A Massively Parallel Sequencing Approach Uncovers Ancient Origins and High Genetic Variability of Endangered Przewalski's Horses

Hiroki Goto[1], Oliver A. Ryder[2], Allison R. Fisher[1], Bryant Schultz[1], Sergei L. Kosakovsky Pond[3], Anton Nekrutenko[4], and Kateryna D. Makova*[,1]

[1]Department of Biology, The Pennsylvania State University

[2]San Diego Zoo Institute for Conservation Research, San Diego Zoo Global, California

[3]Division of Infectious Diseases, Division of Biomedical Informatics, School of Medicine, University of California–San Diego

[4]Department of Biochemistry and Molecular Biology, The Pennsylvania State University

*Corresponding author: E-mail: kdm16@psu.edu.

Accepted: 30 June 2011

Abstract

The endangered Przewalski's horse is the closest relative of the domestic horse and is the only true wild horse species surviving today. The question of whether Przewalski's horse is the direct progenitor of domestic horse has been hotly debated. Studies of DNA diversity within Przewalski's horses have been sparse but are urgently needed to ensure their successful reintroduction to the wild. In an attempt to resolve the controversy surrounding the phylogenetic position and genetic diversity of Przewalski's horses, we used massively parallel sequencing technology to decipher the complete mitochondrial and partial nuclear genomes for all four surviving maternal lineages of Przewalski's horses. Unlike single-nucleotide polymorphism (SNP) typing usually affected by ascertainment bias, the present method is expected to be largely unbiased. Three mitochondrial haplotypes were discovered—two similar ones, haplotypes I/II, and one substantially divergent from the other two, haplotype III. Haplotypes I/II versus III did not cluster together on a phylogenetic tree, rejecting the monophyly of Przewalski's horse maternal lineages, and were estimated to split 0.117–0.186 Ma, significantly preceding horse domestication. In the phylogeny based on autosomal sequences, Przewalski's horses formed a monophyletic clade, separate from the Thoroughbred domestic horse lineage. Our results suggest that Przewalski's horses have ancient origins and are not the direct progenitors of domestic horses. The analysis of the vast amount of sequence data presented here suggests that Przewalski's and domestic horse lineages diverged at least 0.117 Ma but since then have retained ancestral genetic polymorphism and/or experienced gene flow.

Key words: wild horse, next-generation sequencing, mitochondrial DNA, nuclear DNA, phylogeny.

Introduction

Understanding the genetic relationship between domestic and Przewalski's horses is critical for unraveling the domestication history of the former and for formulating conservation and breeding strategies for the latter. Indeed, the endangered Przewalski's horse (*Equus przewalskii*) is the only wild horse living at present time and is the closest extant relative of the domestic horse (*Equus caballus*). Previously inhabiting an extensive range of steppe in both Asia and Europe, Przewalski's horse had become virtually extinct in the wild due to human activity by the middle of the 1960s; however, it had subsequently been bred in captivity and reintroduced to the wild (Ryder and Wedemeyer 1982; Ryder 1993; Bouman and Bouman 1994). The present-day Przewalski's horse population, consisting of over 2,000 animals, originated from 12 founders captured at the turn of the 19th century, a mare captured in 1957 and her descendents, some of which were hybrids with domestic horses. Only four Przewalski's horse matrilines, potentially identifiable as discrete mitochondrial lineages, currently survive (Volf et al. 1991; Oakenfull and Ryder 1998). Przewalski's horse is phenotypically distinct from the domestic horse in having shorter stature and more robust build than the former horse (Sasaki et al. 1999), although there is substantial variation

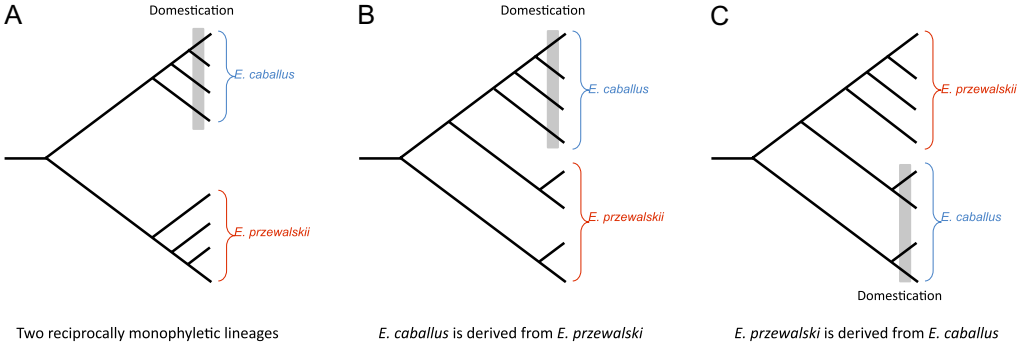

Fig. 1.—Hypothetical scenarios of divergence between *Equus caballus* and *Equus przewalskii*. Gray box indicates horse domestication. (*A*) *E. caballus* and *E. przewalskii* form two reciprocally monophyletic lineages. The two sister species diverged from a common ancestor. (*B*) *E. caballus* is derived from *E. przewalskii*. *E. przewalskii* is the direct ancestor of *E. caballus*. (*C*) *E. przewalskii* is derived from *E. caballus*. *E. caballus* is the direct ancestor of *E. przewalskii*.

among domestic horse breeds. The karyotype of Przewalski's horse ($2n = 66$) differs from that of the domestic horse ($2n = 64$) by a Robertsonian translocation (Benirschke et al. 1965; Bowling and Ruvinsky 2000; Myka et al. 2003; Yang et al. 2003; Ahrens and Stranzinger 2005). Despite these differences, interbreeding between Przewalski's and domestic horses produces fertile offspring (Short et al. 1974).

Previous studies presented contradictory conclusions regarding whether domestic and Przewalski's horses formed monophyletic genetic clades. Although some protein, microsatellite, and Y chromosome analyses have supported phylogenetic separation of the two taxa (Bowling et al. 2003; Wallner et al. 2003, fig. 1*A*), the latest investigations of autosomal (Lau et al. 2009; Wade et al. 2009) and X chromosomal DNA (Lau et al. 2009) have not. For instance, our previous phylogenetic analysis of several autosomal and X chromosomal introns placed Przewalski's horses within the domestic horse clade (Lau et al. 2009). Similarly, phylogenetic separation between Przewalski's versus domestic horses was not observed after typing ~1,000 autosomal single-nucleotide polymorphisms (SNPs) as part of the analysis of the horse genome (Wade et al. 2009). This led to the speculation that the *E. przewalskii* lineage could either be very recently derived from one or more *E. caballus* lineages (fig. 1*C*) or that the two horses intermixed to a limited degree following divergence from a common ancestor (Wade et al. 2009). Only relatively short genomic regions (or a small number of sites) have been analyzed in the reports mentioned above. Studies of mitochondrial DNA (mtDNA) of Przewalski's horse have so far been limited to sequencing the control region (Ishida et al. 1995; Oakenfull and Ryder 1998; Oakenfull et al. 2000).

