On the origin and spread of horse domestication

Vera Warmuth

A dissertation submitted to the University of Cambridge
for the degree of Doctor of Philosophy

Corpus Christi College

August 2011

Author's Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text. No part of this thesis has been submitted to any other university in application for a higher degree. The text does not exceed 300 single pages of double-spaced text.

Summary

For several decades, the origin of domestic horses has been the focus of research across multiple disciplines, yet many aspects of the horse domestication process remain poorly understood. One of the reasons for the difficulty in establishing a coherent scenario of horse domestication is that archaeological, mtDNA, and Y chromosome data have yielded ambiguous results, possibly because each class of markers reflects different aspects of the domestication process. In this thesis, I use large autosomal genetic datasets from horses sampled across Eurasia to investigate the origin and spread of horse domestication.

I begin by characterising genetic diversity of horses from the Eurasian steppes and neighbouring regions, thus laying the groundwork for a more thorough analysis into the demographic history of horses. I then investigate the origin and mode of spread of horse domestication in the Eurasian steppe region using a spatially explicit genetic model. I show that horse domestication was initiated in the western part of the steppes, and that the spread of horse domestication involved both movement of domestic herds and extensive recruitment of wild horses from across this vast region, a scenario which integrates both archaeological and molecular evidence. Having established the route of spread of early domestic horses out of their domestication origin in the western steppe, I go on to investigate the routes and levels of gene flow among Eastern Eurasian horse populations post-domestication. I show that the ancient Silk Roads have played an important role in shaping the genetic structure of Eastern Eurasian horses, facilitating gene flow across deserts and high mountain chains. Finally, I provide further compelling evidence for the persistence of wild horses in the Iberian Peninsula throughout the Holocene period, and the substantial contribution of these local populations to the gene pool of Iberian domestic horses. Together, my results provide a coherent picture of the origin and spread of horse domestication, integrating for the first time previous evidence from archaeology, mtDNA and Y chromosome sequence data.

Acknowledgements

It is difficult to express in words how grateful I am to Andrea Manica for generously welcoming me into his group and for treating me as “one of his own”. Being part of Andrea’s team meant not only that I received the best guidance a PhD student could wish for it also meant that I was part of the Zoology Department, where I felt at home.

I feel I was generally blessed with the best, most supportive, most fun bunch of supervisors and advisors imaginable. In addition to Andrea, I was lucky to have had Robert Asher as my supervisor and Bill Amos as my advisor. Both of them were there for me whenever I needed them and provided me with invaluable advice and encouragement. In the McDonald Institute, I would like to thank Mim Bower for all her support, especially during my first year at Cambridge.

Within the Department, there are so many people I would like to thank just for being there and making this such a wonderful, inspiring place, naming them all would fill several pages. Among those, there are some I would like to thank especially: Aaron Lobo in the Evolutionary Ecology Group, who I very much enjoyed having around; Jon Green in the Conservation Science lab – thank you so much for never failing to fetch me for tea and also for helping me with GIS; Stephanie Prior, who grew to be a dear friend; Richard Wallbank, Richard Merrill, Marta Costa and Steve Montgomery who made my time in the Department special.

Outside the Department, Cambridge was made a wonderful place by Natalia and Giedre: thank you both for inviting me to stay with your families when we were sampling. I am so glad I met you here in Cambridge and I count myself very lucky to have you as my friends.

Throughout my life I have gained much strength from the knowledge that my family and friends would always be there for me, no matter where in the world I was. This applies as much to the times when I travelled to far-away places as to my time in Cambridge. Thank you, Mum, Dad, and Grandma, for always believing in me, and for supporting me throughout these long years of education. Finally, I would like to thank Jürgen, for his love and friendship, for his patience and support, for having been part of my life.

While I had the lead role in all research projects that are part of this thesis, the scale of my projects required me to collaborate with a number of people. In the following, I acknowledge all the collaborators that have contributed to the work undertaken for this thesis

Chapter 3

Has been submitted for publication [at Animal Genetics]

I designed the study, generated and analysed the data, and wrote the manuscript. Samples were obtained through the assistance of Elizabeth Barrett, Mim Bower, Bryan Hanks, Shuicheng Li, Marsha Levine, David Lomitashvili, Rebecca Cassidy, Maria Ochir-Goryaeva, Grigory Sizonov, Vasiliy Soyonov, Vicky Collard, Natalia Vibla and Giedre Keen.

Chapter 4

Has been submitted for publication [at PNAS]

I designed the study, generated and analysed the data, and wrote the manuscript. The samples used in this chapter were obtained through the assistance of the collaborators listed under "Chapter 3". I developed the model together with Anders Eriksson and Andrea Manica, but its implementation in C++ was solely the work of Anders: I do not take credit for this. Anders also created Figures 4.3, 4.4, 4.5, 4.6, and 4.7

Chapter 5

The samples used in this chapter were obtained through the assistance of the collaborators listed under "Chapter 3".

Chapter 6

Has already been published as:

Warmuth V., Eriksson A., Bower M.A., Cañon J., Cothran G., Distl O., Glowatzki-Mullis M.-L., Hunt H., Luís C., do Mar Oom M., Yupanqui I.T., Ząbek T. & Manica A. (2011) European Domestic Horses Originated in Two Holocene Refugia. *PLoS ONE* **6**, e18194.

I designed the study, generated genotyping data for eight breeds, assembled this data with already published datasets (see below), analysed the data, and wrote the manuscript with the assistance of Harriet Hunt. Published genotyping datasets, including between four and seven reference samples each were made available to me by Javier Cañon, Gus Cothran, Marie Louise Glowatzki-Mullis, Tomas Ząbek, Isabel Yupanqui and Cristina Luís. Tomasz Ząbek also contributed DNA samples for one additional breed. Samples for four additional breeds were contributed by Ottmar Distl.

Table of Contents

1 Introduction

The domestication of plants and animals enabled a transition from hunter-gatherer to farming and pastoralist lifeways, thereby profoundly transforming human societies. Animal domestication constitutes a major focus of research in many disciplines, owing to the dramatic effects it had on human societies as well as on the genomes, behaviour, physiology, morphology, and life history of the animals involved.

Horse domestication has been studied for decades across multiple disciplines, yet no consensus has been reached regarding its origin and mode of spread. Previous molecular studies of horse domestication using mitochondrial DNA have revealed that wild horses throughout Eurasia contributed genetic material to the domestic horse gene pool (Cieslak *et al.* 2010); however, it is currently difficult to discern true domestication (i.e. *de novo* establishment of a founder population) from introgression of local wild females into already existing domestic stock. Microsatellite markers have successfully been used to pinpoint the origins and spread of domestication, and to identify zones of introgression in a number of domestic animal species (Hanotte *et al.* 2002; Cymbron *et al.* 2005), yet their use in horses has been restricted to investigate levels of genetic diversity within and between regional populations.

In this thesis, I use microsatellite genotyping data of more than 1500 horses sampled throughout Eurasia to investigate the origin and spread of horse domestication. Using a population genetics approach, I seek to localise primary areas of horse domestication, to determine the mode of spread of horse domestication, and to trace the main routes of gene flow post-domestication.

1.1 Documenting domestication

In the following sections, I describe the most commonly used archaeological and molecular markers in the study of animal domestication and discuss their respective strengths and weaknesses. A review of both archaeological and molecular markers in the study of horse domestication will be presented in a separate section.

1.1.1 Archaeological markers

Archaeological markers of animal domestication can be divided into two major classes, morphological and non-morphological markers. Morphological markers can be further divided into genetically determined markers, which describe morphological changes that arise over many generations as a consequence of changing selection pressures, and ontogenetic markers, which describe morphological changes that manifest themselves within the lifetime of an individual. Non-morphological markers include biogeographic and abundance data, artefacts related to animal husbandry and mortality patterns of animals from archaeological sites.

