ERBB3), ENSECAG00000025029 (C1orf52) |
| FM | 9-10Kb | ENSECAG00000000159 (C16orf78), ENSECAG00000000170, ENSECAG00000000647, ENSECAG00000000721 (MYCBPAP), ENSECAG00000000765, ENSECAG00000001326 (FZD1), ENSECAG00000001479, ENSECAG00000002155 (NUP93), ENSECAG00000003256 (KRTAP8-1), ENSECAG00000003367 (RAD51C), ENSECAG00000003383, ENSECAG00000003776 (ENPP2), ENSECAG00000003791 (NR2C2), ENSECAG00000004649, ENSECAG00000004717, ENSECAG00000005111 (RFX4), ENSECAG00000005132 (OR2D3), ENSECAG00000005676, ENSECAG00000006548 (SPATA3), ENSECAG00000006855, ENSECAG00000007280, ENSECAG00000007543 (LRRC46), ENSECAG00000007881 (IFIH1), ENSECAG00000008528, ENSECAG00000008738 (ID2), ENSECAG00000008893 (PARVB), ENSECAG00000008939 (RBM17), ENSECAG00000008980 (FBXL21), ENSECAG00000009384 (NECAB1), ENSECAG00000009430 (C16orf72), ENSECAG00000009476, ENSECAG00000009759 (TNRC6B), ENSECAG00000009897 (CIPC), ENSECAG00000010242 (MCTP1), ENSECAG00000010259 (TRIM45), ENSECAG00000010367 (RAD51AP1), ENSECAG00000010461 (LSM1), ENSECAG00000010669 (CRB1), ENSECAG00000010723 (ANGEL2), ENSECAG00000010776 (TERF2IP), ENSECAG00000010807 (PXYLP1), ENSECAG00000011222 (NGEF), ENSECAG00000011338 (ABCB5), ENSECAG00000011600 (WNT9B), ENSECAG00000011754 (RPIA), ENSECAG00000011797 (MN-SOD), ENSECAG00000011933, ENSECAG00000012531 (ACYP2), ENSECAG00000012607 (ABTB1), ENSECAG00000012667 (RBFOX1), ENSECAG00000012715, ENSECAG00000012859 (SP2), ENSECAG00000012881 (NME6), ENSECAG00000012942, ENSECAG00000012991, ENSECAG00000013682 (VSNL1), ENSECAG00000014153 (CLK4), ENSECAG00000014610 (PRRT3), ENSECAG00000015003 (PLGRKT), ENSECAG00000015053 (SDCBP), ENSECAG00000015070 (PRPSAP2), ENSECAG00000015133 (FOXK1), ENSECAG00000015567 (FREM1), ENSECAG00000015637 (NREP), ENSECAG00000015821 (ZER1), ENSECAG00000016183 (STARD4), ENSECAG00000016256 (HPCA), ENSECAG00000016463 (TMEM72), ENSECAG00000016989 (OTOL1), ENSECAG00000017115, ENSECAG00000017211 (ASCC1), ENSECAG00000017473 (APPL1), ENSECAG00000017478 (B4GALT6), ENSECAG00000017484 (C20orf194), ENSECAG00000018068 (MROH2B), ENSECAG00000018094 (TBC1D16), ENSECAG00000018341 (CA11), ENSECAG00000018367 (ZWILCH), ENSECAG00000018401 (CNDP2), ENSECAG00000018465 (MS4A5), ENSECAG00000018751, ENSECAG00000018901 (FAM120B), ENSECAG00000019126 (SLC12A5), ENSECAG00000019137 (MAML3), ENSECAG00000019210 (FNDC5), ENSECAG00000019401 (TRIM77), ENSECAG00000019471 (RASSF6), ENSECAG00000019708 (PIK3CG), ENSECAG00000019757, ENSECAG00000019842 (NAA15), ENSECAG00000019847 (NADK2), ENSECAG00000019853 (AMIGO2), ENSECAG00000019882 (CALHM3), ENSECAG00000019955 (ADCK3), ENSECAG00000020025 (HMGN3), ENSECAG00000020392 (AZI2), ENSECAG00000020433 (TENM3), ENSECAG00000020638 (NEB), ENSECAG00000020684 (C12orf49), ENSECAG00000020710 (DRC1), ENSECAG00000020724 (TPR), ENSECAG00000020791 (FRMPD1), ENSECAG00000020843, ENSECAG00000021207 (THRAP3), ENSECAG00000021355 (STAM2), ENSECAG00000021662 (STMN1), ENSECAG00000021867 (MRPS7), ENSECAG00000021889 (ABR), ENSECAG00000022290 (GCA), ENSECAG00000022355, ENSECAG00000023416 (F3), ENSECAG00000023475, ENSECAG00000023616 (DDX4), ENSECAG00000024176 (SIGLECL1), ENSECAG00000024727 (CYTH4), ENSECAG00000024850, ENSECAG00000025029 (C1orf52) |