Because Przewalski's horses are the only truly wild horses existing today, they have been hypothesized to be the direct ancestors of domestic horses (Mohr 1959; Ryder 1994, fig. 1*B*). The history of horse domestication has been investigated largely from mitochondrial and Y chromosome sequences, leading to suggestions of a limited number of patrilines (Lindgren et al. 2004), but numerous matrilines (Vilà et al. 2001) incorporated into the genetic pool of domestic horses. Multiple domestication events have been suggested to occur (Vilà et al. 2001; Jansen et al. 2002). Although genetic studies have yet to identify when and where horse domestication first took place, a recent archaeozoological report indicated the presence of domesticated horses ~5,500 years ago in Kazakhstan (Outram et al. 2009).

Determining the genetic relationship and divergence time between Przewalski's versus domestic horses is expected to inform conservation efforts aimed at preserving the genetic diversity of Przewalski's horses. An accurate picture of genetic diversity is particularly important for endangered species such as Przewalski's horse, for which the number of founding individuals was small, and thus, the effect of inbreeding has been a constant concern. Inbreeding reduces genetic diversity, causes high mortality, and short life span (Ralls and Ballou 1983) and has affected a captive population of Przewalski's horse (Bouman and Bos 1979). Despite efforts to minimize inbreeding of Przewalski's horses and to maintain their current genetic variation (Ryder et al. 1984; Princée et al. 1990; Zimmermann 1997), crucial studies of their genetic diversity have so far been limited. Przewalski's horse genetic diversity has been estimated utilizing blood group and allozyme loci (Bowling and Ryder 1987), mtDNA (Ryder 1994), and single strand conformation polymorphism analysis for major histocompatibility complex genes (Hedrick et al. 1999), leading to substantially disparate values. None of the previous investigations assessed genetic variation of Przewalski's horses from DNA data on a genome-wide scale.

In this study, we estimate the genetic diversity of Przewalski's horses and elucidate their phylogenetic relationship with domestic horses. Using massively parallel sequencing technology, we obtained complete mitochondrial and partial nuclear genomes for four Przewalski's horse individuals

representing all four surviving mitochondrial lineages. We constructed phylogenetic trees and assessed nucleotide diversity of Przewalski's horses based on mitochondrial, autosomal, and X chromosomal data separately. Based on these results, we discuss the genetic relationship of Przewalski's versus domestic horses. In particular, we address whether Przewalski's horses and domestic horses represent two distinct evolutionary genes pools in the diversity of horses. This study is valuable for guiding Przewalski's horse conservation efforts and illuminates the history of horse domestication.

Unlike SNP typing usually affected by ascertainment bias (reviewed in Nielsen 2004), the present method is expected to be largely unbiased. Indeed, most previous studies commenced with an SNP discovery protocol that by definition is limited to a subset of samples or populations. Once discovered, SNPs were typed in a larger sample, whereas variants unique to individuals not utilized in SNP discovery remained unassayed. This represents a serious problem that can now be tackled with the use of massively parallel sequencing technology (e.g., Luikart et al. 2003; Fridjonsson et al. 2011).

Materials and Methods

Horse Samples

Four Przewalski's horse individuals representing all four surviving mitochondrial lineages (Bowling and Ryder 1987) were analyzed: female Belina or OR383, studbook number 319, maternal lineage Staraja II; female Anushka or OR2661, studbook number 668, maternal lineage Bijsk/2; female Bonnette or OR1305, studbook number 339, maternal lineage Bijsk B; and male Bars or KB7674, studbook number 285, maternal lineage Orlica III. For Somali wild ass, we analyzed a female, sample OR3030. All DNA samples were kindly provided by the San Diego Zoo Safari Park; samples OR383, OR1305, and OR3030 came from animals resident in that park, whereas samples OR2661 and KB7674 came from animals originally housed in Minnesota Zoo and Tierpark Hellabrun (Munich), respectively.

DNA Extraction and Next Generation Sequencing

Genomic DNA was isolated using QIAGEN DNeasy Blood & Tissue kit. Paired-end sequences, either 35- or 100-bp reads (for Przewalski's horses and for Somali wild ass, respectively) separated by an ~600-bp interval, were generated with the Illumina/Solexa Genome Analyzer System II. The number of reads obtained per individual is listed in supplementary table S1 (Supplementary Material online). All reads generated in this study were deposited in sequencing trace archive (submission ID: DRA000429; study ID: DRP000437; sample IDs: DRS000782-DRS000786; experiment IDs: DRX000819-DRX000823; run IDs: DRR001222-DRR001226; analysis ID: DRZ000051).

Sequencing Read Mapping and Filtering

We used Burrows-Wheeler Aligner (Li and Durbin 2009) with default parameters and allowing no more than two mismatches to map full-length (untrimmed) paired-end reads to the horse nuclear genome (eca2) and its reference mtDNA (NC_001640). Reads were mapped against each individual chromosome separately, and reads mapping in discordance to their mate pair relationships were discarded (i.e., when both reads of a pair mapped to the same strand and/or the distance between the two mapped reads was greater than 2,000 bp). Only pairs of reads that mapped to unique positions in the horse genome were retained. Bases with Illumina sequencing quality score below 20 were eliminated after mapping.

For mtDNA, a consensus sequence was constructed for each sequenced individual (GenBank accession numbers AP012267–AP012271). Only sites supported by at least three (by at least two for Bars) uniquely mapped reads were used in the subsequent analysis. Sites with deletions, as compared with reference, were excluded. Sites possessing polymorphisms with frequency $\geq$0.28 (supported by at least two reads in Bars) were excluded (a total of 28 sites, see supplementary table S2, Supplementary Material online, for a summary of heteroplasmic sites). The most frequent base was taken as the consensus for the other polymorphic sites. For nuclear DNA, we only used sites supported by at least two uniquely mapped sequencing reads with an identical call, and all polymorphic sites were discarded.

Estimation of Divergence Time and Phylogenetic Reconstruction

For the mtDNA alignment, divergence time was estimated by Bayesian "relaxed molecular clock" approach implemented in BEAST (Drummond and Rambaut 2007). We fitted a Tamura–Nei sequence evolution model (Tamura and Nei 1993), assuming a relaxed molecular clock with uncorrelated rates sampled from the log-normal distribution and the Yule process tree prior. For the Bayesian analysis, we used Markov chain Monte Carlo sampling for 10 million generations (burn-in 1,000 generations). Because of excess transitions, unequal nucleotide frequencies, and variation of substitution rate among different sites for mtDNA, the Tamura–Nei model is thought to be the appropriate model for such data (Tamura and Nei 1993). Moreover, because the evolutionary distances among analyzed samples are very small, the choice of the model is not expected to influence the results substantially. Previously estimated divergence time between domestic horse and Somali wild ass of 2.0 Ma was used as a calibration point (Forstén 1992).