Morphological markers

Genetically determined morphological markers

Traits that are selected for in the wild differ greatly from those selected for in animals undergoing domestication. For example, in the wild, sexual selection favours traits which increase the chances of males to successfully compete against other males and to attract females, such as large body size, large canine teeth, horns, or antlers. In managed populations, relaxation of the selection pressures acting in the wild, human control over mating, and adaptation to the captive environment tend to lead to morphological changes that can be used as markers of domestication. Among the most widely observed morphological changes following domestication are reduced body size (Uerpmann 1978) and, in bovids, an increase in the variability of the size and shape of horns (Grigson 1978; Zeder 2006).

An increase in the variability of body size in horses has been used to infer beginning horse domestication by some (Uerpmann 1990; Benecke 1994b), but has been cautioned against by others (Olsen 2006a). Similarly, while a reduction in body size has been argued to be a robust indicator of beginning animal domestication (Uerpmann 1978), the generality of this phenomenon has been questioned by recent studies which failed to find a marked downward shift in body size in goats (Zeder & Hesse 2000; Zeder 2006) and donkeys (Rossel *et al.* 2008) until at least 1000 years after their presumed initial domestication.

Virtually all domestic animals have undergone a series of morphological changes relative to their wild ancestors which are thought to be a direct evolutionary consequence of selection for less aggressive behaviour (Trut 1999). These include a decrease in cranial capacity, a broadening of the forehead, a shortening of the face, and a reduction in the size and number of teeth (Zeder 2006) (so called paedomorphic, or juvenile, characteristics). While changes in tooth size and number appear relatively early on in the domestication of pigs (Albarella *et al.* 2006; Dobney & Larson 2006) and dogs (Morey 1992; Clutton-Brock 1999), most genetically determined morphological changes manifest themselves only very late in the domestication process, owing to a considerable time-lag between changing selection pressures and clear changes in morphological characters (Dobney & Larson 2006; Zeder 2006).

Ontogenetic markers

Events and processes that occur within the lifetime of domestic animals can also cause morphological changes that are visible in the archaeological record. Ontogenetic markers mostly refer to pathologies due to management practices [e.g. confinement, tethering, penning, (Olsen 2006a)] or pathologies resulting from the use of animals for work.

In Equids (horses and donkeys), certain vertebral pathologies [e.g., abnormalities of the posterior thoracic vertebrae (Levine *et al.* 2000)], joint diseases [e.g.,spondylo-arthropathies (Rossel *et al.* 2008)], compression and inclination of neural spines (Rossel *et al.* 2008) as well as wear on the lower second premolar caused by the use of a bit (Anthony & Brown 1991; Brown & Anthony 1998; Outram *et al.* 2009, but see Olsen 2006a) is used as evidence for riding; in cattle, a pronounced development of deformations on the metapodials and phalanges (e.g. eburnation, exostoses, spavin) has been shown to be common in draught cattle (De Cupere *et al.* 2000).

The exploitation of domestic animals for secondary products, such as wool, milk, traction, riding, and carrying loads, is thought to have started in the sixth millennium BP, in the course of the so called secondary products revolution (Sherratt 1981; Sherrat 1983). Before that, domestic animals were primarily used as sources of meat. Consequently, the earliest evidence of the use of cattle for traction appears around 5500 BP in the form of plough-marks, pictorial evidence of oxen and carts, and paired-ox burials, (Sherrat 1983). Thus, work-related pathologies in cattle appear no earlier than the mid-6th millennium BP, i.e. at least 4000 years after the initial domestication of cattle. In contrast, osteopathologies in donkeys, which were domesticated around the time animals started to be used for work, have been shown to appear before any changes in genetically determined morphological characteristics become visible (Rossel *et al.* 2008). Thus, the usefulness of work-related pathologies in identifying early stages of domestication depends on the point in the domestication process at which a species begins to be exploited for work as well as the type of work and its intensity.

Non-morphological markers

Demographic pattern

At the level of archaeological sites, demographic profiling of animal bone assemblages can provide insights into the subsistence strategies pursued by the site occupants. Raising animals for meat consumption, for instance, often involves culling of young males. Demographic profiles characterised by an excess of young males is therefore often used to infer domestication, particularly in sheep and goats (Zeder & Hesse 2000). In cattle, the presence in archaeological assemblages of castrated males is thought to imply the use of these animals for draught purposes, whereas a large proportion of mature females is thought to indicate stock breeding, and, after the 6th millennium BP, potentially milking (Sherrat 1983). However, when animals were raised for several purposes, when several hunting techniques were employed, or when both hunting an animal husbandry were practiced, a distinction between hunting and animal husbandry often becomes impossible, especially when sample sizes are small and bone material is poorly preserved, as is often the case.

Artefacts

Animal domestication can also be inferred from the presence of building structures associated with animals, such as pens or corrals; art work such as figurines and drawings of animals in a domestic context; artefacts documenting the use of animals for work, such as ploughs, chariots, and yokes (Sherratt 1981); and evidence for the use of their secondary products such as milk residues in pottery (Olsen 2006a).

1.1.2 Molecular markers

In the past two decades, molecular markers have been successfully used to complement archaeological data. Molecular markers of animal domestication can be divided into two broad classes, uniparental markers (mitochondrial DNA, Y chromosomal DNA) and autosomal markers . Whereas the latter are inherited from both parents, mitochondrial DNA is exclusively transmitted from mothers to their offspring (both male and female), and the Y chromosome is only transmitted from fathers to sons.

Autosomal Markers

Microsatellites

Among the most widely used autosomal markers in population genetic studies are microsatellites and single nucleotide polymorphisms (SNPs). Microsatellites are tandem repeats of non-coding DNA sequence, typically 1-5 bases in length and distributed throughout eukaryote chromosomes. The number of microsatellite repeats is often highly variable due to "slippage" during DNA replication (Levinson & Gutman 1987; Tautz & Schloetterer 1994). Microsatellite mutation rates range from 10^{-6} to 10^{-2} per generation (Schlötterer 2000), making them the most variable type of DNA sequence in the nuclear genome (Weber 1990; Schlötterer 2000).

Single nucleotide polymorphism (SNP)

A single nucleotide polymorphism (SNP) is a type of sequence variation in which a single nucleotide differs between homologous chromosomes within an individual or between individuals of the same species. SNP genotyping is becoming increasingly more efficient and cost-effective than microsatellite genotyping, and since SNPs are even more abundant in genomes than microsatellites, the typing of large numbers of SNPs provides broad genome coverage and high quality data. However, SNPs are often more affected by ascertainment bias than microsatellites (Moragues *et al.* 2010), leading, in the worst case, to a serious underestimation of the genetic diversity in populations not involved in the SNP discovery process. Microsatellite ascertainment bias, while still an issue, is less likely to have this effect since even microsatellites that are highly polymorphic in populations from which they were isolated can exhibit new alleles when genotyped in other populations (Boyko *et al.* 2009).

The most widely used approach to infer domestication origins from autosomal genetic data is based on the assumption that population movements away from centres of origin will cause a loss in genetic diversity due to repeated founder effects; as a consequence, primary areas of domestication are expected to harbour high levels of genetic diversity (Luikart *et al.* 2006), whereas areas where domestic stock was introduced are characterised by comparatively lower diversity.