## Table S7.9. GO term enrichment for genes harbouring adaptive alleles at their 0-1Kb and 0-10Kb site categories.

| Site category | Panel | Model organism | GO Term |
|---|---|---|---|
| 0-10Kb | DOM | mouse | ATP binding (0.0206) |
| | | | dipeptidase activity (0.0206) |
| | | | adenyl nucleotide binding (0.0206) |
| | | | small molecule binding (0.0206) |
| | | | adenyl ribonucleotide binding (0.0206) |
| | | | nucleotide binding (0.0248) |
| | | | translation factor activity, nucleic acid binding (0.0248) |
| | | | anion binding (0.0248) |
| | | | nucleoside phosphate binding (0.0248) |
| | | | ribonucleoside binding (0.0314) |
| | FM | mouse | metabolic process (0.0016) |
| | | | single-organism metabolic process (0.0027) |
| | | | cellular metabolic process (0.0027) |
| | | | organic substance metabolic process (0.0039) |
| | | | primary metabolic process (0.0039) |
| | | | anatomical structure morphogenesis (0.0168) |
| | | | biological regulation (0.0180) |
| | | | localization (0.0180) |
| | | | cartilage condensation (0.0323) |
| | | | RNA binding (0.0091) |
| | | | R-SMAD binding (0.0411) |
| | | | cell periphery (0.0073) |
| | | | plasma membrane (0.0073) |
| | | | membrane (0.0080) |
| | | | intracellular organelle (0.0080) |
| | | | organelle (0.0080) |
| | | | cytoplasm (0.0080) |

## Table S7.10. KEGG pathway enrichment for genes harbouring adaptive alleles at their 0-1Kb and 0-10Kb site categories.

| Site category | Panel | Model organism | KEGG Pathway |
|---|---|---|---|
| 0-1Kb | DOM | mouse | Olfactory transduction (0.0100) |
| | FM | human | Purine metabolism (0.0159) |
| | | | RNA transport (0.0159) |
| | | | Ubiquitin mediated proteolysis (0.0159) |
| | | mouse | RNA transport (0.0145) |
| | | | Ubiquitin mediated proteolysis (0.0145) |
| | | | Purine metabolism (0.0145) |
| 0-10Kb | DOM | human | Graft-versus-host disease (0.0077) |
| | | | RNA transport (0.0077) |
| | | | Type I diabetes mellitus (0.0077) |
| | | | Olfactory transduction (0.0097) |
| | | | Apoptosis (0.0182) |
| | | mouse | RNA transport (0.0083) |
| | | | Graft-versus-host disease (0.0083) |
| | | | Type I diabetes mellitus (0.0083) |
| | | | Apoptosis (0.0130) |
| | | | Olfactory transduction (0.0170) |
| | | | Protein processing in endoplasmic reticulum (0.0317) |
| | | | Chemokine signaling pathway (0.0320) |
| | | | Cytokine-cytokine receptor interaction (0.0470) |
| | | | MAPK signaling pathway (0.0491) |
| | FM | human | RNA transport (0.0002) |
| | | | Pathways in cancer (0.0002) |
| | | | Metabolic pathways (0.0109) |
| | | | Melanoma (0.0143) |
| | | | Endocytosis (0.0148) |
| | | | Small cell lung cancer (0.0219) |
| | | | Histidine metabolism (0.0340) |
| | | | Amoebiasis (0.0420) |
| | | | Purine metabolism (0.0459) |
| | | | MAPK signaling pathway (0.0459) |
| | | mouse | Metabolic pathways (0.0001) |
| | | | Endocytosis (0.0042) |
| | | | Melanoma (0.0042) |
| | | | Purine metabolism (0.0042) |
| | | | Small cell lung cancer (0.0059) |
| | | | Histidine metabolism (0.0092) |
| | | | MAPK signaling pathway (0.0092) |
| | | | Complement and coagulation cascades (0.0196) |