For the large-scale pairwise alignments of autosomes and chromosome X, we calculated maximum likelihood pairwise distances utilizing the Tamura–Nei sequence evolution model (Tamura and Nei 1993) and estimated their sampling errors using nonparametric bootstrap. The Tamura–Nei model was chosen because it has a simple closed form solution for estimating the distance from the matrix of pairwise difference counts (Tamura and Nei 1993). A custom utility in C was

developed to permit rapid estimation and bootstrapping from long pairwise alignments; source code is available for download from the HyPhy subversion code repository. Estimated distances were used then to build phylogenetic trees with the Neighbor-Joining method (Saitou and Nei 1987) as implemented in HyPhy (Kosakovsky Pond and Muse 2006). Bootstrap test (with 1,000 replicates) was performed to obtain the statistical support for each clade of the tree observed. The joint bootstrapping procedure constructed 1,000 pairwise distance matrices by independently sampling each distance from its corresponding bootstrap distribution. This procedure is expected to overestimate topological variance, when compared with the standard multiple sequence alignment method, and correspondingly lower conservative values for phylogenetic clade support.

To test whether the obtained phylogenetic trees have topologies significantly superior to alternative topologies, we utilized two approaches. For the mtDNA data, we used the Kishino–Hasegawa, Shimodaira–Hasegawa, weighted Kishino–Hasegawa, and weighted Shimodaira–Hasegawa tests (Shimodaira and Hasegawa 1999) as implemented in CONSEL (Shimodaira and Hasegawa 2001). For the nuclear DNA data, which did not possess the form of a multiple sequence alignment, but rather a series of largely nonoverlapping pairwise alignments, standard tests for phylogenetic support do not apply; hence, we adopted a simulation approach. In order to estimate the parameters of the nucleotide substitution process, we fitted the GTR model with site-to-site rate variation modeled by a 3-bin general discrete distribution (which is more flexible than the standard discretized gamma, Kosakovsky Pond and Frost 2005) to the 5-way alignment of horse DNA sequences (415,452 bases). We next simulated genomic alignments of four Przewalski's horse, the Thoroughbred horse, and the Somali wild ass sequences (depicted in fig. 3*A*) under a collection of alternative topologies and subsampled ($N = 100$ times) pairwise alignments of lengths equal to those obtained from real data and reran the neighbor-joining tree construction (NJ + TN93) procedure on simulated data. We asked the following questions using our simulations: a) assuming that the data are generated using the inferred tree (e.g., fig. 3*A* and *B*) what is the frequency at which it is recovered from pairwise alignments using NJ + TN93? and b) how often is the tree inferred from real data recovered by NJ + TN93 if the Thoroughbred lineage is placed at any alternative location in the Przewalksi horses clade?

Capillary Sequencing

To confirm the observed variation in four mtDNA lineages of Przewalski's horse, we analyzed eight additional Przewalski's horse samples: Basil, Henrietta, Hermonia (descendents of Staraja II), Bertland, Bonar, Nadiushka, Rolmar (descendents of Bijsk B), and Kuporovitch (descendent of Bijsk/2). We performed capillary sequencing for two mtDNA regions, corresponding to positions 2981–3647 and 15506–15860 in Xu and Arnason (Xu and Arnason 1994). These positions were selected because they contained multiple sites differentiating the haplotypes. Details are available in the Supplementary Material online.

Results

Sequencing Complete Mitochondrial Genomes from All Surviving Przewalski's Horse Matrilines

Using the paired-end module of an Illumina Genome Analyzer II, we sequenced genomic DNA (mtDNA and nuclear DNA), of four Przewalski's horses (three females: Belina, Anushka, and Bonnette, and one male: Bars) representing all four surviving mitochondrial lineages (Staraja II, Bijsk/2, Bijsk B, and Orlica III, respectively). Genomic DNA from a Somali wild ass (*Equus africanus somaliensis*), an outgroup (Beja-Pereira et al. 2004), was also sequenced. For four of the five sequenced individuals, DNA was isolated from the heart muscle, resulting in enrichment for mtDNA. Heart tissue was not available for Bars, and thus, his blood was utilized instead.

Although we sequenced complete mtDNA in all five individuals, our subsequent analysis was limited to the sequencing reads that were uniquely mapped to mtDNA, in order to exclude potential mtDNA insertions in the nuclear genome (*numts*; Bensasson et al. 2003). Mutations specific to individual lineages of Przewalski's horse were confirmed by Sanger sequencing. The final data set included 10,840 base pairs (bp) covered in all five individuals and aligned to the reference mtDNA sequence of the domestic horse (supplementary tables S1 and S3, Supplementary Material online). Mean mtDNA read coverage per site ranged from 60x to 260x for DNA isolated from heart and was 7x for DNA isolated from blood.

Two mtDNA sequences (heteroplasmic and indel sites were excluded, see Supplementary Material online) were identical (haplotype I, observed in Anushka and Belina), the third sequence (haplotype II, present in Bars) was very similar to haplotype I, whereas the fourth one (haplotype III, observed in Bonnette) was substantially different from both of them (supplementary table S3, Supplementary Material online). Partial sequencing of mtDNA from additional horses belonging to the four original matrilines confirmed these results (see Supplementary Material online). Haplotypes I and II differed by only two substitutions in 10,840 bp (both substitutions were located outside of the control region). In contrast, sequence divergence (corrected for multiple hits) between haplotype I (or II) and haplotype III was much higher, 59 ± 6.8 substitutions in 10,840 bp or $0.548 \pm 0.064\%$ (outside the control region, there were 47 ± 6.3 substitutions in 10,085 bp or $0.466 \pm 0.063\%$).

Phylogenetic Analysis of mtDNA Sequences

Bayesian phylogenetic analysis of Przewalski's and publicly available domestic horse mtDNA sequences (sequences

