

SUPPLEMENTARY INFORMATION

METHODS

The Y chromosome fragments used for SNP screening were derived in three different ways.

(i) A Y-specific horse BAC clone was isolated by screening an equine BAC library for clones containing the *ZFY* gene. Following subcloning and DNA sequencing, PCR primers were designed to generate short fragments (typically 300-500 base pairs) for polymorphism screening. These regions were augmented with (ii) introns from Y-linked genes¹ and (iii) anonymous, non-coding horse Y chromosome sequence² available in GenBank.

PCR screening of BAC library. Primers (AAGCAGAGAAGGCTATCGAA and CACAGTACTGACAGTGGTAT) were designed for the equine zinc finger protein Y gene (*ZFY*; accession no. AF133198) to amplify a 297 bp fragment from within one of the exons; the design of these primers was that they should not amplify the homologous gene on the X chromosome, *ZFX*. A horse bacterial artificial chromosome (BAC) library was screened by PCR for the presence of clones containing the *ZFY* gene. Initially, primers were applied to amplify DNA from the 78 superpools of the library; the superpools exhibiting an amplification product of expected length were further analysed by performing PCR on the corresponding plate / row / column pools.

Subcloning and sequencing of a ZFY-positive BAC clone. BAC DNA was isolated from 10 ml 2xTY (containing 25µg/ml chloramphenicol) overnight cultures by standard methods using GET lysis buffer (50mM Glucose, 10 mM EDTA, 25mM Tris pH 8.0). Fragmentation of BAC DNA was achieved by restriction endonuclease digestion with *Sau3AI*. DNA fragments in the size range of 500-1000 bp were selected on a 1% agarose gel and extracted from a gel slice using QIAquick gel extraction kit (Qiagen). These fragments were subsequently ligated into phosphatased *BamH1* M13 vector (Appligene) and transformed into

Novagen competent cells. Preparation of single strand phage DNA was performed by using the standard Sanger 96 well protocol described at (<http://www.sanger.ac.uk/Teams/Team51/Triton2.shtml>). Cycle sequencing was performed using dye terminator chemistry (PE Biosystems) with reactions analysed on an Applied Biosystems 3700 instrument. The BLAST program was used to search GenBank with the obtained subclone sequences to identify possible homologies to equine repeat elements.

Amplification of Y-specific fragments. Primers (all primers are available upon request) were designed to amplify unique BAC subclone sequences, yielding amplicons of approximately 300-500 bp in length (16 fragments). We amplified introns (another 16 fragments) from the Y-linked genes *AMELY*, *DBY*, *SMCY* and *SRY* using conserved exonic primers (“YCATS”; Y chromosome conservative anchored tagged sequences¹) or horse-specific internal introns primers. Finally, five fragments from anonymous, non-coding Y chromosome sequence were also amplified². Each 50 µl PCR reaction contained 1.25 units of AmpliTaq Gold DNA polymerase (Perkin-Elmer), 200 µM dNTPs, 1.5-2.5 mM MgCl₂, 50 mM KCl, 10 mM Tris-HCl (pH 8.3), 10-25 pmoles of each primer, and 100 ng genomic DNA. The standard PCR profile consisted of one cycle of 94°C for 5 min to activate AmpliTaq Gold, 58°C for 30 s, and 72°C for 30 s, followed by 29 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 30 s. Alternatively, a touch down PCR profile was used, where the annealing temperature ranged from 55-45°C with a decrease of 0.5°C/cycle for 20 cycles, followed by a constant annealing temperature of 45°C for 10 cycles. All primers were confirmed to amplify Y chromosome-specific products by using female horse DNA samples as negative controls; no marker amplified female DNA.

Samples. A total of 52 male horses were screened, including four of Arab, Ardennais, Connemara, Exmoor, Fjord, Gotland, Icelandic, Shetland, North-Swedish, and Thoroughbred, three of Khuzestan Arab and Caspian Pony, and two of Akhal Teké, Malwari

and Thai Pony. A single male Przewalski's horse, derived from Helsinki zoo, Finland, was also analysed, as was a male donkey. DNA was prepared from blood or serum samples.

Amplification of X-specific fragments. Conserved primers were designed from mammalian *DBX* and *SMCX* gene sequences and were used to amplify X-linked intron sequences (Supplementary Table 1). Nineteen female horses, mainly from the same breeds as used in the screening for Y chromosome variability (four of Standardbred, three of Gotland, two of Exmoor, Icelandic, North-Swedish, Shetland, and Thoroughbred, and one of Arab and Ardennais), and one female donkey were used for amplification with these primers.

SNP screening by DNA sequencing. Amplification products were purified with QIAquick PCR purification columns (Qiagen). The corresponding amplimers were used as sequencing primers in both forward and reverse sequencing, with dye terminator chemistry, and analysed on an ABI377 instrument. Alternatively, amplification products were purified with EXO-SAP-IT (USB), sequenced with Amersham's ET dye terminator kit and recorded on a MegaBace 1000 instrument. The obtained sequences were aligned and analysed for polymorphic sites using AutoAssembler, Sequence Navigator (Applied Biosystems), or Sequencher (Gene Codes). Two of six polymorphic positions identified in the Przewalski's horse have been independently confirmed in a study of 1.8 kb of the same Y chromosome sequence as analysed here². This study also screened the 1.8 kb for polymorphism among 21 male horses but found no variation; due to the limited amount of equine sequence surveyed this does not represent statistically significant evidence for lower Y chromosome variation in domestic horses than in humans.

As X chromosome sequences were surveyed in female horses we encountered cases of heterozygote positions, identified as double peaks in sequence reads. Besides being confirmed

from forward and reverse sequencing, such heterozygote positions were also confirmed by independent amplification and sequencing reactions.

Data analysis. Average pairwise difference between alleles or haplotypes, nucleotide diversity (π), and its standard deviation was estimated in DnaSP³. The same program was used for the Hudson-Kreitman-Aquadro (HKA) test⁴. This test contrasts levels of interspecific divergence and intraspecific polymorphism in two or more genomic regions. Divergence between domestic horse and donkey was 1.19% (95% confidence interval, CI, 1.01-1.37) for Y chromosome sequences, and 0.97% (95% CI: 0.54-1.40) for X chromosome sequences.

1. Hellborg, L. & Ellegren, H. *Mol. Ecol.* **12**, 283-291 (2003).
2. Wallner, B. *et al. Anim. Genet.* **34**, 453-456 (2003).
3. Rozas, J. & Rozas, R. *Bioinformatics* **15**, 174-175 (1999).
4. Hudson, R. R. *et al. Genetics* **116**, 153-159 (1987).