# Table S7.11. Wiki pathways enrichment for genes harbouring adaptive alleles at their 0-1Kb and 0-10Kb site categories.

| Site category | Panel | Model organism | Wikipathway |
|---|---|---|---|
| 0-1Kb | DOM | mouse | GPCRs, Class A Rhodopsin-like (0.0005) |
| | | | GPCRs, Other (0.0005) |
| | FM | human | Muscle cell TarBase (0.0278) |
| | | mouse | GPCRs, Other (0.0378) |
| | | | GPCRs, Class A Rhodopsin-like (0.0378) |
| 0-10Kb | DOM | human | FAS pathway and Stress induction of HSP regulation (0.0180) |
| | | | TSH signaling pathway (0.0180) |
| | | | TCR Signaling Pathway (0.0266) |
| | | | Focal Adhesion (0.0450) |
| | | | MAPK signaling pathway (0.0450) |
| | | mouse | GPCRs, Class A Rhodopsin-like (0.0010) |
| | | | Iron Homeostasis (0.0010) |
| | | | FAS pathway and Stress induction of HSP regulation (0.0063) |
| | | | MAPK signaling pathway (0.0324) |
| | | | Focal Adhesion (0.0324) |
| | | | GPCRs, Other (0.0324) |
| | | | Chemokine signaling pathway (0.0324) |
| | FM | human | Muscle cell TarBase (0.0195) |
| | | | B Cell Receptor Signaling Pathway (0.0325) |
| | | mouse | ESC Pluripotency Pathways (0.0058) |
| | | | Non-odorant GPCRs (0.0058) |
| | | | mRNA processing (0.0058) |
| | | | serotonin and anxiety (0.0101) |
| | | | Complement and Coagulation Cascades (0.0128) |
| | | | GPCRs, Class A Rhodopsin-like (0.0155) |
| | | | EPO Receptor Signaling (0.0191) |
| | | | Focal Adhesion (0.0210) |
| | | | Odorant GPCRs (0.0219) |
| | | | miRs in Muscle Cell Differentiation (0.0238) |

## Table S7.12. Phenotype term enrichment for genes harbouring adaptive alleles at their 0-1Kb and 0-10Kb site categories.

| Site category | Panel | Model organism | Phenotype |
|---|---|---|---|
| 0-1Kb | DOM | mouse | abnormal hematopoietic cell number (0.0064) |
| | | | abnormal hematopoiesis (0.0072) |
| | | | abnormal blood cell morphology/development (0.0072) |
| | | | abnormal immune system morphology (0.0072) |
| | | | abnormal hematopoietic system morphology/development (0.0078) |
| | | | increased hematopoietic cell number (0.0078) |
| | | | hematopoietic system phenotype (0.0078) |
| | | | increased circulating glucose level (0.0096) |
| | | | abnormal leukocyte cell number (0.0146) |
| | | | increased T cell number (0.0160) |
| 0-10Kb | DOM | human | Prominent interphalangeal joints (0.0116) |
| | | | Dumbbell-shaped long bone (0.0155) |
| | | | Anterior rib cupping (0.0155) |
| | | | Cupped ribs (0.0232) |
| | | | Edema (0.0232) |
| | | | Flat acetabular roof (0.0232) |
| | | | Abnormal tarsal ossification (0.0232) |
| | | | Abnormal foot bone ossification (0.0261) |
| | | | Arthropathy (0.0284) |
| | | | Abnormality of fluid regulation (0.0371) |
| | | mouse | platyspondylia (0.0140) |
| | FM | mouse | mortality/aging (0.0031) |
| | | | abnormal survival (0.0336) |
| | | | abnormal nervous system physiology (0.0459) |

## Table S7.13. PheWas term enrichment for genes harbouring adaptive alleles at their 0-1Kb and 0-10Kb site categories.

| Site category | Panel | PheWas |
|---|---|---|
| 0-1Kb | FM | Biliary cirrhosis (0.0184) |
| | | Adverse effects of opiates and related narcotics in therapeutic use (0.0184) |
| | | Viral infection (0.0184) |
| | | First degree AV block (0.0184) |
| | | Cellulitis and abscess of trunk (0.0184) |
| | | Peripheral arterial disease (0.0219) |
| | | Superficial cellulitis and abscess (0.0391) |
| | | Atrial fibrillation & flutter (0.0460) |
| | | Disorders of fluid, electrolyte, and acid-base balance (0.0460) |
| | | Chronic sinusitis (0.0460) |
| 0-10Kb | DOM | Lump or mass in breast (0.0420) |
| | | Other rheumatic heart disease (0.0420) |
| | | Chronic rheumatic disease of the heart valves (0.0420) |
| | | Torticollis (0.0420) |
| | | Retinal detachments and defects (0.0420) |
| | | Other peripheral nerve disorders (0.0420) |
| | | Posterior pituitary disorders (0.0420) |
| | | Malignant neoplasm of ovary (0.0420) |
| | | Cancer of other female genital organs (0.0420) |
| | | Enthesopathy (0.0420) |