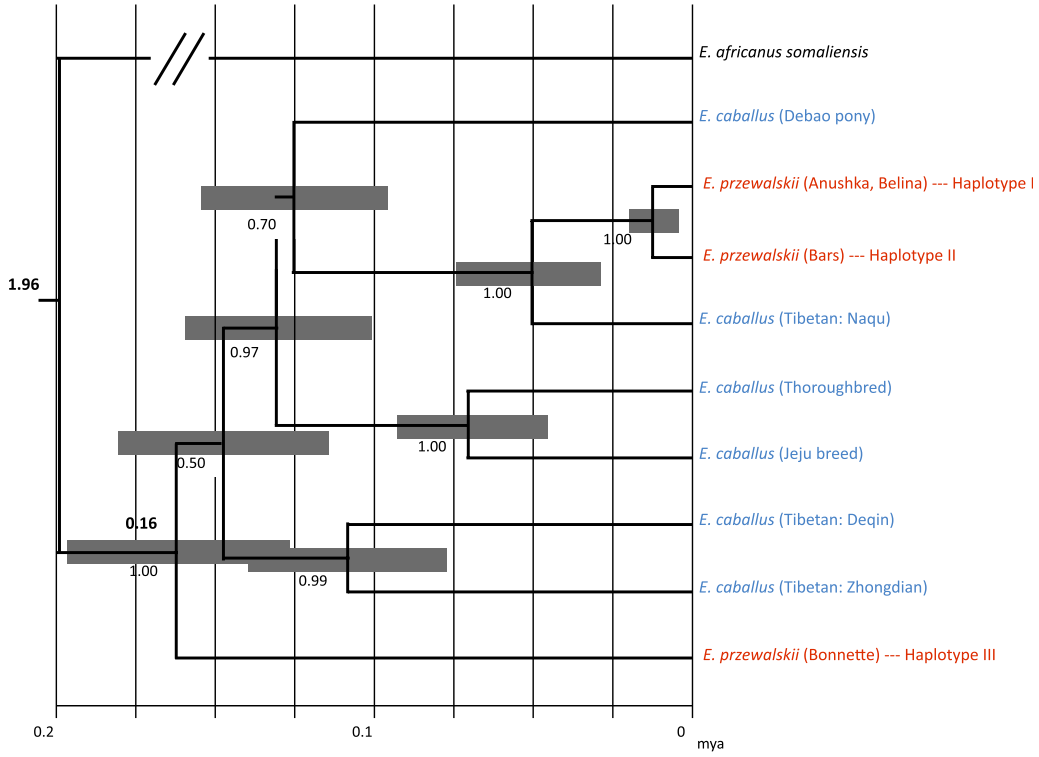

Fig. 2.—Phylogenetic tree and divergence time estimates for mtDNA sequences of Przewalski's and domestic horses. The results of analysis carried out in BEAST. Based on a fossil record, we assumed that Somali wild ass and domestic horse diverged 2.0 Ma (Forstén 1992). Numbers below nodes are posterior probabilities, and bold numbers above nodes are the estimated divergence times. The 95% highest posterior density estimates for each clade are represented by bars.

located outside the control region were utilized because they produced the most reliable alignment), with Somali wild ass as an outgroup, confirmed substantial differences between the Przewalski's horse haplotypes I/II versus III (fig. 2). Specifically, these haplotypes did not form a monophyletic clade but instead were separated on the phylogenetic tree: the mtDNA sequences of haplotypes I/II grouped with this from the Tibetan breed (from Naqu), whereas haplotype III was more basal; previous studies have shown that domestic horse mtDNA haplotypes are not breed specific (Vilà et al. 2001; Jansen et al. 2002). The observed topology (fig. 2) was significantly better than alternative topologies assuming that (1) domestic and Przewalski's horses form two monophyletic clades, respectively, (2) domestic horse is an ancestor to Przewalski's horse, or (3) Przewalski's horse is an ancestor to domestic horses (supplementary table S4, Supplementary Material online). Similar results were obtained when ~600 bp of the mtDNA control region of 37 domestic horse breeds were included in the phylogenetic analysis (supplementary fig. S1, Supplementary Material online). These observations rely on a substantially longer sequence and thus a more informative alignment than in previous studies (e.g., Jansen et al. 2002), implying that Przewalski's horses are not monophyletic for mtDNA.

Assuming that the divergence time between horse and Somali wild ass was 2 Ma (Forstén 1992), we estimated that haplotypes I/II and III separated from each other ~0.156 Ma (confidence interval [CI] is between 0.117 and 0.186 Ma). Therefore, these two Przewalski's horse mtDNA lineages have diverged from a common ancestral sequence anciently.

Diversity and Phylogenetic Analysis of Portions of the Nuclear Genome of Przewalski's Horses

Our approach fortuitously provided partial nuclear DNA sequences for Przewalski's horses and Somali wild ass (see Materials and Methods). The total number of analyzed bases ranged from 24.5 to 106.2 million for autosomes and from 0.8 to 5.6 million for the X chromosome, depending on individual (supplementary table S3, Supplementary Material online), a vast increase over previous studies. Nevertheless, due to inclusion of the nuclear genome from only one domestic horse breed (the Thoroughbred, the only domestic horse breed for which large-scale nuclear genome data are currently available) and due to low sequence coverage of the nuclear genomes of Przewalski's horses and Somali wild ass, caution should be exercised when interpreting these results.

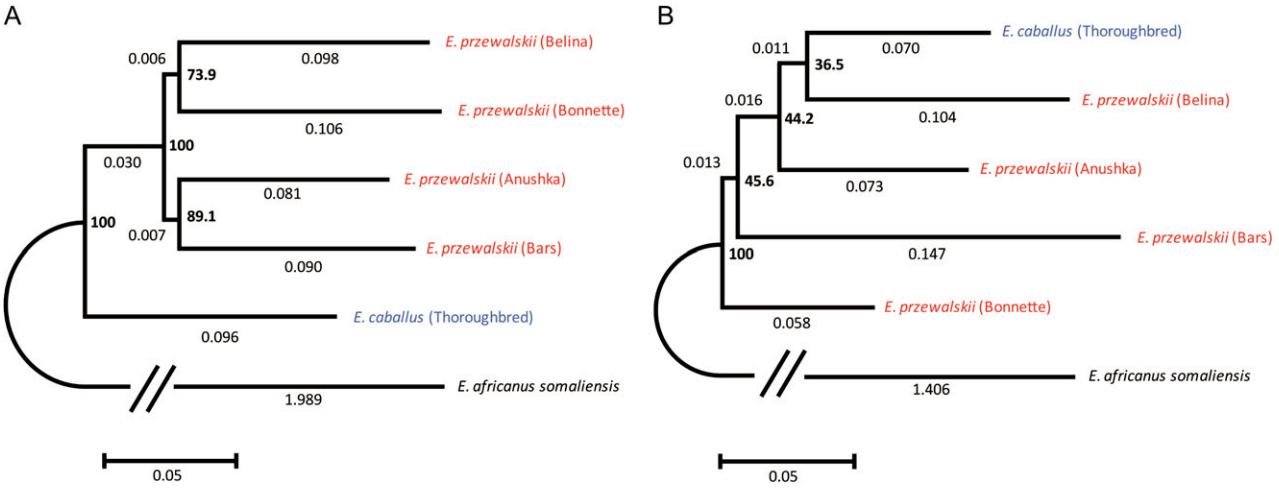

Fig. 3.—Neighbor-joining trees of autosomal sequences (*A*) and chromosome X sequences (*B*) based on pairwise genetic distances. Numbers at the nodes represent bootstrap support values and numbers on the branches indicate genetic distances. Somali wild ass (*Equus africanus somaliensis*) was used as an outgroup.

Utilizing these data, we built pairwise alignments for all possible combinations of six individual nuclear genomes (five genomes partially sequenced here plus the reference horse genome) and estimated the corresponding pairwise nucleotide genetic distances separately for chromosome X and autosomes based on the Tamura and Nei model of sequence evolution (Tamura and Nei 1993; supplementary table S3, Supplementary Material online). The distances between the Somali wild ass sequence and any of the horse sequences were substantially higher than the other comparisons, reaffirming that Somali wild ass is an adequate outgroup. The analysis below is based on pairwise alignments because few bases were expected to align in all five sequenced animals, given low nuclear genome coverage (in fact, using the proportions of genomes covered in the sequenced regions for each individual from supplementary table S1, Supplementary Material online, and using 2.5 Gb [Wade et al. 2009] as the size of the horse genome, we only expect ~2.5 bases to be shared among all five sequenced individuals).

