

## Asian horses deepen the MSY phylogeny

S. Felkel<sup>\*,†</sup>, C. Vogl<sup>\*</sup>, D. Rigler<sup>\*</sup>, V. Jagannathan<sup>‡</sup>, T. Leeb<sup>‡</sup> , R. Fries<sup>§</sup>, M. Neuditschko<sup>¶</sup>,  
S. Rieder<sup>¶</sup> , B. Velie<sup>\*\*</sup>, G. Lindgren<sup>\*\*</sup>, C.-J. Rubin<sup>††</sup>, C. Schlötterer<sup>‡‡</sup>, T. Rattei<sup>§§</sup>, G. Brem<sup>\*</sup> and  
B. Wallner<sup>\*</sup>

<sup>\*</sup>Institute of Animal Breeding and Genetics, University of Veterinary Medicine Vienna, Vienna, Austria. <sup>†</sup>Vienna Graduate School of Population Genetics, Vienna, Austria. <sup>‡</sup>Institute of Genetics, Vetsuisse Faculty, University of Bern, Bern, Switzerland. <sup>§</sup>Lehrstuhl für Tierzucht, Technische Universität München, Freising, Germany. <sup>¶</sup>Agroscope, Swiss National Stud Farm, Avenches, Switzerland. <sup>\*\*</sup>Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden. <sup>††</sup>Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden. <sup>‡‡</sup>Institut für Populationsgenetik, University of Veterinary Medicine Vienna, Vienna, Austria. <sup>§§</sup>Department of Microbiology and Ecosystem Science, Division of Computational Systems Biology, University of Vienna, Vienna, Austria.

### Summary

Humans have shaped the population history of the horse ever since domestication about 5500 years ago. Comparative analyses of the Y chromosome can illuminate the paternal origin of modern horse breeds. This may also reveal different breeding strategies that led to the formation of extant breeds. Recently, a horse Y-chromosomal phylogeny of modern horses based on 1.46 Mb of the male-specific Y (MSY) was generated. We extended this dataset with 52 samples from five European, two American and seven Asian breeds. As in the previous study, almost all modern European horses fall into a crown group, connected via a few autochthonous Northern European lineages to the outgroup, the Przewalski's Horse. In total, we now distinguish 42 MSY haplotypes determined by 158 variants within domestic horses. Asian horses show much higher diversity than previously found in European breeds. The Asian breeds also introduce a deep split to the phylogeny, preliminarily dated to  $5527 \pm 872$  years. We conclude that the deep splitting Asian Y haplotypes are remnants of a far more diverse ancient horse population, whose haplotypes were lost in other lineages.

**Keywords** Equus, breeding, paternal lineage, Y chromosome, haplotype, ancestry, diversity, history, oriental origin

Domestication of the horse started about 5500 years ago in Central Asia (Outram *et al.* 2009) and revolutionized mobility, transport, economy and warfare of ancient human cultures (Kelekna 2009). Today, the horse is an invaluable companion in many human societies, and numerous breeds listed today are witnesses of the long-term relationship between humans and horses. The origin of the horse and its diversity has been studied intensively with autosomal DNA and maternally inherited mitochondrial DNA (mtDNA) (Vilà *et al.* 2001; Achilli *et al.* 2012; Petersen *et al.* 2013). The male-specific Y (MSY) can be used to trace the origin of stallion lines, as it is inherited strictly through the paternal line. This was, however, precluded until recently due to the low sequence variability on the horse MSY (Lindgren *et al.* 2004; Wallner *et al.* 2013).

#### Address for correspondence

B. Wallner, Institute of Animal Breeding and Genetics, University of Veterinary Medicine Vienna, Vienna, Austria.  
E-mail: barbara.wallner@vetmeduni.ac.at