## Table S7.14. Disease enrichment for genes harbouring adaptive alleles at their 0-1Kb and 0-10Kb sites categories.

| Site category | Panel | Disease |
|---|---|---|
| 0-1Kb | FM | Kidney Neoplasms (0.0057) |
| | | Gallbladder Neoplasms (0.0057) |
| | | Drug-induced chronic hepatitis (0.0057) |
| | | Encephalomyelitis (0.0057) |
| | | Gallbladder Diseases (0.0057) |
| | | Endocrine disorder NOS (0.0057) |
| | | Endocrine System Diseases (0.0057) |
| | | Endocrine disturbance NOS (0.0057) |
| | | Anemia, Refractory (0.0057) |
| | | Hypothyroidism (0.0090) |
| 0-10Kb | DOM | Arthritis, Rheumatoid (0.0033) |
| | | Occupational Diseases (0.0033) |
| | | Autoimmune Diseases (0.0033) |
| | | Arthritis (0.0033) |
| | | Lens Diseases (0.0095) |
| | | Wounds and Injuries (0.0095) |
| | | Sunburn (0.0095) |
| | | Liver Failure, Acute (0.0111) |
| | | Liver Cirrhosis, Alcoholic (0.0111) |
| | | Degeneration of lumbar intervertebral disc (0.0111) |
| | FM | Demyelinating Diseases (0.0010) |
| | | Tooth, Supernumerary (0.0018) |
| | | Encephalomyelitis (0.0018) |
| | | Hypophosphatemia, Familial (0.0018) |
| | | Osteomalacia (0.0018) |
| | | Genetic Predisposition to Disease (0.0021) |
| | | Adhesion (0.0024) |
| | | Multiple Sclerosis (0.0036) |
| | | Hereditary Hypophosphatemic Rickets with Hypercalciuria(HHRH) (0.0036) |
| | | Chronic Disease (0.0036) |

## REFERENCES

1. Pakendorf B, et al. (2006) Investigating the effects of prehistoric migrations in Siberia: genetic variation and the origins of Yakuts. *Hum Genet* 120(3):334–353.

2. Crubézy E, et al. (2010) Human evolution in Siberia: from frozen bodies to ancient DNA. *BMC Evol Biol* 10(1):25.

3. Fedorova SA, et al. (2013) Autosomal and uniparental portraits of the native populations of Sakha (Yakutia): implications for the peopling of Northeast Eurasia. *BMC Evol Biol* 13(1):127.

4. Keyser C, et al. (2015) The ancient Yakuts: a population genetic enigma. *Philos Trans R Soc Lond B Biol Sci* 370(1660):20130385.

5. West BA (2009) *Encyclopedia of the Peoples of Asia and Oceania* (Facts On File, Incorporated).

6. Ferret C (2009) *Une civilisation du cheval* (BELIN). BELIN edition.

7. Bonnie LH (1995) International Encyclopedia of Horse Breeds.

8. Solomonov NG, Anufriev AI, Yadrikhinskii VF, Isaev AP (2009) Body temperature changes in purebred and hybrid Yakut horses under the conditions of Yakutia. *Dokl Biol Sci* 427(1):358–361.

9. Grigoreva NN, Koryakina LP (2008) Dynamics of products of a metabolism glycoproteide and activity of enzymes AcAT, AAT at different type the Yakut breed of a horse. *Dokl Ross Akad Selskokhozyaistvennykh Nauk* (3):44–46.

10. Sarkissian C Der, et al. (2014) Shotgun microbial profiling of fossil remains. *Mol Ecol* 23(7):1780–1798.

11. Gilbert MTP, et al. (2004) Ancient mitochondrial DNA from hair. *Curr Biol CB* 14(12):R463–464.

12. Vilstrup JT, et al. (2013) Mitochondrial Phylogenomics of Modern and Ancient Equids. *PLoS ONE* 8(2):e55950.

13. Orlando L, et al. (2013) Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* 499(7456):74–78.

14. Meyer M, Kircher M (2010) Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harb Protoc* 2010(6):db.prot5448.

15. Pedersen JS, et al. (2014) Genome-wide nucleosome map and cytosine methylation levels of an ancient human genome. *Genome Res* 24(3):454–466.