Uniparental markers

Mitochondrial DNA (mtDNA)

Mitochondrial genomes are small (around 20kb) plasmids that only occur in mitochondria. Mitochondrial DNA (mtDNA), especially the so called control region, evolves extremely rapidly so that differences among populations can accumulate rapidly. In the presence of highly divergent, highly localised mtDNA haplogroups, mtDNA sequence data can be used to identify the wild populations that contributed to a domestic species' gene pool by tracing the geographic origin of the haplotype(s) found in these populations. A case for domestication is also often made when ancestral haplogroups show evidence for a demographic expansion. Demographic expansions have been shown to result in a star-like topology of haplotype networks and a unimodal distribution of the number of pairwise sequence differences (mismatch distributions) (Slatkin & Hudson 1991; Rogers & Harpending 1992). Based on the assumption that domestication (but not introgression) is followed by rapid population growth, an absence of expansion signals has been used by some distinguish between true domestication (establishment of domestic founder populations in a geographically defined area) and secondary introgression (Cieslak et al. 2010). However, while indicative, summary statistics based on mismatch distributions have been shown to be very conservative and frequently fail to detect expansions (Ramos-Onsins & Rozas 2002, Ramirez-Soriano *et al*. 2008), whereas other statistics commonly used to test deviation from the null hypothesis of stable population size (e.g. Fu's F_S (Fu 1997), Fu and Li's *D* (Fu and Li 1993)) will only detect expansions when effective population sizes post-expansion are large (Ray et al. 2003).

One of the major advantages of mtDNA is that it is available in much higher copy numbers than nuclear DNA. (There are up to several thousand mitochondria in each cell, but only one nucleus). This helps with the amplification of mtDNA from ancient samples (Hofreiter *et al.* 2001). mtDNA sequence data from populations of the extinct wild ancestors of cattle (Beja-Pereira *et al.* 2006; Bollongino *et al.* 2006; Edwards *et al.* 2007) and horses (Cieslak *et al.* 2010), for instance, have provided important insights into spatio-temporal patterns of female-mediated gene flow in these species, and have helped document early expansions of domestic stock and the replacement of original domestic stock by populations domesticated at a later stage (Larson *et al.* 2007; Larson *et al.* 2010). Owing to its exclusively maternal

mode of inheritance (Hutchison *et al.* 1974), mtDNA only documents the genetic history of females. While the capture of wild females is expected to have formed an integral part of the initial establishment of founder populations of domestic animals, increasing evidence suggests that male-mediated gene flow may have been just as important as female-mediated gene flow in shaping the genetic make-up of domestic populations (Bradley *et al.* 1994; Hanotte *et al.* 2000; Kantanen *et al.* 2009; Perez-Pardal *et al.* 2010).

In addition, demographic reconstructions made from effectively single markers, such as mtDNA and Y-chromosomal DNA have considerable uncertainty due to the substantially higher rates of stochastic genetic drift compared to autosomal loci. Phylogenetic inference based on mtDNA and Y chromosome sequence data is effectively based on a single segregating locus, which is subject to specific evolutionary, and atypical population dynamics; thus, the genetic patterns observed in these uniparental markers might not be representative of the genome as a whole (Bruford *et al.* 2003; Bradley & Magee 2006).

Y-chromosomal DNA

Y-chromosomal DNA, in particular Y-specific microsatellites, Y-specific single nucleotide polymorphisms (SNPs) and stretches of sequence data covering the non-recombining region of the Y chromosome (NRY), are often used to complement mtDNA data because they reflect male-mediated gene flow. In contrast to the mitochondrial genome and the autosomes, nucleotide diversity on the Y chromosomes is generally low. Several reasons have been put forward to account for the low levels of variability of Y chromosomes, including their low effective size (one quater that of autosomes), which makes them susceptible to selective forces, such as bottlenecks, strong selective pressure on the functional genes that lie on the Y chromosome together with a general lack of recombination, and, in mammals with a polygynous mating system, low male effective population sizes (Ellegren 2003). In addition to low levels of variability, the non-recombining region of the Y chromosome is also characterised by an abundance of highly repetitive "junk" DNA, which makes the development of informative Y chromosomal markers difficult. While Y chromosomal DNA suffers from the same limitation as mtDNA in behaving as a single marker, its use can reveal important aspects of the demographic history of a species or population.

The inherently stochastic nature of evolutionary/demographic processes can cause loci with similar histories to show different genetic patterns (Sunnucks 2000). Genetic diversity patterns that are estimated by averaging across multiple, independently evolved loci are therefore both statistically more powerful and more accurate than those estimated from (effectively) single loci, such as mtDNA and Y chromosomal DNA. Since high levels of genetic diversity are also expected in areas where genetically differentiated domestic populations hybridise (Luikart *et al.* 2006), the question of whether high levels of genetic diversity in a particular region may be caused by admixture should be addressed before concluding that such areas represent primary areas of domestication.

1.2 Domestication of cattle, sheep, goats and pigs

1.2.1 Goats

Goats (*Capra hircus*) were among the first livestock animals to be domesticated in Eurasia. It is now widely accepted that the wild ancestor of goats was the bezoar (*Capra aegagrus*, Manceau *et al.* 1999; Luikart *et al.* 2006), the natural range of which stretches from Anatolia in the west to Pakistan in the east. Archaeological evidence traces the domestication of goats to Eastern Anatolia around 10,500 cal BP, and the Central Zagros Mountains between 9900 and 9500 cal BP (Zeder & Hesse 2000). Further support for a domestication origin of goats in Eastern Anatolia comes from the high prevalence of a certain mtDNA haplogroup (haplogroup A) in wild bezoar from this region (Naderi *et al.* 2008), which is significant in so far as more than 90% of modern domestic goats fall into this haplogroup (Naderi *et al.* 2007). While there is so far no corroborating molecular evidence for an origin of goat domestication in the Central Zagros Mountains, Naderi *et al.* (2008) found a genetic signal of a population expansion in bezoar of the C haplogroup, which likely originated in the region comprising the Southern Zagros Mountains and the Central Iranian Plateau. Since the timing of this expansion roughly corresponds to the period when goat domestication took place, the authors suggested placing the second domestication event in the Southern, not the Central Zagros Mountains. However, based on low prevalence of the C haplogroup in modern goats (1.4%, Naderi *et al.* 2007) the contribution of early domestic goats from Southern Zagros to the current domestic goat gene pool has been suggested to have been low.

Following their domestication in southwest Asia, goats were imported to Europe (Price 2000). In line with expectations based on population genetic theory, the spread of domestic goats out of their domestication origin in southwest Asia into Europe left a strong genetic signature in the form of a southwest to northeast decline in autosomal genetic diversity (Cañón *et al.* 2006) and a cline in allele frequencies (Laloë *et al.* 2010). In contrast, mtDNA haplotype diversity is similarly high in goats from across Eurasia (Naderi *et al.* 2007).

1.2.2 Sheep

According to archaeological evidence, sheep (*Ovis orientalis aries*) were domesticated from Asiatic Mouflon in the Fertile Crescent region in southwest Asia between 9000 and 8000 BP (Clutton-Brock 1999; Hiendleder *et al.* 2002). Using a family of endogenous retroviruses as genetic markers, Chessa *et al.* (2009) have recently shown that there were at least two population expansions of domestic sheep out of southwest Asia into Europe. The first expansion involved the originally domesticated stock, the descendants of which survive in some breeds from north-western and northern Europe (e.g. Soay sheep, Hebrideans, Orkney sheep, Icelandic, and Nordic breeds) as well as in Mouflon of Sardinia, Corsica, and Cyprus (Chessa *et al.* 2009); the second population expansion brought the lineages of wool sheep to Europe, the descendants of which are widespread today.

The spread of domestic sheep from their domestication centre in southwest Asia has left a genetic signature in modern domestic sheep in the form of a southeast to northwest oriented decline in autosomal genetic diversity as estimated using both SNPs (Tapio *et al.* 2010) and microsatellites (Peter *et al.* 2007), and a cline in allele frequencies (Laloë *et al.* 2010). Phylogeographic structure in domestic sheep is very weak (Meadows *et al.* 2005; Handley *et al.* 2007a; Kijas *et al.* 2009; Tapio *et al.* 2010), likely reflecting high levels of cross-breeding, especially in Europe.