Accepted for publication 26 November 2017

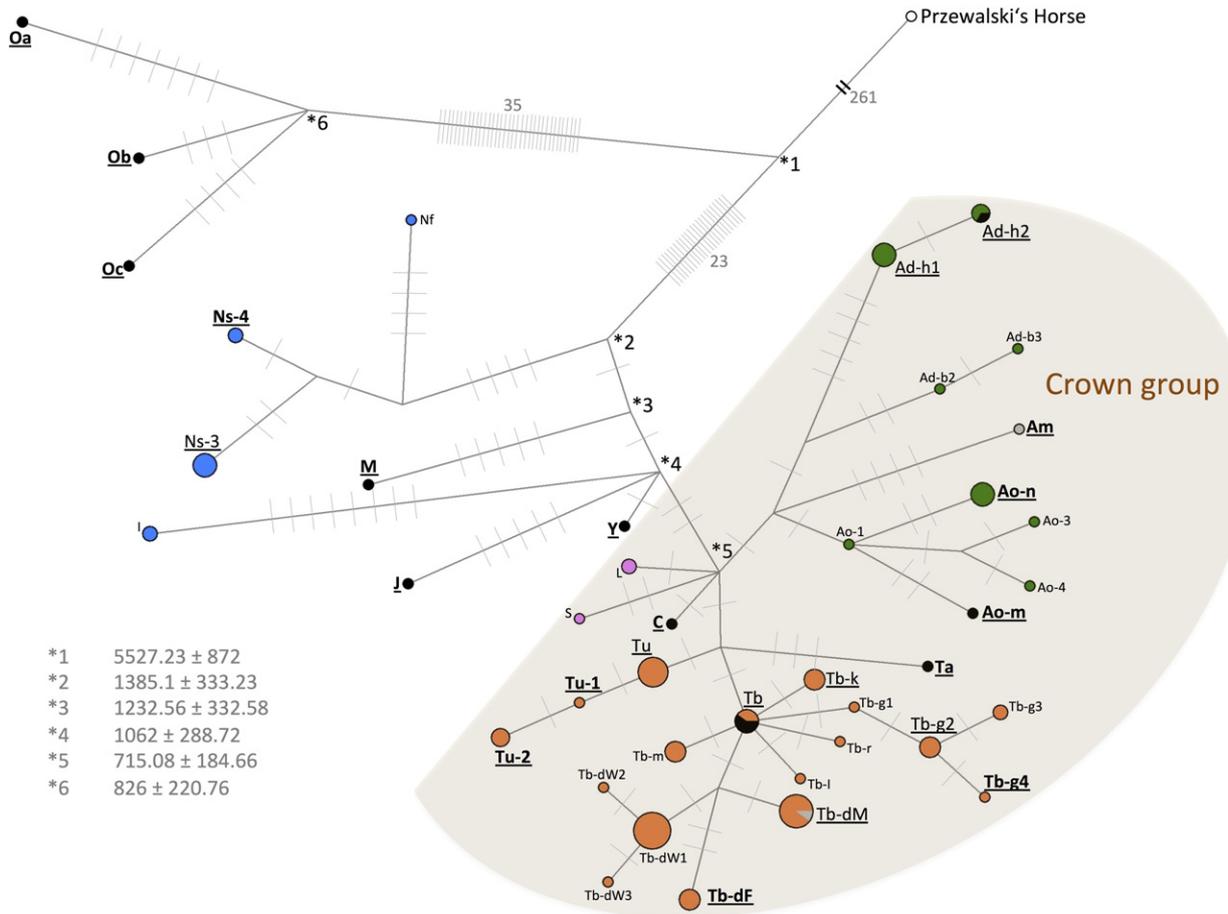
In a recent study, Wallner *et al.* (2017) created a domestic horse MSY phylogeny based on variants found in 52 horses of 21 breeds on a 1.46 Mb singlecopy MSY reference. They showed that, apart from few Northern European lineages, the MSY haplotypes of modern horse breeds cluster together in a roughly 700-year-old haplogroup. They claimed that the predominance of this MSY crown group is a signature of the intense male-biased breeding practices during the past 400–500 years and concluded that these haplotypes were introduced to Europe and America by imported Oriental stallions. So far, the short MSY phylogeny mirrors only recent stallion population dynamics. Autosomal and mtDNA studies, as well as MSY studies on ancient horses, indicate more genetic variability in rural breeds (Lippold *et al.* 2011; Warmuth *et al.* 2012; Librado *et al.* 2017). As these breeds, often from remote regions, are generally not intensively selected, they could retain private variation already lost in strongly selected modern breeds.

To test whether this is also evident for Y-chromosomal diversity we augmented the dataset from Wallner *et al.*

(2017) with horse breeds from different continents for our analyses. We analysed 52 additional horses from five European breeds (Franches-Montagnes, Noriker, Haflinger, Swedish Draft and Warmblood Horse), two American breeds (Mangalarga Marchador and Tennessee Walking Horse) and seven Asian breeds (Chakouyi, Marwari, Akhal-Teke, Arabian, Mongolian, Yakutian Horse and Jeju Pony). In total we used whole-genome next-generation sequenced (NGS) data from 105 males, including a Przewalski's Horse as an outgroup (see detailed sample description in Table S1), for MSY variant ascertainment. Twenty whole-genome sequenced samples were generated *de novo* as part of this study, and 85 were derived from publicly available databases. We mapped the data (genome coverage ranging from 8–32×) to the single copy MSY reference contigs (Genbank WGS master accession no. MPVR01000000) from Wallner *et al.* (2017), called variants and inferred MSY haplotypes—all according to previously published protocols