The average pairwise autosomal genetic distance between the Thoroughbred domestic and Przewalski's horse was 0.226% (95% CI = 0.225–0.227%), slightly higher than 0.18% reported recently in a study analyzing a smaller data set (Wade et al. 2009). The average autosomal divergence between Somali wild ass and either the Thoroughbred domestic or Przewalski's horse was ~1.1%. Assuming strict molecular clock and 2 Ma divergence between horse and Somali wild ass (Forstén 1992), this corresponds to the rate of 2.75×10^{-9} substitutions per site per year. Using this rate, we computed the coalescence time between the Thoroughbred domestic horse and sequenced Przewalski's horses to be ~0.411 Ma (95% CI = 0.409–0.413). The average autosomal pairwise diversity among the four Przewalski's horses was 0.195% (95% CI = 0.189–0.199%; supplementary table S3, Supplementary Material online), suggesting a coalescence time of 0.353 Ma (95% CI = 0.344–0.362), an even more ancient origin than estimated from shorter mtDNA sequences.

To investigate divergence in the nuclear genome between domestic and Przewalski's horses, we constructed Neighbor-Joining (Saitou and Nei 1987) phylogenies from pairwise distances, separately for autosomes and the X chromosome (fig. 3). Note that these phylogenetic trees need to be corroborated in future studies including additional domestic horse breeds. Nevertheless, from our results, in contrast to the mtDNA tree indicating two distinct Przewalski's haplotype groups that were intermingled with domestic horse haplotypes (fig. 2), on the autosomal tree Przewalski's horses formed a monophyletic clade (fig. 3*A*). The parametric simulation test described in the Materials and Methods indicated that 1) the data generated under the tree in figure 3*A* led to the recovery of the correct tree in 100% of cases using NJ + TN93 methodology and 2) the probability of inferring the tree in figure 3*A* given any other alternative placement of Przewalski's horses (i.e., where these horses are not monophyletic) was <0.01 (based on 100 replicates). These results suggest that the low level of pairwise divergence and very long sequences allow NJ + TN93 to accurately recover the underlying topology even when only pairwise alignments are available.

The average X chromosome divergence between Somali wild ass and either the Thoroughbred domestic horse or Przewalski's horses was ~0.817% (supplementary table S3, Supplementary Material online), lower than for autosomes, in agreement with the phenomenon of male mutation bias (Makova and Li 2002) and a smaller effective population size for the X chromosome. Assuming molecular clock and 2 Ma divergence between horse and Somali wild

ass (Forstén 1992), this corresponds to the X chromosomal rate of 2.04×10^{-9} substitutions per site per year. The X chromosomal distances between horses were also lower than the autosomal distances in most cases (supplementary table S3, Supplementary Material online). The X chromosomal pairwise distances including data from the only male used in this analysis, Bars, were unusually high, potentially due to the mapping of some Y chromosome sequences to the domestic horse X chromosome. The average nucleotide diversity among Przewalski's horse X chromosomes was 0.211% (0.180% excluding Bars). The average pairwise distance between the X chromosome sequences of Przewalski's horse Bonnette versus other horses (this is the deepest root among horse sequences) was 0.186% (95% CI = 0.164–0.210%) corresponding to coalescence time of 0.455 Ma (95% CI = 0.402–0.514 Ma). Excluding Bars, this value was 0.179% (95% CI = 0.157–0.202%), corresponding to coalescence time of 0.439 Ma (95% CI = 0.385–0.495 Ma), largely in agreement with our autosomal results. Thus, all three types of data—autosomal, X chromosomal, and mitochondrial—point toward ancient genetic origins of Przewalski's horses sequenced here.

On the X chromosomal tree, the Thoroughbred domestic horse sequence was intermingled with Przewalski's horse sequences, and no clustering was significant (fig. 3*B*). Parametric simulations of topological signal showed that the correct tree is recovered with 100% accuracy, and the probability of inferring the tree in figure 3*B* given any other alternative placement of the Thoroughbred lineage is <0.01 (based on 100 simulations).

Discussion

Evolution of Przewalski's and Domestic Horses

According to our data, Przewalski's horse is not the direct progenitor of the domestic horse. If domestic horses had been derived from the Przewalski's horse, then domestic horse sequences would have been embedded within the Przewalski's horse phylogenetic clade (fig. 1*B*; Ryder 1994). This expectation is contradicted by the analysis of mtDNA data. Indeed, mtDNA haplotypes of Przewalski's horses are intermingled with domestic horse sequences (fig. 2, also see below). Additionally, Przewalski's horse autosomal sequences form a separate monophyletic clade excluding the Thoroughbred domestic horse (fig. 3*A*), although this result will have to be reevaluated when nuclear sequences of additional domestic horses become available.

The hypothesis of a recent origin of Przewalski's horses from domestic horses (Oakenfull and Ryder 1998; Wade et al. 2009), according to which Przewalski's sequences are expected to be embedded within the domestic horse phylogenetic clade (fig. 1*C*), is also contradicted by our data because our analysis placed one of the Przewalski's horse haplotypes at the deepest branching point among currently available complete mtDNA caballine sequences (fig. 2) and at one of the deepest branching points among currently available mtDNA control region sequences (supplementary fig. S1, Supplementary Material online). Autosomal sequences of additional domestic horse breeds are needed to test the origin of Przewalski's horses from domestic horses more explicitly.

Do Przewalski's and domestic horses represent two distinct evolutionary gene pools in the diversity of horses? Several recent studies based on the analyses of the mtDNA control region (Oakenfull and Ryder 1998; Ishida et al. 1995; Vilà et al. 2001; Jansen et al. 2002; Kim et al. 1999) and autosomal SNPs (Wade et al. 2009) failed to separate Przewalski's and domestic horse sequences in molecular phylogenies. The reports demonstrating genetic differentiation between Przewalski's and domestic horses were based on either nuclear-encoded data sets (e.g., Bowling et al. 2003) or Y chromosomal DNA (Wallner et al. 2003). In the latter study, the split between the two lineages was estimated to occur 0.123–0.241 Ma, a date significantly preceding horse domestication, and the karyotype present in domestic horses was suggested to represent the derived condition. Recently, based on a well-resolved phylogeny of the Perissodactyla, Steiner and Ryder (Steiner CC, Ryder OA, submitted) asserted that the ancestral state of the Robertsonian translocation between *E. caballus* and *E. przewalskii* was the unfused elements and corresponding higher diploid number currently present in *E. przewalskii*, in contrast with the results of Myka et al. (2003) and of Trifonov et al. (2008) that relied upon an alternate phylogeny of *Equus*.

It has been suggested that in Przewalski's horses, the genealogy of nuclear DNA might be different from the genealogy of mtDNA (Ishida et al. 1995), as was observed for African elephants (Roca et al. 2005). Our results indeed indicate the presence of just such differences leading in different answers to the question of separation of genetic pool between domestic and Przewalski's horses. The mtDNA analysis resulted in intermingling between Przewalski's and Thoroughbred horse sequences, with one of Przewalski's horse haplotypes located at the most basal position among the available complete mtDNA caballine sequences. In contrast, all Przewalski's horses formed a separate clade on the autosomal tree with high bootstrap support. The latter conclusion results from the unbiased analyses of SNP variation using distance methods. Although not definitive, because a large number of domestic horses have not been investigated, the results obtained using nuclear data are consistent with a scenario that is quite distinct from the view that emerges from analysis of mtDNA variation.