1.2.3 Cattle

Current evidence from archaeology and genetics suggests that at least two genetically divergent lineages of aurochs (*Bos primigenius*) were independently domesticated, one in southwest Asia around 9000 BP (*B. p. primigenius*, the ancestor of European taurine, or

humpless cattle), and one in the Indian subcontinent, possibly in present-day Pakistan (*B. p. namadicus*, the ancestor of zebu, or humped cattle) (Loftus *et al.* 1994).

Following their domestication in southwest Asia, taurine cattle were introduced to Europe, most likely as part of the expansion of early Neolithic farmers out of southwest Asia (Epstein & Mason 1984). Archaeological data suggests that the spread of cattle across Europe followed two distinct main routes, one along the coast of the Mediterranean Sea (the "Mediterranean route"), and one following the inland course of the river Danube into the plains of central and northern Europe (the "Danubian route") (Bogucki 1996). Interestingly, the differential migration routes are reflected as different relationships between genetic and geographic distances in European cattle breeds from northern Europe and the Mediterranean region, respectively (Cymbron *et al.* 2005). As in sheep and goats, the expansion of domestic cattle out of southwest Asia is reflected in a southwest to northeast decline in autosomal genetic diversity (Loftus *et al.* 1999; Cymbron *et al.* 2005; Medugorac *et al.* 2009) and a cline in allele frequencies (Laloë *et al.* 2010). A reduction of genetic diversity with increasing distance from southwest Asia can also be observed in the mtDNA of European cattle populations (Troy *et al.* 2001).

While the derived southwest Asia origin of the vast majority of European taurine cattle has received support from numerous studies employing both mtDNA (Bradley *et al.* 1996; Troy *et al.* 2001; Edwards *et al.* 2007; Achilli *et al.* 2009) and microsatellites (Loftus *et al.* 1999; Cymbron *et al.* 2005; Medugorac *et al.* 2009), recent mtDNA studies suggest that wild females of local European aurochs populations, notably from Italy, central, and northern Europe, also contributed to the gene pool of European domestic cattle, although to a very limited extent (Beja-Pereira *et al.* 2006; Achilli *et al.* 2008; Achilli *et al.* 2009). The extent to which European aurochs populations contributed to modern stock via male-mediated gene flow remains unclear (Groeneveld *et al.* 2010).

There is increasing evidence that African cattle may have been domesticated from indigenous African aurochsen (Grigson 1991; Bradley *et al.* 1996; Wendorf & Schild 1998; Troy *et al.* 2001; Hanotte *et al.* 2002; Perez-Pardal *et al.* 2010), or that indigenous aurochsen at least contributed considerable genetic material to African domestic cattle. Alternatively, African cattle descend from stock that was domesticated in southwest Asia and subsequently introduced to Africa (Epstein 1957; Beja-Pereira *et al.* 2006). Regardless

of whether the initial herds were the product of local aurochsen domestication or not, cattle pastoralism spread both southward and westward across Africa from a centre of origin in eastern Africa (Hanotte *et al.* 2002).

1.2.4 Pigs

The domestic pig originates from the Eurasian wild boar (*Sus scrofa*). Both archaeological and molecular data have long suggested two separate domestications of differentiated subspecies of wild boar in Central Europe and Central China (Kijas & Andersson 2001; Larson *et al.* 2005; Larson *et al.* 2010). However, recent research using mtDNA has revealed additional areas where wild boar were either domesticated or contributed to local domestic stock, including Italy (Larson *et al.* 2005), the Indian subcontinent (Larson *et al.* 2005; Tanaka *et al.* 2008), mainland (Burma, Thailand), and insular Southeast Asia (Larson *et al.* 2005). Using ancient DNA, Larson (Larson *et al.* 2007) showed that domestic pigs of southwest Asian ancestry were initially introduced to Europe, and that the descendants of this early expansion were later replaced by stock that was domesticated from local wild boar.

1.3 The evolutionary history of horses – an overview

1.3.1 Evolution and spread of caballine horses

The genus *Equus* shares common ancestry with *Hyracotherium*, a small (about 25–45 cm in height), forest-dwelling animal that lived in the early and middle Eocene between 55 and 52 mya (Froehlich 2002; MacFadden 2005). While species of the genus *Hyracotherium* were found across the Northern Hemisphere (Froehlich 2002), the evolution of the genus *Equus* occurred (mainly) in North America (Lewis 1937).

As the primary vegetation in North America gradually changed from (tropical) forest to grasslands, the lineage that would give rise to modern horses evolved high-crowned, large-sized teeth with pronounced ridges and longer, more slender legs with a reduced number of toes, adaptations which would allow them to break up tough grasses and to escape from predators on open plains.

The common ancestor of all *Equus*, a species named *Equus simplicidens*, first appears in the Hagermann Fauna in Idaho, North America between 3.7 and 3.2 mya (Savage & Russel 1983; Forsten 1992). Morphologically, early fossil *Equus* belonged to a lineage that would give rise to stenonine horses (today represented by zebras and asses). Caballine, or true horses (today represented by domestic horses and the Przewalski's horse), are thought to have split from this ancestral stenonine lineage between 1.9 (Forsten 1988; 1992) and 2.3 mya (Oakenfull *et al.* 2000).

During the first major glaciations in the late Pliocene, about 2.6 mya, stenonine horses dispersed to Eurasia via the Bering landbridge (Lindsay *et al.* 1980; Azzaroli 1983) as part of a faunal exchange between North America and Eurasia which also brought mammoths *(Mammuthus meridionalis)* and gazelles *(Gazella borbonica)* to Eurasia (Azzaroli 1983). In Eurasia, the stenonine lineage diversified into zebras and asses, probably in the course of one or several rapid radiation events (Oakenfull *et al.* 2000). Caballines are said to appear with *Equus scotti* Gidley in North America between 1.9 (Kurtén & Anderson 1980; Savage & Russel 1983) and 1.4 mya (Prothero & Schoch 2002). Caballine horses dispersed to Eurasia towards the end of the Villafranchian, where they replaced stenonine lineages over wide geographic areas (Forsten 1988). By about 1 mya, caballine horses had reached central-east Europe, and by 900 kya they had begun to differentiate into various ecotypes (Prothero & Schoch 2002). It is currently unknown how many full species of caballine horses there were in Eurasia. However, it is now widely accepted that the large, heavy forms [e.g. *E. germanicus* (=*E. latipes*) and *E. mosbachensis*], which were initially widespread, died out sometime in the Late Pleistocene, leaving only a relatively small species (Forsten 1988; Prothero & Schoch 2002; Olsen 2006a), which is commonly referred to as *Equus ferus* Boddaert, 1785 (Nobis 1971). Further evidence for the presence of a single caballine species in Eurasia by the end of the Pleistocene has recently been provided based on genetic evidence showing that all modern caballine horses (i.e domestic horses and the Przewalski's horses) descend from a single wild species (Weinstock *et al.* 2005).

The temporal origin of *Equus ferus* in Eurasia is unknown. It has been argued by some that all caballine horses in Eurasia, from the first appearance of caballines about 1 mya through to the Holocene period, belonged to a single species (Cramer 2002; van Asperen 2010b), implying an arrival time of *E. ferus* in Eurasia about 1 mya (see above). In contrast, Forsten (1988) states that *E. ferus*, was one of at least three caballine species in Eurasia that

"succeeded each other and partly overlapped chronologically", with *Equus ferus* being the most recent form. The notion of a rather late arrival of *Equus ferus* in Eurasia is also shared by Prothero & Schoch (2002), who suggest that *Equus ferus* first appeared around 200,000 years ago in the fossil record of East Asia.

During the Upper Pleistocene, much of northern Eurasia was dominated by treeless tundra and prairie–steppe, the so called Pleistocene cold-steppe or Mammoth Steppe (Tarasov *et al.* 2000). During that time, the known distribution of caballine horses ranged from 75 N to 35 N and from 130 E to 10 W, dependent on the extent of glacial cover (Eisenmann 1996). The beginning of the Holocene period was marked by dramatic changes in climate and therefore vegetation. In Europe, an increase in temperature and precipitation resulted in a successive replacement of steppe-like vegetation by birch-pine, and eventually by the mixed oak forests that would become the dominant forest type in much of Central Europe (Lang 1994).