16. Schubert M, et al. (2014) Prehistoric genomes reveal the genetic foundation and cost of horse domestication. *Proc Natl Acad Sci U A* 111(52):E5661–9.

17. Seguin-Orlando A, et al. (2013) Ligation Bias in Illumina Next-Generation DNA Libraries: Implications for Sequencing Ancient Genomes. *PLoS ONE* 8(10):e78575.

18. Der Sarkissian C, et al. (2015) Evolutionary Genomics and Conservation of the Endangered Przewalski's Horse. *Curr Biol*. doi:10.1016/j.cub.2015.08.032.

19. Do K-T, et al. (2014) Genomic characterization of the Przewalski′s horse inhabiting Mongolian steppe by whole genome re-sequencing. *Livest Sci* 167(0):86–91.

20. Schubert 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.

21. Botchkareva NV, Ahluwalia G, Shander D (2006) Apoptosis in the Hair Follicle. *J Invest Dermatol* 126(2):258–264.

22. Rasmussen M, et al. (2010) Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* 463(7282):757–762.

23. Olalde I, et al. (2014) Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. *Nature*. doi:10.1038/nature12960.

24. Enk JM, et al. (2014) Ancient whole genome enrichment using baits built from modern DNA. *Mol Biol Evol* 31(5):1292–1294.

25. Briggs AW, et al. (2007) Patterns of damage in genomic DNA sequences from a Neandertal. *Proc Natl Acad Sci* 104(37):14616–14621.

26. Jónsson H, Ginolhac A, Schubert M, Johnson PLF, Orlando L (2013) mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics* 29(13):1682–1684.

27. Star B, et al. (2014) Palindromic Sequence Artifacts Generated during Next Generation Sequencing Library Preparation from Historic and Ancient DNA. *PLoS One* 9(3):e89676.

28. Reich D, et al. (2010) Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* 468(7327):1053–1060.

29. Wade CM, et al. (2009) Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* 326(5954):865–867.

30. Jónsson H, et al. (2014) Speciation with gene flow in equids despite extensive chromosomal plasticity. *Proc Natl Acad Sci U A* 111(52):18655–18660.

31. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4):357–359.

32. Segata N, et al. (2012) Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods* 9(8):811–814.

33. Suzuki R, Shimodaira H (2006) Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* 22(12):1540–1542.

34. The Human Microbiome Project Consortium (2012) Structure, function and diversity of the healthy human microbiome. *Nature* 486(7402):207–214.

35. Fierer N, et al. (2012) Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proc Natl Acad Sci U A* 109(52):21390–21395.

36. McLaren W, et al. (2010) Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* 26(16):2069–2070.

37. Baye TM, Tiwari HK, Allison DB, Go RC (2009) Database mining for selection of SNP markers useful in admixture mapping. *BioData Min* 2(1):1.

38. Nicholas FW (2003) Online Mendelian Inheritance in Animals (OMIA): a comparative knowledgebase of genetic disorders and other familial traits in non-laboratory animals. *Nucleic Acids Res* 31(1):275–277.

39. Doan R, et al. (2012) Whole-genome sequencing and genetic variant analysis of a Quarter Horse mare. *BMC Genomics* 13:78.

40. Signer-Hasler H, et al. (2012) A genome-wide association study reveals loci influencing height and other conformation traits in horses. *PLoS One* 7(5):e37282.

41. Alkan C, et al. (2009) Personalized copy number and segmental duplication maps using next-generation sequencing. *Nat Genet* 41(10):1061–1067.

42. Quinlan AR (2014) BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinforma* 47:11.12.1–11.12.34.

43. Wang W, et al. (2014) Genome-wide detection of copy number variations among diverse horse breeds by array CGH. *PLoS One* 9(1):e86860.

44. Ghosh S, et al. (2014) Copy number variation in the horse genome. *PLoS Genet* 10(10):e1004712.

45. Watterson GA (1979) Estimating and Testing Selection: The Two-Alleles, Genic Selection Diffusion Model. *Adv Appl Probab* 11(1):14–30.