Different topologies for horse mitochondrial versus nuclear DNA observed here might be explained by distinct evolutionary histories of horse matrilines versus patrilines. Genetic introgression between domestic and Przewalski's horses may have been largely female mediated, in agreement with the hypothesis proposed by Wallner et al. (2003). This explanation

would be consistent with the nonmonophyletic placement of Przewalski's horse sequences in the mtDNA tree (exclusively maternal) as well as their intermingling in the X chromosomal tree (predominantly maternal). Male-mediated admixture between Przewalski's and domestic horses may have been limited. Indeed, the male-specific Y chromosome contains fixed differences separating Przewalski's and domestic horses (Wallner et al. 2003). Also, the present analysis of Przewalski's horse autosomal sequences that have a substantial paternal contribution groups them in a monophyletic clade. The dichotomous findings obtained by us for autosomal versus mtDNA data can potentially be resolved in the future by investigations of ancient DNA and greater analysis of sequence variation in domestic horses, especially utilizing methods developed in the present study (that do not introduce ascertainment bias from the individuals used for SNP discovery; reviewed in Nielsen 2004).

Divergent mtDNA Haplotypes among Przewalski's Horses

The detailed analysis of complete mitochondrial genomes from all four surviving maternal lineages of Przewalski's horses indicated their ancient nonmonophyletic origins. We identified three mtDNA haplotypes; two haplotypes (I and II) were very similar to each other, whereas the third one (III) was markedly distinct from the other two haplotypes. Why are haplotypes I/II so divergent from haplotype III, and why do not the three haplotypes form a monophyletic clade, even though the horses harboring them went through a severe genetic bottleneck (Volf et al. 1991) and do not exhibit morphological variation?

First, the observed haplotypes could exemplify the genetic polymorphism present in the ancestral horse population that existed prior to the divergence of Przewalski's and other major modern horse lineages (Jansen et al. 2002). The deep phylogenetic separation between haplotypes I/II and III could represent the natural variation within a single species, driven by an early maternal lineage split. We cannot exclude the possibility that selection or geographic isolation contributed to the separation of the haplotypes. Regardless of the mechanism, Przewalski's horses (as well as possibly domestic horses) could have retained such ancestral variation in their current population. This scenario would interfere with inferring the monophyletic origins of Przewalski's horses.

Second, some haplotypes could have been introduced from domestic into Przewalski's horses via interbreeding. In particular, haplotypes I/II are very similar in sequence to and cluster together with domestic horse haplotypes and, thus, might have been acquired by Przewalski's horse through introgressive hybridization. Przewalski's horse haplotype III might represent the "true" Przewalski's horse mtDNA. However, additional sequencing of mtDNA from modern domestic horses from Eurasia may identify mtDNA haplotypes similar in sequence to Przewalski's horse haplotype III. A combination of these two scenarios—some haplotypes acquired via introgression and other haplotypes inherited from the ancestral horse population—is also possible.

The presence of highly divergent mtDNA haplotypes is unexpected for Przewalski's horses because they have gone through a genetic bottleneck. However, this is not unprecedented because some other mammalian species also exhibit distinct mtDNA haplotypes within their continuous populations, for example, moose (Hundertmark et al. 2002), reindeer (Flagstad and Røed 2003), elephant (Fleischer et al. 2001), and mammoth (Gilbert et al. 2008).

A Model of Divergence between Przewalski's and Domestic Horses

The analyses of all three types of genomic data (mtDNA, autosomal, and X chromosomal) indicate that Przewalski's and domestic horse lineages diverged significantly preceding horse domestication, thought to have occurred ~5,000–6,000 years ago (Outram et al. 2009). In fact, mtDNA haplotypes of Przewalski's horses coalesce 0.117–0.187 Ma, that is, at least a hundred thousand years prior to horse domestication. Moreover, Przewalski's horse autosomal sequences, as well as X chromosomal sequences, coalesce several hundred thousand years preceding horse domestication. These observations are at variance with the hypothesis that Przewalski's horse population represents the wild stock from which the domestic horses were bred, even though our results suggest a close genetic relationship between mtDNA haplotypes of some Przewalski's and domestic horses (see discussion below). Note that, if both groups of horses retained substantial levels of ancestral polymorphism, this would interfere with our estimates of the divergence of their lineages. Nevertheless, the drastic difference between horse domestication time (~5,000–6,000 years ago) and our coalescent estimates (at least 117,000 years ago) is unlikely to be the result of retained ancestral polymorphism alone.

From our phylogenetic analysis (see above), we concluded that domestic horse is neither derived from Przewalski's horse nor the opposite. We propose a model according to which Przewalski's horse and domestic horse are descendents of two lineages that diverged potentially as early as ~0.150 Ma. This is consistent with our mtDNA and nuclear analyses as well as with the published Y chromosomal results (Wallner et al. 2003). Indeed, a monophyletic grouping of Przewalski's horses on an autosomal tree is not anticipated in a recently diverged genetic pool containing substantial shared genetic variation. Nevertheless, the initial divergence event between the two lineages could have been followed by the retention of ancestral polymorphism and/or introgressive hybridization. The signatures of these events are particularly conspicuous in the mtDNA data due to the absence of recombination.

The monophyly of Przewalski's horse nuclear sequences contradicts the recent findings of Wade et al. (2009) who

suggested that only few Przewalski's horse-specific mutations are absent from the domestic horse population. Why do the results of these two studies differ? First, Wade and colleagues used a smaller data set that was based on SNPs derived from domestic horses and thus was prone to ascertainment bias (reviewed in Nielsen 2004). Second, the monophyly of Przewalski's horse sequences presented here might be in part influenced by sparse sampling of autosomal domestic horse genomes; future studies will have to evaluate this possibility. Note that three of the Przewalski's horses sequenced here (Bars, Belina, and Bonette) are thought to have no recent (since the bottleneck) domestic horse genetic contribution to their known pedigrees, whereas 16% of Anushka's DNA can be traced back to a Mongolian domestic mare (Ballou 1994). Nevertheless, Anushka's autosomal sequences formed a clade with those from Bars and did not have a greater genetic distance to Przewalski's horses without recent domestic horse contribution. These results suggest that a more ancient gene flow might have occurred between ancestral populations of Przewalski's wild horse and Asian domestic horse breeds, especially since the past distribution of Przewalski's horse overlaps with the present-day Mongolian horse distribution (Ishida et al. 1995) and corroborate our findings based on mtDNA data.