Steppe vegetation is thought to have primarily persisted in southwest and central Asia (Huntley 1988; Prentice *et al.* 1996; Tarasov *et al.* 1998). Palaeovegetation data increasingly suggests that open landscapes also persisted in some parts of the Iberian Peninsula, especially in the semi-arid and high elevation plateaus in south (Pantaléon-Cano *et al.* 2003; Carrión *et al.* 2001) and central (Preece 1991) Spain.

The climatic changes at the Pleistocene-Holocene boundary have been held, at least partly, responsible for the extinction crisis that affected numerous large land vertebrates in Eurasia and America (Hofreiter & Stewart 2009). Horses were no exception: in America, all equid species became extinct towards the end of the Pleistocene, with the last fossil findings of caballine horses in Alaska dated to between 14,180 and 14,960 cal BP (Koch & Barnosky 2006). While a survival of horses in Alaska up until 10,500 BP has recently been proposed based on sedimentary DNA (Haile *et al.* 2009), this date needs to be viewed cautiously as dates obtained from sedimentary DNA are associated with large errors. In Eurasia, the geographic distribution of *E. ferus* contracted considerably. In line with expectations based on the ecology of (caballine) horses, the geographic range of *E. ferus* contracted eastward into southwest Asia and Central Asia (Stewart 2007), where large tracts of steppe vegetation persisted (Tarasov *et al.* 1998). In Eastern Eurasia, *E. ferus* maintained a more or less contiguous range stretching from the Carpathian Mountains to Mongolia, largely

coinciding with the present day distribution of the Eurasian steppe (Olsen 2006b). In contrast, west of the Carpathian Mountains, horses became locally extinct over large parts of its former range, probably as a consequence of widespread habitat loss as the Pleistocene cold-steppe was gradually replaced by dense forest (Olsen 2006b).

Based on the spatio-temporal distribution of wild horse remains in Europe between 10,300 and 4800 BP, wild horses were largely absent in much of Europe and Britain between 9000 and 5500 BP (Benecke 2006; Boyle 2006; Steppan 2006; Sommer *et al.* 2011). From 5000 BP, horse numbers slowly started to increase again in areas where Neolithic farmers had begun to clear forests to obtain grazing grounds for their livestock (Kalis *et al.* 2003; Sommer *et al.* 2011), suggesting a strong link between landscape "openness" and wild horse abundance. Interestingly, the only geographic region in Western Europe for which a continuous fossil record for horses appears to exist until at least 5000 BP, is the Iberian Peninsula (Uerpmann 1990; Sommer *et al.* 2011) (see above). However, because most of the fossil material for horses in Iberia is not directly dated (Olsen 2006a; Sommer *et al.* 2011), the question of how long wild horses survived in the Iberian Peninsula – with all its implications for horse domestication in Iberia, remains open.

1.4 Horse Domestication

1.4.1 Archaeological evidence

Archaeological evidence increasingly points towards the steppes of modern-day Ukraine, southwest Russia and Kazakhstan (the western Eurasian steppe) as the area where horses were first domesticated (Anthony & Brown 2000, 2003; Outram *et al.* 2009; reviewed in Olsen 2006a), although claims for local horse domestication in Europe have also been made (Uerpmann 1990; Benecke 1994a), particularly in the Iberian Peninsula, and southwest France (Uerpmann 1990).

Until recently, any claims for horse domestication before 4000 BP, the time when the first horse-drawn chariots appeared in the Ural steppe (Anthony 1995), were based on indirect, in some instances highly ambiguous lines of evidence. However, recent findings of horse milk residues in pottery from Botai, a settlement in north-Kazakhstan, provide the first

uncontested evidence for horse husbandry in the western part of the Eurasian steppe by around 5500 BP (Outram *et al.* 2009).

Other, more indirect lines of evidence for horse domestication in the western steppes include, but are not limited to, an increase in the number of horses in archaeological sites (Bibikova 1975; Bibikova 1986b), a reduction of body size accompanied by an increase in variability (Uerpmann 1990; Benecke 1993), osteometric data (Outram *et al.* 2009), mortality patterns thought to reflect the selective slaughter of subadult male horses and thus herd management (Bibikova 1986a; Bibikova 1986b; but see Levine 1999a), tooth crowding (Bökönyi 1993), tooth pathologies thought to reflect bit wear (Brown & Anthony 1998; Anthony & Brown 2000; Outram *et al.* 2009; but see Levine 1999b; Levine 2004; Kosintsev 2006; Olsen 2006a), the appearance of horses in burials together with other domesticates and/or humans (Bökönyi 1993; Anthony & Brown 2000), and the presence cultural indicators of horse control (Olsen 2006a).

Among the most important criteria for horse domestication in the western steppes has been an increase in the absolute number as well as the proportion of horses in faunal assemblages of Chalcolithic (Copper Age) sites located east of the Dnepr river, but not west of it (Bibikova 1975): whereas the ratio of horse bones to the overall number of animal remains in faunal assemblages west of the Dnepr remained far below 10% until about 4500 BP (see also Boyle 2006)), horses made up 25% of the animal remains found in Moljukhov Bugor, 27% in Alexandrija, 61% in Dereivka (all three located in southern Ukraine), 80% in Repin (on the southern Don river), 66% in Kozhai, and 99% in Botai (both located in north Kazakhstan) (Bökönyi 1993; Kosintsev 2006). While some argue that the temporal context and the geographic pattern of horse numbers in archaeological sites indicates the presence of equine livestock (Bibikova 1986b; Olsen 2006a), others have suggested that an increased number of horses in the steppes, an ecological region where wild horses were abundant, may also reflect an intensified hunt for wild horses, probably as a response to the decline of other commonly hunted wild animals (Uerpmann 1990; Levine 1999a).

At a number of archaeological sites in the steppes, horse remains have been reported to exhibit increased body size variability (Bökönyi 1974), which is considered a classic indicator of domestication by some (Uerpmann 1990; Benecke 2006). However, it has been argued that changes in body size variability may also reflect changes in exploitation

patterns, for example from specialist hunting techniques targeting prime adults to the hunting of whole herds (Uerpmann 1990; Levine 1999a). A notable east-to-west decline in the body size of wild horses from Ukraine towards the Iberian Peninsula (Nobis 1971; Uerpmann 1990) may further blur the pattern, as any east-west oriented migration of wild horses would have led to an increase in body size variability in areas where horses from East and West co-occurred. In this case, inference of beginning horse domestication from increased size variability may result in an overestimation of the onset of domestication.

Similarly, while the large proportion of subadult males and the absence of aged individuals in Dereivka and Botai have been argued to reflect the selective slaughter of young males in a husbanded herd (Bibikova 1986a; Kuz'mina 1993), an excess of (young) males may also indicate specialised hunting techniques targeting bachelor groups (Levine 1990; Uerpmann 1990; Levine 1999a; Kosintsev 2006). For a number of reasons, the usefulness of mortality patterns in horses has been questioned: for one, culling of young males may have been less common than in other livestock species. In some cultures, stallions were highly regarded as objects of prestige and cult (Olsen 2006a). In addition, male horses under the age of 4-5 are very difficult to distinguish from females because size differences are marginal and because the large canines which distinguish males from females are not present in juvenile males (Olsen 2006a). The only other reliable criterion by which males can be distinguished from females is the shape of the pelvis, which is often not available in sufficient numbers to carry out a statistically sound population analysis (Olsen 2006a). If juvenile males cannot be distinguished from juvenile females, mortality patterns characterised by a high proportion of juveniles could therefore also reflect a hunting strategy which targets whole herds.