(see details in Appendix S1). In addition to the 57 domestic and 284 Przewalski's variants (13 indels and 271 single nucleotide variants) described previously (Wallner *et al.* 2013, 2017), we found 101 new variants resulting in a total of 158 variants and 42 haplotypes in domestics. We randomly selected 33 of the newly identified single nucleotide variants and successfully validated their Y specificity as well as their allelic state, when samples were available, using LGC-KASP technology (see Table S2). The relationship of the resulting haplotypes was visualized in a haplotype network with NETWORK 4.615 (Fig. 1) and in a maximum parsimony tree with PAUP 4.0a (Fig. S1; see details in Appendix S1). Dating of the most important nodes of the tree was performed as described in Wallner *et al.* (2017).

The median-joining network (Fig. 1) shows that all European breeds, as well as the Mangalarga Marchador and the Tennessee Walking Horse, cluster into the recently



**Figure 1** MSY haplotype network of 105 male horses (52 domestic horses from Wallner *et al.* 2017, 52 additional samples, 1 Przewalski's horse). Hatchmarks along branches correspond to number of mutations; the number of mutations is shown for long branches. Asian samples are shown in black, American samples in grey and Northern European samples in blue. Haplotypes within the shaded area are part of the crown group (ALST) previously defined by Wallner *et al.* (2017). Haplogroup (HG) A is shown in green, HGs L and S in pink and HG T in orange. Haplotypes detected in the appended dataset are underlined, and haplotypes defined in this study are bold. Datings of nodes (marked by asterisks) and standard deviation were calculated by assuming a generation time of 7 years, as described in Appendix S1. The network is rooted with the Przewalski's Horse. Detailed information on samples, variants and haplotypes is given in Fig. S2 and Tables S1 & S2.

**Table 1** An increased Watterson's Theta is observed after inclusion of Asian horses.

Clade	Watterson's $\theta$
Whole phylogeny	
Without Asians	$3.07 \times 10^{-6}$
With Asians	$5.17 \times 10^{-6}$
Only Asians	$5.37 \times 10^{-6}$
Crown group	
Without Asians	$1.52 \times 10^{-6}$
With Asians	$1.70 \times 10^{-6}$
Only Asians	$2.59 \times 10^{-6}$

defined ALST crown group (Wallner *et al.* 2017). Within the crown group, the Noriker and Haflinger samples fall into the A clade, represented by imported Arabian Horses and draft horse breeds. All four Franches-Montagnes lineages represented in our dataset are located in the T clade and form four separate clusters (see Table S1). Only a few new crown haplotypes were introduced by these samples (Fig. 1). The observation that most European breeds cluster closely together within the recently established polytomic crown group underlines the predominance of these haplotypes in modern horses. Additional samples from Northern Europe (three Shetland Ponies and two North Swedish Draft Horses) confirm the private lineages (N) of these breed groups.

Out of the 101 newly described variants, 61 were detected solely in Asian breeds, resulting in a total of eight private haplotypes for Asian horses (Fig. 1). Only seven of the 13 Asian samples fell into the crown group, out of which three introduced private haplotypes (C, Ta, Ao-m, shown in Fig. 1). The Asian crown haplotypes correspond to 'intensively bred breeds from the Asian continent' and agree with their breeding history. We found a haplotype of the Ao group in the Marwari Horse, i.e., a connection to the Arabian Horse. The observation of Tb haplotypes in the Akhal-Teke reflects the close relationship to the English Thoroughbred lines that all carry Tb haplotypes (Wallner *et al.* 2017).

The Asian horses from rural breeds significantly increase MSY diversity (Table 1) and tremendously deepen the horse Y phylogeny. Three Asian samples represent private haplotypes (Y, M and J, shown in Fig. 1), all branching off after the Northern European group. The remaining three Asian horses group far outside the previous MSY phylogeny and support a new haplogroup (HG) O defined by 35 variants (Figs 1 & S2). HG-O introduces a deep,  $5527 \pm 872$ -year-old split from the rest of the lineages.

We find it remarkable that the four sequenced Yakutian samples carry mainly private haplotypes from all over the phylogeny. We detected HG-O (Ob, Oc), a private intermediate (Y) and a recent crown (Ad-h2) haplotype shared with Coldblood Horses, like Noriker and South German Coldblood. This indicates an origin of the Yakutian horse

population from many sources, which was already evident from autosomal data (Librado *et al.* 2015).