Relatively High Nucleotide Diversity among Przewalski's Horses

The high average nucleotide diversity observed here in the nuclear and mtDNA of Przewalski's horses was unexpected, given that their population had dwindled to a mere dozen individuals only 40 years ago (Volf et al. 1991) and was subsequently subject to inbreeding. Mean autosomal diversity in Przewalski's horses (0.195%; this study) was higher than that in several breeds of domestic horses studied by us previously (0.1%; Lau et al. 2009) as well as in the sequenced Thoroughbred horse (0.05%; Wade et al. 2009). Average X chromosomal diversity was similarly higher for Przewalski's than for domestic horses (0.182% estimated here versus 0.1% in domestic horses as estimated by us; Lau et al. 2009). Note that our estimates of nucleotide diversity from nuclear DNA data are likely deflated because we required each analyzed site to be supported by two or more identical sequencing reads. For mtDNA, Przewalski's horse nucleotide diversity was also relatively high. When estimated from total mtDNA, the nucleotide diversity between divergent haplotypes (0.54%; this study) was comparable to that estimated for Tibetan horse breeds (0.66%; Xu et al. 2007). When estimated from the mtDNA control region, the nucleotide diversity between divergent haplotypes (1.6%; this study) was only slightly lower than that observed in natural populations of moose (2.5%; Hundertmark et al. 2002) and Asian elephant (1.8%; Roca et al. 2005).

What can explain the relatively high genetic diversity in the modern population of Przewalski's horses, despite a recent and severe genetic bottleneck? First, the Przewalski's horse population that existed prior to the bottleneck might have possessed substantial levels of genetic diversity and some of this diversity may have carried across. Second, interbreeding with domestic horses, known to have occurred for at least some Przewalski's horses after the bottleneck (Volf et al. 1991), and also, potentially, more anciently (see above), might have elevated genetic diversity in Przewalski's horses. One or both of these factors have likely counterbalanced the effects of inbreeding.

Implications for Breeding Strategies

Our results have direct implications for the strategies of breeding Przewalski's horses, hundreds of which are kept in captivity and are being released to the wild via reintroduction programs. The major goal of managing the Przewalski's horse population is to maintain it in sufficient size and genetic diversity to protect the species from extinction. To achieve this goal, a careful analysis of past, current, and future genetic characteristics of the population is required (Ballou 1994). From this perspective, first, our results demonstrate the existence of two highly divergent mitochondrial haplogroups in Przewalski's horses. It is imperative to ensure the survival of both these haplotypes in growing populations of Przewalski's horses because all reintroduction projects should include representation of the entire species gene pool (Ryder 1993). Second, our analysis points toward substantial genetic diversity persisting in the current population and likely present in the founders of the surviving population. Nevertheless, inbreeding should be kept at a minimum to preserve this genetic diversity. Third, and albeit indirectly, our results suggest ancient introgression of domestic horse genes into Przewalski's horse genes. This questions the need to separate "pure" Przewalski's horses (e.g., the Munich line) from "non-pure" Przewalski's horses (i.e. with known domestic horse contributions) because even the former lineage likely experienced some, perhaps more ancient, admixture with domestic horses. Moreover, interbreeding with domestic horses might have elevated the nucleotide diversity of Przewalski's horses.

To further illuminate the natural history of Przewalski's horses, it will be necessary to investigate mtDNA haplotypes as well as nuclear DNA from preserved skin specimens present in museum collections around the world. This would allow one to evaluate the genetic diversity of Przewalski's horses prior to the bottleneck and, based on collection sites, to correlate this diversity with the geographic distribution. Additionally, the complete sequencing of the Przewalski's horse genome (or exome) and its detailed comparison with the domestic horse genome is expected to facilitate the discovery of genotypic differences of phenotypic consequences distinguishing these two closely related species.

Supplementary Material

Supplementary tables S1–S4, figure S1 and other supplementary materials are available at *Genome Biology and Evolution* online.

Acknowledgments

We thank Wen-Yu Chung, Benjamin Dickins, and Guruprasad Ananda for assistance with computational analyses; Cynthia Steiner, Melissa Wilson Sayres, Hie Lim Kim, Masafumi Nozawa, Chungoo Park, and Lydia Krasilnikova for helpful comments; and Leona Chemnick for her assistance in sample handling and preparation. This work is supported by start-up funds from the Eberly College of Science at The Pennsylvania State University to K.D.M., by National Science Foundation and National Institutes of Health grants to A.N., and by National Institutes of Health grants to S.K.P. Additional funding is provided, in part, under a grant with the Pennsylvania Department of Health using Tobacco Settlement Funds. The Department specifically disclaims responsibility for any analyses, interpretations, or conclusions. Additional support from the Sonny Foundation and the John and Beverly Stauffer Foundation is gratefully acknowledged.

Literature Cited

Ahrens E, Stranzinger G. 2005. Comparative chromosomal studies of *E. caballus* (ECA) and *E. przewalskii* (EPR) in a female F1 hybrid. J Anim Breed Genet. 122(Suppl 1):97–102.

Ballou JD. 1994. Population biology. In: Houpt KA, editor. Przewalski's horse: the history and biology of an endangered species. Albany (NY): The State University of New York Press. p. 93–113.

Beja-Pereira A, et al. 2004. African origins of the domestic donkey. Science 304:1781.

Benirschke K, Malouf N, Low RJ, Heck H. 1965. Chromosome complement: differences between *Equus caballus* and *Equus przewalskii*, poliakoff. Science 148:382–383.

Bensasson D, Feldman MW, Petrov DA. 2003. Rates of DNA duplication and mitochondrial DNA insertion in the human genome. J Mol Evol. 57:343–354.

Bouman I, Bouman J. 1994. The history of the Przewalski's horse. In: Boyd L, Houpt KA, editors. Przewalski's horse: the history and biology of an endangered species. Albany (NY): The State University of New York Press. p. 5–38.

Bouman JG, Bos H. 1979. Two symptoms of inbreeding depression in Przewalski horses living in captivity. In: de Boer LEM, Bouman J, Bouman I, editors. Genetics and hereditary diseases of the Przewalski horse. Rotterdam (The Netherlands): Foundation for the Preservation and Protection of the Przewalski horse. p. 165–168.

Bowling AT, Ruvinsky A. 2000. Genetic aspects of domestication, breeds, and their origins. In: Bowling AT, Ruvinsky A, editors. The genetics of the horse. Wallingford (UK): CABI Publishing. p. 25–51.

Bowling AT, Ryder OA. 1987. Genetic studies of blood markers in Przewalski's horses. J Hered. 78:75–80.

Bowling AT, et al. 2003. Genetic variation in Przewalski's horses, with special focus on the last wild caught mare, 231 Orlitza III. Cytogenet Genome Res. 101:226–234.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 7:214.

Flagstad O, Røed KH. 2003. Refugial origins of reindeer (Rangifer tarandus L.) inferred from mitochondrial DNA sequences. Evolution 57:658–670.

Fleischer RC, Perry EA, Muralidharan K, Stevens EE, Wemmer CM. 2001. Phylogeography of the asian elephant (Elephas maximus) based on mitochondrial DNA. Evolution 55:1882–1892.

Forstén A. 1992. Mitochondrial-DNA time-table and the evolution of *Equus*: comparison of molecular and paleontological evidence. Ann Zool Fennici. 28:301–309.