The appearance of sacrificial horses in human burials in the western steppe has also been put forward as evidence for an onset of horse domestication in this area. The earliest findings of horses appeared in burials in Khvalynsk and S'yezzhe, two sites in the Volga region dating around 6500 BP (Anthony *et al.* 2006). In this region, the presence of wild animals in human graves was rare, whereas clearly domestic animals such as cattle, sheep and dogs were frequently interred with humans (Olsen 2006a).

The observation that horses occurring in human graves were buried in exactly the same way as other domestic animals (so called "head and hoof" burials), has been suggested to reflect the domestic status of these sacrificial horses (Olsen 2006a).

Finally, some sites in the western steppes have been claimed to contain evidence for horseback riding. In Dereivka, for instance, six antler tine artefacts have been recovered which are of the same shape as antler tines unambiguously identified as bridle parts (Telegin 1986). However, based on findings of very similar artefacts, in areas where horses were absent at the time, it has been suggested that antler tines could have been used for many purposes, questioning their use as evidence for horse riding (Uerpmann 1990). At two sites, Botai and Kozhai, horse teeth (lower second premolars) with bevels of more than 3 mm on their mesial part were found (Brown & Anthony 1998). Riding experiments have shown that the habitual outfitting of horses with a bit, even with organic materials such as leather and hemp, can cause wear patterns similar to those observed in Botai and Kozhai, leading (Brown & Anthony 1998; Anthony *et al.* 2006) to suggest that at least some horses in these sites were ridden and thus domesticated. However, inference of horse-back riding solely on bit wear has been cautioned against by (Olsen 2006a), who found very similar wear patterns in several specimens of unambiguously wild horses dating to the Upper Pleistocene, suggesting that bevels on the lower premolars are not necessarily caused by bit wear but may reflect natural dietary wear (see also Levine 1999b; Levine 2004; Kosintsev 2006).

In a number of areas west of the Carpathian Mountains, notably the Iberian Peninsula and southwest France, but also in some areas in Central Europe (Benecke 1994a), local horse domestication has been inferred from the observation that (supposedly) local domestic horses were more similar in (their) body size to (supposedly) local wild horses than to wild horses from the steppes, the latter having been considerably larger than wild horses in Europe (Uerpmann 1990). However, (Bökönyi 1993) has argued that local domestication in Europe, at least in Central Europe, cannot have been on a great scale, because both wild horses and late Neolithic/early Bronze Age (i.e. early domestic) horses were rare in these regions. Conversely, the presence of large horses outside the steppes, such as in Central Europe (Uerpmann 1990; Benecke 2006), the North Caucasus, Transcaucasia, and Eastern Anatolia (Bökönyi 1993) is seen as evidence for their Eurasian steppe ancestry.

1.4.2 Molecular evidence

So far, molecular studies of horse domestication have found no evidence for a geographically restricted origin of horse domestication in the western Eurasian steppe or anywhere in Eurasia. Furthermore, while matrilineal diversity in Eurasia-wide populations of domestic horses is high, patrilineal diversity is extremely low. Taken together, this suggests that archaeological, mtDNA, and Y chromosomal data reflect different aspects of the domestication process and that data from additional sources will be required to more accurately and comprehensively characterise the horse domestication process. In the following sections I present the key findings of molecular studies of horse domestication.

Using mtDNA sequence data from 29 individuals belonging to 14 domestic horse breeds and Przewalski's horses, Lister *et al.* (1998) documented considerable haplotype diversity in domestic horse breeds but not in Przewalski's horses. From the observed low haplotype diversity in Przewalski's horses, Lister et al. inferred that wild horse populations must have been genetically rather homogeneous. Based on this assumption, they concluded that the high haplotype diversity observed in modern horses reflects the genetic contribution of wild stock distributed over "a moderately extensive geographical region".

Low levels of haplotype diversity in eight wild horses from a permafrost site in Alaska dated to between 12,000 to 28,000 years ago led Vilà *et al.* (2001) to draw similar conclusions: a single, geographically restricted wild population could not have contained enough haplotype diversity within it to account for the large number of haplotypes in domestic horses. However, while Lister *et al.* (1998) envisaged a scenario in which wild females were captured within a moderately large area, but domesticated in only one or very few locations, Vilà *et al.* (2001) proposed a scenario whereby wild horses from a wide geographic area were domesticated on numerous, possibly independent, occasions, and that this was made possible by the spread of technique for horse domestication.

Using mtDNA sequence data from 652 modern domestic horses, Jansen *et al.* (2002) found a total of 81 haplotypes most of which fell into 17 clusters that formed 7 major haplogroups. In contrast, only three, closely related haplotypes were found in Przewalski's horses despite increased sample sizes when compared to the previous studies (Lister *et al.* 1998; Vilà *et al.* 2001; Jansen *et al.* 2002). Based on the assumption that *Equus* is 1 my old

(the development of a cranial character common to all extant equid species, according to the authors the latest date for the appearance of *Equus*), and an onset of domestication around 11,500 BP (the end of the glacial period; according to the authors the earliest possible date for horse domestication), Jansen *et al.* (2002) estimated an mtDNA mutation rate of 1 mutation per 100,000 years. Accounting for the number of lineages that evolved in the time since domestication, the authors estimated that at least 77 successfully breeding wild females must have been incorporated into the gene pool of domestic horses.

For the first time, there appeared to be evidence for an association of some haplotype clusters with particular breeds or geographic regions, notably Jansen *et al.*'s cluster C1, which was strongly associated with northern European pony breeds, and cluster D1, which appeared to be most strongly represented in Iberian and North African horses. Based on their results, Jansen *et al.* (2002) argued for numerous spatially and temporally localised domestication events. Like Vilà *et al.* (2001), Jansen *et al.* (2002) regarded it as unlikely that human societies in different parts of Eurasia would have acquired the techniques and skills needed to successfully capture, tame, and breed horses independently of one another, therefore proposing cultural transmission as the main mechanism by which the use of horses spread.

By covering major gaps in the geographic coverage of sampling locations, especially in Eastern Eurasia, McGahern *et al.* (2006) and Lei *et al.* (2009), showed that mtDNA sequences in Europe were, in fact, not associated with any one breed or geographic region as suggested by Jansen *et al.* (2002). Instead, both studies found clear east-west oriented frequency clines concerning two of the seven major haplogroups, with a higher prevalence of their haplogroup F in the Middle East and Asia and a higher prevalence of their haplogroup D in Europe. While only explaining 2.71% of the total variation (McGahern *et al.* 2006), this geographic organisation was found to be statistically significant (McGahern *et al.* 2006; Lei *et al.* 2009).

In the most comprehensive study to date, Cieslak *et al.* (2010) analysed mtDNA sequences of 1754 modern and 207 ancient samples from across Eurasia, representing different time slices. The enlarged dataset (which now comprised 87 haplotypes belonging to 19 haplogroups) revealed for the first time that mtDNA variability in Eurasian wild horses was not organised into genetically distinct, geographically localised clusters, as previously

assumed (Vilà *et al.* 2001; Jansen *et al.* 2002). On the contrary, population sub-structuring in Holocene wild horses was generally low, with a noticeable subdivision only between populations from Iberia and the Eurasian steppe region ($0.1 \leq F_{ST} \leq 0.4$). The results by (Cieslak *et al.* 2010) further showed that haplotype diversity in wild horses was not much lower than that of modern domestic horses (0.978 ± 0.035 and 0.600 ± 0.131 in wild horses from northeast Siberia and Iberia, respectively, versus 0.994 ± 0.019 - 0.842 ± 0.041 in nine modern Chinese breeds), questioning previous assumptions of low haplotype diversity in individual wild horse populations and the inferences drawn from it (i.e. multiple domestications).

From the spatio-temporal distribution of mtDNA haplotypes in Eurasian populations of wild and domestic horses, and from the geographic association of some of these haplotypes, (Cieslak *et al.* 2010) suggested that wild horses from many regions in Eurasia contributed to domestic stock, and that different processes could have caused the observed high levels of matrilineal diversity in domestic horses, including the domestication from a single, diverse wild population, multiple domestications, and/or widespread introgression of local wild horses into domestic stock.