46. Korneliussen TS, Albrechtsen A, Nielsen R (2014) ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* 15(1):356.

47. Korneliussen TS, Moltke I, Albrechtsen A, Nielsen R (2013) Calculation of Tajima’s D and other neutrality test statistics from low depth next-generation sequencing data. *BMC Bioinformatics* 14:289.

48. Prüfer K, et al. (2014) The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* 505(7481):43–49.

49. Killick R, Eckley I (2014) changepoint:an R package for changepoint analysis. *J Stat Softw* 58(3):1–19.

50. Hill EW, Gu J, McGivney BA, MacHugh DE (2010) Targets of selection in the Thoroughbred genome contain exercise-relevant gene SNPs associated with elite racecourse performance. *Anim Genet* 41 Suppl 2:56–63.

51. Bellone RR, et al. (2010) Fine-mapping and mutation analysis of TRPM1: a candidate gene for leopard complex (LP) spotting and congenital stationary night blindness in horses. *Brief Funct Genomics* 9(3):193–207.

52. Tryon RC, White SD, Bannasch DL (2007) Homozygosity mapping approach identifies a missense mutation in equine cyclophilin B (PPIB) associated with HERDA in the American Quarter Horse. *Genomics* 90(1):93–102.

53. Brooks SA, et al. (2010) Whole-Genome SNP Association in the Horse: Identification of a Deletion in Myosin Va Responsible for Lavender Foal Syndrome. *PLoS Genet* 6(4):e1000909.

54. Brault LS, Cooper CA, Famula TR, Murray JD, Penedo MCT (2011) Mapping of equine cerebellar abiotrophy to ECA2 and identification of a potential causative mutation affecting expression of MUTYH. *Genomics* 97(2):121–129.

55. Marklund L, Moller MJ, Sandberg K, Andersson L (1996) A missense mutation in the gene for melanocyte-stimulating hormone receptor (MCIR) is associated with the chestnut coat color in horses. *Mamm Genome* 7(12):895–899.

56. Wagner H-J, Reissmann M (2000) New polymorphism detected in the horse MC1R gene. *Anim Genet* 31(4):289–290.

57. Brooks SA, Bailey E (2005) Exon skipping in the KIT gene causes a Sabino spotting pattern in horses. *Mamm Genome* 16(11):893–902.

58. Brooks SA, Terry RB, Bailey E (2002) A PCR-RFLP for KIT associated with tobiano spotting pattern in horses. *Anim Genet* 33(4):301–303.

59. Makvandi-Nejad S, et al. (2012) Four Loci Explain 83% of Size Variation in the Horse. *PLoS ONE* 7(7):e39929.

60. Wijnberg ID, et al. (2012) A missense mutation in the skeletal muscle chloride channel 1 (CLCN1) as candidate causal mutation for congenital myotonia in a New Forest pony. *Neuromuscul Disord* 22(4):361–367.

61. Spirito F, et al. (2002) Animal Models for Skin Blistering Conditions: Absence of Laminin 5 Causes Hereditary Junctional Mechanobullous Disease in the Belgian Horse. *J Invest Dermatol* 119(3):684–691.

62. Hauswirth R, et al. (2012) Mutations in MITF and PAX3 Cause "Splashed White" and Other White Spotting Phenotypes in Horses. *PLoS Genet* 8(4):e1002653.

63. Hauswirth R, et al. (2013) Novel variants in the *KIT* and *PAX3* genes in horses with white-spotted coat colour phenotypes. *Anim Genet* 44(6):763–765.

64. Brunberg E, et al. (2006) A missense mutation in PMEL17 is associated with the Silver coat color in the horse. *BMC Genet* 7:46.

65. Shin EK, Perryman LE, Meek K (1997) A kinase-negative mutation of DNA-PK(CS) in equine SCID results in defective coding and signal joint formation. *J Immunol* 158(8):3565–3569.

66. Aleman M, et al. (2004) Association of a mutation in the ryanodine receptor 1 gene with equine malignant hyperthermia. *Muscle Nerve* 30(3):356–365.

67. Gu J, et al. (2010) Association of sequence variants in CKM (creatine kinase, muscle) and COX4I2 (cytochrome c oxidase, subunit 4, isoform 2) genes with racing performance in Thoroughbred horses: SNP association with elite racing performance. *Equine Vet J* 42:569–575.

68. McCue ME, et al. (2012) A High Density SNP Array for the Domestic Horse and Extant Perissodactyla: Utility for Association Mapping, Genetic Diversity, and Phylogeny Studies. *PLoS Genet* 8(1):e1002451.