Waves of migration and founder effects have led to a severe reduction of haplotype diversity in modern breeds. The high diversity observed in Asian samples suggests that they have been less subjected to this demographic force and, hence, have retained autochthonous variation. This finding enables inference of the male history of horses further back in time. We hypothesize that the distant Asian haplotypes (HG-O in Fig. 1) are remnants of a far more diverse early-domesticated horse population (Librado *et al.* 2017). Further, the polytomy leading to the crown group and the three autochthonous Northern European and Asian lines (I, J and Y) are dated at  $1062 \pm 289$  years before present. The star-like structure of haplotypes leading to these lines must overlap with a certain cultural expansion together with horses (e.g. Zerjal *et al.* 2003) and should be investigated in detail in future studies.

Technically, the MSY can be investigated to trace horse male migration routes to a similar level as is possible in humans. Because of the very recent expansion of male lineages in most modern horse breeds, more data on extensively managed breeds need to be included in future studies to obtain more detailed insights into their male origin. Analysis of fossil data will provide a clearer picture of the early stages of domestication and the exact timing and location and how modern breeds are linked to their wild ancestors.

## Data submission

Data from the previous study can be obtained from SRA archive PRJNA353951. Mapped NGS reads of all new samples in this paper have been submitted to SRA archive PRJNA400560.

## Acknowledgements

Funding was provided by the Innovation Fund of the Austrian Academy of Sciences (ÖAW) and the Austrian Federal Ministry of Agriculture, Forestry, Environment and Water Management (DAFNE, 101184).

## Competing financial interests

The authors declare no competing financial interests.

## Ethics statement

The study was approved by the Ethics Committee of the University of Veterinary Medicine Vienna.

## References

Achilli A., Olivieri P., Soares H. *et al.* (2012) Mitochondrial genomes from modern horses reveal the major haplogroups that

- underwent domestication. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 2449–54.
- Kelekna P. (2009). *The Horse in Human History*. Cambridge University Press, Cambridge, UK.
- Librado P., Der Sarkissian C., Ermini L. *et al.* (2015) Tracking the origins of Yakutian horses and the genetic basis for their fast adaptation to subarctic environments. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 6889–97.
- Librado P, Gamba C., Gaunitz C. *et al.* (2017) Ancient genomic changes associated with domestication of the horse. *Science* **445**, 442–5.
- Lindgren G., Backström N. & Swinburne J. (2004) Limited number of patrilineages in horse domestication. *Nature Genetics* **36**, 335–6.
- Lippold S., Knapp M., Kuznetsova T. *et al.* (2011) Discovery of lost diversity of paternal horse lineages using ancient DNA. *Nature Communications* **2**, 450.
- Outram A, Stear N., Bendrey R., Olsen S., Kasparov A., Zaibert V., Thorpe N. & Evershed R.P. (2009) The earliest horse harnessing and milking. *Science* **323**, 1332–5.
- Petersen J., Mickelson J., Rendahl A. *et al.* (2013) Genome-wide analysis reveals selection for important traits in domestic horse breeds. *PLoS Genetics* **9**, e1003211.
- Vilà C., Leonard J., Götherström A., Markland S., Sandberg K., Lidén K., Wayne R.K. & Ellegren H. (2001) Widespread origins of domestic horse lineages. *Science* **291**, 474–7.
- Wallner B., Vogl C., Shukla P., Burgstaller J.P., Druml T., Brem G (2013) Identification of genetic variation on the horse Y chromosome and the tracing of male founder lineages in modern breeds. *PLoS One* **8**, 0060015.
- Wallner B., Palmieri N., Vogl C. *et al.* (2017) Y chromosome uncovers the recent oriental origin of modern stallions. *Current Biology* **10**, 2029–35.
- Warmuth V., Eriksson A., Bower M. *et al.* (2012) Reconstructing the origin and spread of horse domestication in the Eurasian steppe. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 8202–6.
- Zerjal T., Xue Y., Bertorelle G. *et al.* (2003) The genetic legacy of the Mongols. *American Journal of Human Genetics* **72**, 717–21.

### Supporting information

Additional supporting information may be found online in the supporting information tab for this article:

**Table S1.** Information on all samples analysed in this study.

**Table S2.** Information on all new variants identified.

**Appendix S1.** Supplemental methods.

**Figure S1.** A maximum parsimony tree with bootstrap values is shown. The tree is rooted with the donkey.

**Figure S2.** Haplotype network as shown in Fig. 1 but with readable variant IDs.