The exceptionally high levels of matrilineal diversity contrast with the presence of only one segregating site on the horse Y chromosome (Ling *et al.* 2010a). To explain this pattern, many researchers have invoked a sex-bias towards females in horse breeding (Lindgren *et al.* 2004; Wallner *et al.* 2004; Vilà *et al.* 2006; Kavar & Dovc 2008; Cieslak *et al.* 2010; Ling *et al.* 2010a). However, a strong reproductive skew in male wild horses (which is expected in harem-holding species such as horses) could be an equally plausible explanation if only few wild males in a population get the chance to pass on their genes, wild populations are expected to be characterised by high matrilineal but low patrilineal diversity.

The spread of domestic horses out of geographically restricted domestication origins would have resulted in a further loss of patrilineal, but not matrilineal diversity if spreading herds were re-stocked with female wild horses (Lindgren *et al.* 2004; Vilà *et al.* 2006). However, while any of these explanations, whether individually or in combination, may have reduced levels of equine Y chromosome diversity to a certain extent, the reason(s) for the presence of only two haplotypes in Eurasian populations of domestic horses (Ling *et al.* 2010a)

requires further investigation, especially in light of historical records documenting the widespread practice of having domestic females covered by wild males (Jankovich 1971).

The prevalence of a particular mtDNA haplogroup (haplogroup D in Jansen) in modern Iberian horses was long taken as evidence for local horse domestication in Iberia (Jansen *et al.* 2002; Royo *et al.* 2005). While it has recently been shown that this particular haplogroup only occurred in Iberia from the Middle Ages onward (Lira *et al.* 2010), the presence of pre-domestic Iberian lineages in both early and modern domestic horses of Iberian descent (Cieslak *et al.* 2010; Lira *et al.* 2010) suggests that Iberian wild stock was involved in the domestication process. However, owing to a lack of evidence for a demographic expansion between 5000 and 600 BP (Cieslak *et al.* 2010; Lira *et al.* 2010), as well as a lack of a genetic signal for an increase in coat colour variation [a signal of intentional human selection (Fang *et al.* 2009)](Ludwig *et al.* 2009), it remains unclear whether Iberian wild horses form the basis of Iberian domestic stock or whether local mares were incorporated into already domesticated stock, possibly having come from elsewhere (Cieslak *et al.* 2010; Lira *et al.* 2010).

1.5 Objectives and structure of thesis

Despite decades of research across multiple disciplines, many aspect regarding the origin and spread of domestic horses remain poorly understood. In this thesis, I use autosomal genetic data from more than 1500 horses sampled throughout much of Eurasia to investigate the origin and spread of horse domestication as well as subsequent population movements.

In Chapter 3, I investigate geographic patterns of genetic variation in previously understudied horse populations from Eastern Eurasia, a vast geographic region which played an important role in the early history of horses. I find a significant decline in genetic diversity with increasing distance from the easternmost sampling location in my dataset, Mongolia, consistent with an expansion of horses out of East Asia.

In Chapter 4, I reconstruct both the population genetic structure of the extinct wild progenitor of domestic horses (*E. ferus*) and horse domestication by parameterising a spatially explicit model with genetic data of horse populations from throughout the steppes. I show that horse domestication was initiated in the western part of the steppes, and that the spread of horse domestication involved both movement of domestic herds and extensive recruitment of wild horses from throughout the steppes, a scenario which integrates for the first time archaeological and molecular evidence.

Having established the route of spread of early domestic horses out of their domestication origin in the western steppe, I investigate the routes and levels of gene flow among Eastern Eurasian horse populations post-domestication (Chapter 5). I show that the Silk Roads have played an important role in shaping the genetic structure of Eastern Eurasian horses, facilitating gene flow across deserts and high mountain chains.

Finally, in Chapter 6, I address the long-standing debate surrounding horse domestication in Europe. I show that traditional horse breeds from areas characterized by at least partly open landscapes in the mid-Holocene, the Iberian Peninsula and southwest Asia, harbour high levels of genetic diversity, whereas horses from areas that were densely forested during that time (i.e. Central Europe and Britain) were characterized by very low diversity.

Since I found no indication for higher levels of admixture in Iberian horses compared to horses from the rest of Europe, the results presented in this chapter corroborate and amplify previous evidence for the persistence of wild horses in the Iberian Peninsula throughout the Holocene period and indicate that Iberian wild stock contributed substantially to the gene pool of Iberian domestic horses.

Together, my results provide a coherent picture of the origin and spread of horse domestication, integrating for the first time previous evidence from archaeology, mtDNA and Y chromosome sequence data.

2 General Methods

2.1 Datasets

Genotyping data from Western and Eastern Eurasia had to be analysed separately, because the former, which included already published data, was genotyped at fewer loci than the latter. In addition, while the samples from Eastern Eurasia consisted of randomly sampled local horses that did not belong to any specific breed, samples from Western Eurasia did belong to specific breeds; the two datasets were therefore not directly comparable.

2.1.1 Non-breed horses from Eastern Eurasia

Geographic scope

Eastern Eurasia, in this thesis, is defined as the geographic region stretching from latitude 57N (Lithuania) to 25N (southwest China) and from longitude 24E (western Ukraine) to 103E (Mongolia).

Sampling

The dataset covering Eastern Eurasian consists of 455 non-breed horses from 17 sampling locations spanning eight countries. The sampling locations are shown in Figure 2.1. Further information on the samples can be found in Table 2.1.