Fridjonsson O, et al. 2011. Detection and mapping of mtDNA SNPs in Atlantic salmon using high throughput DNA sequencing. BMC Genomics 12:179.

Gilbert MT, et al. 2008. Whole-genome shotgun sequencing of mitochondria from ancient hair shafts. Science 317:1927–1930.

Hedrick PW, Parker KM, Miller EL, Miller PS. 1999. Major histocompatibility complex variation in the endangered Przewalski's horse. Genetics 152:1701–1710.

Hundertmark KJ, et al. 2002. Mitochondrial phylogeography of moose (Alces alces): late pleistocene divergence and population expansion. Mol Phylogenet Evol. 22:375–387.

Ishida N, Oyunsuren T, Mashima S, Mukoyama H, Saitou N. 1995. Mitochondrial DNA sequences of various species of the genus Equus with special reference to the phylogenetic relationship between Przewalskii's wild horse and domestic horse. J Mol Evol. 41:180–188.

Jansen T, et al. 2002. Mitochondrial DNA and the origins of the domestic horse. Proc Natl Acad Sci U S A. 99:10905–10910.

Kim KI, et al. 1999. Phylogenetic relationships of Cheju horses to other horse breeds as determined by mtDNA D-loop sequence polymorphism. Anim Genet. 30:102–108.

Kosakovsky Pond SL, Frost SD. 2005. A simple hierarchical approach to modeling distributions of substitution rates. Mol Biol Evol. 22:223–34.

Kosakovsky Pond SL, Muse SV. 2006. HyPhy: hypothesis testing using phylogenies. Bioinformatics 21:676–679.

Lau AN, et al. 2009. Horse domestication and conservation genetics of Przewalski's horse inferred from sex chromosomal and autosomal sequences. Mol Biol Evol. 26:199–208.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.

Lindgren G, et al. 2004. Limited number of patrilines in horse domestication. Nat Genet. 36:335–336.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P. 2003. The power and promise of population genomics: from genotyping to genome typing. Nat Rev Genet. 4(12):981–94.

Makova KD, Li WH. 2002. Strong male-driven evolution of DNA sequences in humans and apes. Nature 416:624–626.

Mohr E, 1959. Das Urwildpferd. Die Neue Brehm-Bücherei. Wittenberg (Lutherstadt): A. Ziemsen Verlag, p. 144.

Myka JL, Lear TL, Houck ML, Ryder OA, Bailey E. 2003. FISH analysis comparing genome organization in the domestic horse (*Equus caballus*) to that of the Mongolian wild horse (*E. przewalskii*). Cytogenet Genome Res. 102:222–225.

Nielsen R. 2004. Population genetic analysis of ascertained SNP data. Hum Genomics. 1(3):218–224.

Oakenfull EA, Lim HN, Ryder OA. 2000. A survey of equid mitochondrial DNA: implications for the evolution, genetic diversity, and conservation of *Equus*. Conserv Genet. 1:341–355.

Oakenfull EA, Ryder OA. 1998. Mitochondrial control region and 12S rRNA variation in Przewalski's horse (*Equus przewalskii*). Anim Genet. 29:456–459.

Outram AK, et al. 2009. The earliest horse harnessing and milking. Science 323:1332–1335.
Princée FPG, Zimmerman W, Ryder OA, Dolan JM. 1990. The phenotypic approach I genetic management of Przewalski's horse. In: Seal US, Foose TJ, Lacy RC, Zimmerman W, Ryder OA, Princée FPG, editors. Przewalski's horse draft global conservation plan. Apple Valley (MN): CBSG/SSC/IUCN.
Ralls K, Ballou J. 1983. Extinction: lessons from zoos. In: Schonewald-Cox CM, Chambers SM, MacBryde B, Thomas WL, editors. Genetics and conservation. San Francisco (CA): The Benjamin/Cummings Publishing Company, Inc. p. 164–184.
Roca AL, Georgiadis N, O'Brien SJ. 2005. Cytonuclear genomic dissociation in African elephant species. Nat Genet. 37:96–100.
Ryder OA. 1993. Przewalski's horse: prospects for reintroduction into the wild. Conserv Biol. 7:13–15.
Ryder OA. 1994. Genetic studies of Przewalski's horses and their impact on conservation. In: Boyd L, Houpt KA, editors. Przewalski's horse: the history and biology of an endangered species. Albany (NY): The State University of New York Press. p. 75–92.
Ryder OA, et al. 1984. Genetics of *Equus przewalskii* Poliakov 1881: analysis of genetic variability in breeding lines, comparison of equid DNAs and a brief description of a cooperative breeding program in North America. Equus 2:207–227.
Ryder OA, Wedemeyer EA. 1982. A cooperative breeding program for the Mongolian wild horse, *Equus przewalskii*, in the United States. Biol Conserv. 22:259–271.
Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 4:406–425.
Sasaki M, Endo H, Yamagiwa D, Yamamoto M, Arishima K, Hayashi Y. 1999. Morphological character of the shoulder and leg skeleton in Przewalski's horse (*Equus przewalskii*). Ann Anat. 181:403–407.
Shimodaira H, Hasegawa M. 1999. Multiple comparisons of Log-likelihoods with applications to phylogenetic inference. Mol Biol Evol. 16:1114–1116.
Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17:1246–1247.
Short RV, Chandley AC, Jones RC, Allen WR. 1974. Meiosis in interspecific equine hybrids II. The Przewalski horse/domestic horse hybrid (*Equus przewalskii* X *E. caballus*). Cytogenet Cell Genet. 13:465–478.
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. 10:512–526.
Trifonov VA, et al. 2008. Multidirectional cross-species painting illuminates the history of karyotypic evolution in Perissodactyla. Chromosome Res. 16:89–107.
Vilà C, et al. 2001. Widespread origins of domestic horse lineages. Science 291:474–477.
Volf J, Kus E, Prokopová L. 1991. General studbook of the Przewalski horse (Zoological Garden Prague, Prague, Czech Republic) [cited 2011 Jul 1]. Available from: http://przwhorse.pikeelectronic.com/
Wade CM, et al. 2009. Genome sequence, comparative analysis, and population genetics of the domestic horse. Science 326: 865–867.
Wallner B, Brem G, Müller M, Achmann R. 2003. Fixed nucleotide differences on the Y chromosome indicate clear divergence between *Equus przewalskii* and *Equus caballus*. Anim Genet. 34:453–456.
Xu S, et al. 2007. High altitude adaptation and phylogenetic analysis of Tibetan horse based on the mitochondrial genome. J Genet Genomics. 34:720–729.
Xu X, Arnason U. 1994. The complete mitochondrial DNA sequence of the horse, Equus caballus: extensive heteroplasmy of the control region. Gene 148:357–362.
Yang F, et al. 2003. Karyotypic relationships of horses and zebras: results of cross-species chromosome painting. Cytogenet Genome Res. 102:235–243.
Zimmermann W. 1997. Die Bedeutung von Semire-servaten für das EEP Przewalskipferd. Zoo Magazin Nordrhein-Westfalen. 3:70–75.

Associate editor: Michael Purugganan