69. Cannon SC, Hayward LJ, Beech J, Brown RH (1995) Sodium channel inactivation is impaired in equine hyperkalemic periodic paralysis. *J Neurophysiol* 73(5):1892–1899.

70. Christopherson PW, Santen VL, Livesey L, Boudreaux MK (2007) A 10-Base-Pair Deletion in the Gene Encoding Platelet Glycoprotein IIb Associated with Glanzmann Thrombasthenia in a Horse. *J Vet Intern Med* 21(1):196–198.

71. Orr N, et al. (2010) Genome-wide SNP association-based localization of a dwarfism gene in Friesian dwarf horses: SNP association of a dwarfism gene. *Anim Genet* 41:2–7.

72. Cook D, Brooks S, Bellone R, Bailey E (2008) Missense Mutation in Exon 2 of SLC36A1 Responsible for Champagne Dilution in Horses. *PLoS Genet* 4(9):e1000195.

73. Hansen M, Knorr C, Hall AJ, Broad TE, Brenig B (2007) Sequence analysis of the equine &lt;i&gt;SLC26A2&lt;/i&gt; gene locus on chromosome 14q15&amp;rarr;q21. *Cytogenet Genome Res* 118(1):55–62.

74. Yang G (1998) A dinucleotide mutation in the endothelin-B receptor gene is associated with lethal white foal syndrome (LWFS); a horse variant of Hirschsprung disease. *Hum Mol Genet* 7(6):1047–1052.

75. Tozaki T, et al. (2010) A genome-wide association study for racing performances in Thoroughbreds clarifies a candidate region near the MSTN gene: A genome-wide scan for racing performances. *Anim Genet* 41:28–35.

76. W. Hill E, P. Ryan D, E. MacHugh D (2012) Horses for Courses: a DNA-based Test for Race Distance Aptitude in Thoroughbred Racehorses. *Recent Pat DNA Gene Seq* 6(3):203–208.

77. Mariat D, Taourit S, Guérin G (2003) A mutation in the MATP gene causes the cream coat colour in the horse. *Genet Sel Evol* 35(1):119.

78. Rieder S, Taourit S, Mariat D, Langlois B, Gu&#x000E9;rin G (2001) Mutations in the agouti (ASIP), the extension (MC1R), and the brown (TYRP1) loci and their association to coat color phenotypes in horses (Equus caballus). *Mamm Genome* 12(6):450–455.

79. Andersson LS, et al. (2012) Mutations in DMRT3 affect locomotion in horses and spinal circuit function in mice. *Nature* 488(7413):642–646.

80. Fox-Clipsham LY, et al. (2011) Identification of a Mutation Associated with Fatal Foal Immunodeficiency Syndrome in the Fell and Dales Pony. *PLoS Genet* 7(7):e1002133.

81. Révay T, Villagómez DAF, Brewer D, Chenier T, King WA (2012) GTG Mutation in the Start Codon of the Androgen Receptor Gene in a Family of Horses with 64,XY Disorder of Sex Development. *Sex Dev* 6(1-3):108–116.

82. Towers RE, et al. (2013) A Nonsense Mutation in the IKBKG Gene in Mares with Incontinentia Pigmenti. *PLoS ONE* 8(12):e81625.

83. 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.

84. Li H, et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.

85. Keane T, Creevey C, Pentony M, Naughton T, Mclnerney J (2006) Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol Biol* 6(1):29.

86. Schwarz G (1978) Estimating the Dimension of a Model. *Ann Stat* 6(2):461–464.

87. Guindon S, et al. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59(3):307–321.

88. Anisimova M, Gascuel O (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol* 55(4):539–552.

89. Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* 22(2):160–174.

90. Huson DH, Scornavacca C (2012) Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst Biol* 61(6):1061–1067.

91. Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29(8):1969–1973.

92. Rambaut A, Suchard MA, Xie D, Drummond AJ (2014) *Tracer v1.6*.

93. Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian Coalescent Inference of Past Population Dynamics from Molecular Sequences. *Mol Biol Evol* 22(5):1185–1192.

94. Minin VN, Bloomquist EW, Suchard MA (2008) Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol Biol Evol* 25(7):1459–1471.

95. Wickham H (2009) *ggplot2: Elegant Graphics for Data Analysis* (Springer New York).

96. Wallner B, et al. (2013) Identification of Genetic Variation on the Horse Y Chromosome and the Tracing of Male Founder Lineages in Modern Breeds. *PLoS One* 8(4):e60015.

97. Lippold S, Matzke NJ, Reissmann M, Hofreiter M (2011) Whole mitochondrial genome sequencing of domestic horses reveals incorporation of extensive wild horse diversity during domestication. *BMC Evol Biol* 11:328.

98. Flicek P, et al. (2013) Ensembl 2013. *Nucleic Acids Res* 41(D1):D48–D55.

99. Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21):2688–2690.

100. Goto H, et al. (2011) A massively parallel sequencing approach uncovers ancient origins and high genetic variability of endangered Przewalski's horses. *Genome Biol Evol* 3:1096–1106.

101. Achilli A, et al. (2012) Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. *Proc Natl Acad Sci U S A* 109(7):2449–2454.

102. Xu S, et al. (2007) High altitude adaptation and phylogenetic analysis of Tibetan horse based on the mitochondrial genome. *J Genet Genomics* 34(8):720–729.

103. Jiang Q, et al. (2011) The complete mitochondrial genome and phylogenetic analysis of the Debao pony (Equus caballus). *Mol Biol Rep* 38(1):593–599.

104. Sawyer S, Krause J, Guschanski K, Savolainen V, Pääbo S (2012) Temporal Patterns of Nucleotide Misincorporations and DNA Fragmentation in Ancient DNA. *PLoS ONE* 7(3):e34131.

105. Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 10(3):512–526.

106. Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* 475(7357):493–496.

107. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet* 5(10):e1000695.

108. Outram AK, et al. (2009) The Earliest Horse Harnessing and Milking. *Science* 323(5919):1332–1335.

109. Terhorst J, Song YS (2015) Fundamental limits on the accuracy of demographic inference based on the sample frequency spectrum. *Proc Natl Acad Sci* 112(25):7677–7682.

110. Yang MA, Harris K, Slatkin M (2014) The projection of a test genome onto a reference population and applications to humans and archaic hominins. *Genetics* 198(4):1655–1670.

111. Danecek P, et al. (2011) The Variant Call Format and VCFtools. *Bioinformatics*. doi:10.1093/bioinformatics/btr330.

112. Price AL, et al. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* 38(8):904–909.

113. Fumagalli M, Vieira FG, Linderoth T, Nielsen R (2014) ngsTools: methods for population genetics analyses from next-generation sequencing data. *Bioinformatics* 30(10):1486–1487.

114. Green RE, et al. (2010) A Draft Sequence of the Neandertal Genome. *Science* 328(5979):710–722.

115. Patterson N, et al. (2012) Ancient Admixture in Human History. *Genetics* 192(3):1065–1093.

116. Pickrell JK, Pritchard JK (2012) Inference of Population Splits and Mixtures from Genome-Wide Allele Frequency Data. *PLoS Genet* 8(11):e1002967.

117. Skotte L, Korneliussen TS, Albrechtsen A (2013) Estimating individual admixture proportions from next generation sequencing data. *Genetics* 195(3):693–702.

118. Durand EY, Patterson N, Reich D, Slatkin M (2011) Testing for ancient admixture between closely related populations. *Mol Biol Evol* 28(8):2239–2252.

119. Petersen JL, et al. (2013) Genome-wide analysis reveals selection for important traits in domestic horse breeds. *PLoS Genet* 9(1):e1003211.

120. Wittkopp PJ, Kalay G (2011) Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet* 13(1):59–69.

121. King MC, Wilson AC (1975) Evolution at two levels in humans and chimpanzees. *Science* 188(4184):107–116.

122. Carroll SB (2005) Evolution at two levels: on genes and form. *PLoS Biol* 3(7):e245.

123. Carroll SB (2008) Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134(1):25–36.

124. Hoekstra HE, Coyne JA (2007) The locus of evolution: evo devo and the genetics of adaptation. *Evolution* 61(5):995–1016.

125. Fraser HB (2013) Gene expression drives local adaptation in humans. *Genome Res* 23(7):1089–1096.

126. Ohno S (1970) *Evolution by Gene Duplication:* (Springer Berlin Heidelberg).

127. Nei M (1987) *Molecular Evolutionary Genetics* (Columbia University Press).

128. Zhang B, Kirov S, Snoddy J (2005) WebGestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res* 33(Web Server issue):W741–8.

129. Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B Stat Methodol* 57(1):289–300.