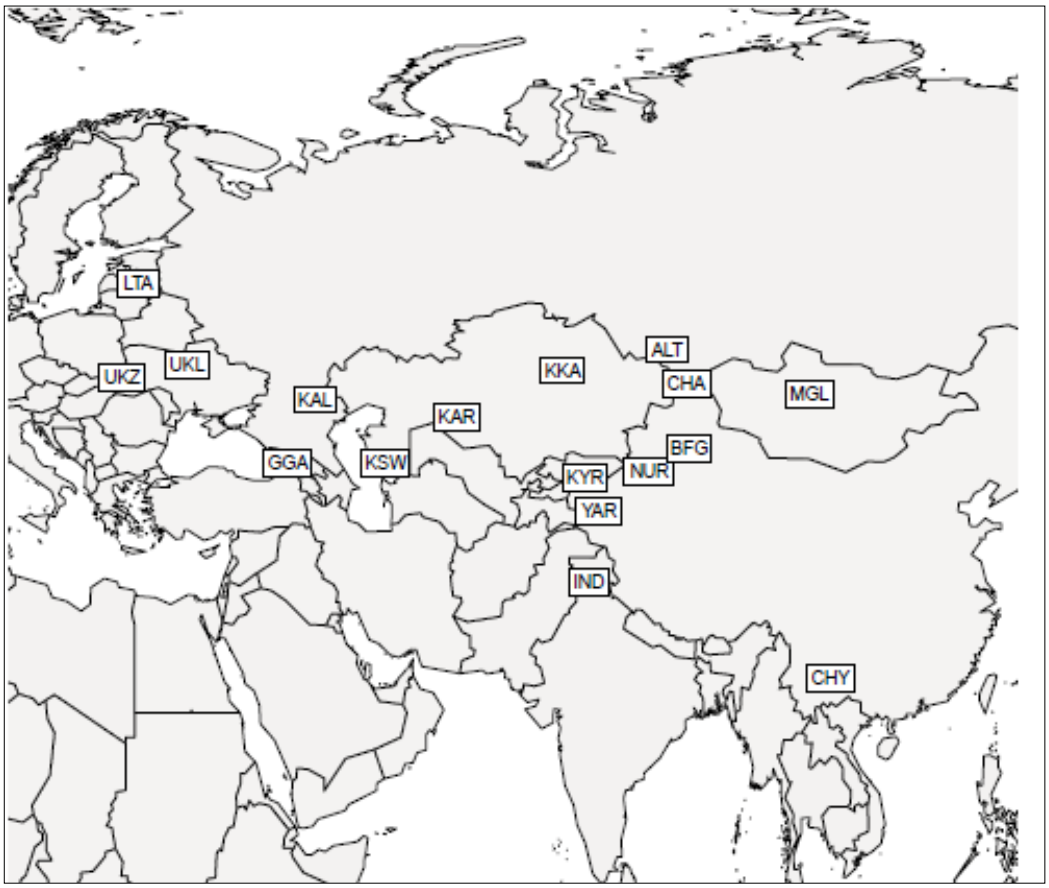

Figure 2.1

Sampling locations in Eastern Eurasia. ALT – Russia, Republic of Altai, BFG – China, Xinjiang Province, CHA, China, Xinjiang Province, CHY – China, Yunnan Province, GGA – Georgia, Samegrelo and Zemo Svaneti region, IND – India, Jammu and Kashmir State, KAL –Russia, Republic of Kalmykia, KAR – Kazakhstan, Kyzylorda Oblast, KKA – Kazakhstan, Karagandy Oblast, KSW – Kazakhstan, Mangystau Oblast, KYR – Kyrgyzstan, Naryn Oblast, LTA – Lithuania, MGL – Mongolia, Övörkhangai Aimag, NUR – China, Xinjiang Province, UKL – Ukraine – Lviv Oblast, UKZ – Ukraine, Zakarpattia Oblast, YAR – China, Xinjiang Province.

Table 2.1

Geographic origin of samples included in Chapters 3 and 5. Samples included in Chapter 4 are indicated by an asterisk.

ID	Country	Administrative division	Latitude	Longitude	N
MGL*	Mongolia	Övörkhangai	48.0	101.0	44
CHA*	China	Xinjiang	48.7	87.0	34
ALT*	Russia	Altai	51.6	85.0	40
KAL*	Russia	Kalmykia	47.5	45.3	22
KYR*	Kyrgyzstan	Naryn	41.1	75.7	20
KSW*	Kazakhstan	Mangystau	42.3	53.2	24
KAR*	Kazakhstan	Kyzylorda	46.0	61.3	35
KKA*	Kazakhstan	Karagandy	50.0	73.0	25
UKL*	Ukraine	Lviv	50.3	30.9	21
UKZ*	Ukraine	Zakarpattia	49.2	23.6	18
LTA*	Lithuania	Vilnius	56.9	25.4	21
GGA*	Georgia	Samegrelo	42.3	42.3	24
CHY	China	Yunnan	24.8	103.3	33
BFG	China	Xinjiang	43.5	87.4	24
NUR	China	Xinjiang	41.6	82.9	23
YAR	China	Xinjiang	38.4	77.3	24
IND	India	Jammu and Kashmir	32.6	76.1	24

N=Sample size

The sampling strategy for Eastern Eurasia focused on animals from remote areas that were mainly used for everyday work. Interviews with horse owners from across the sampling area revealed that their working horses were predominantly bred locally. While I cannot rule out that there might be traces of admixture from foreign or western horses in the types of horses that were sampled, I am confident that the risk of having sampled individuals with high levels of admixture was reduced as far as is possible in horses. Sampling locations were chosen to be evenly distributed across Eastern Eurasia. Great care was taken to obtain a representative sample of the genetic variability within sampling locations. To achieve this, samples were collected in different villages and towns with a maximum geographic distance of 100km between any two horses from the same sampling location. Horses from the same sampling location are henceforth referred to as populations, although it is appreciated that this term might not apply in the strict biological sense.

Sampling was performed by plucking approximately 50 hairs from the manes of individual horses, provided their owners gave their (verbal) consent. In cases where several horses were owned by the same person, special care was taken not to sample related individuals. The sampling procedure described here was approved by the ethics committee of the University of Cambridge as non-regulated (approval code 10/Z03).

Microsatellite markers

A total of 26 microsatellite loci with known genomic assignments were amplified in two multiplex reactions (Table 2.2) using the Type-it Microsatellite PCR kit from Qiagen. PCR amplification was carried out in a total volume of 12.5µl, with 9µl of the Type-it master mix, 1µl template DNA, 1.25µl Q-solution, and 1.25µl of a 1:10 dilution of primer mix. PCR reactions were performed on a thermal cycler under the following cycling conditions: 95°C for 6 min; 32 cycles of 95°C for 30 sec, 58°C for 90 sec, 72°C for 30 sec; 60°C for 30 min. Quality control measures included independent amplification and typing of each sample until the same genotype was obtained at least twice (multiple-tubes approach, Taberlet 1996) as well as checking for potential genotyping errors (presence of null alleles, large allele dropout, stuttering) using MICROCHECKER (Van Oosterhout *et al.* 2004).

Table 2.2

Summary of microsatellite markers used in Chapters 3, 4, and 5.

Locus	ECA	Primer 5'-3'	Multiplex	Size range	Dye	Reference
VHL20	30	CAAGTCCTCTTACTTGAAGACTAG AACTCAGGGAGAATCTTCCTCA	1	82-102	FAM	(Van Haeringen *et al.* 1994)
HTG4	9	CTATCTCAGTCTTGATTGCAGGAC GCTCCCTCCCTCCCTCTGTTCTC	1	123-137	FAM	(Ellegren *et al.* 1992)
AHT4	24	AACCGCCTGAGCAAGGAAGT GCTCCCAGAGAGTTTACCCT	1	151-169	FAM	(Binns *et al.* 1995)
HMS7	1	CAGGAAACTCTCATGTTGATACCATC GTGTTGTTGAAACATACCTTGACTGT	1	172-186	FAM	(Guérin *et al.* 1994)
COR18	25	AGTCTGGCAATATTGAGGATGT AGCAGCTACCCTTTGAATACTG	1	263-277	FAM	(Hopman *et al.* 1999)
AHT5	8	ACGGACACATCCCTGCCTGC GCAGGCTAAGGAGGCTCAGC	1	125-141	VIC	(Binns *et al.* 1995)
HMS6	4	CTCCATCTTGTGAAGTGTAACTCA GAAGCTGCCAGTATTCAACCATTG	1	159-171	VIC	(Guérin *et al.* 1994)
ASB23	3	ACATCCTGGTCAAATCACAGTCC GAGGGCAGCAGGTTGGGAAGG	1	183-215	VIC	(Breen *et al.* 1997)
TKY312	6	AACCTGGGTTTCTGTTGTTG GATCCTTCTTTTTATGGCTG	1	100-126	FAM	(Tozaki *et al.* 2001a)
TKY343	11	TAGTCCCTATTTCTCCTGAG AAACCCACAGATACTCTAGA	1	143-173	NED	(Tozaki *et al.* 2001b)
LEX33	4	TTTAATCAAAGGATTCAGTTG GGGACACTTTCTTTACTTTC	1	191-217	NED	(Coogle *et al.* 1996)
HMS3	9	CCAACTCTTTGTCACATAACAAGA GCCATCCTCACTTTTTCACTTTGTT	1	151-171	PET	(Guérin *et al.* 1994)
COR58	12	CACCAGGCTAAGTAGCCAAG GGGAAGGACGATGAGTGAC	1	210-234	PET	(Ruth *et al.* 1999)
HMS5	5	TAGTGTATCCGTCAGAGTTCAAGG GCAAGGAAGTCAGACTCCTGGA	2	98-104	FAM	(Guérin *et al.* 1994)
EB2E8	26	TTCTGTGTTAGGGGTTGTG GTATGAGCCAGTTCTTGAT	2	125-139	FAM	(Gralak *et al.* 1994)
TKY321	20	TTGTTGGGTTTAGGTATGAAGG GTGTCAATGTGACTTCAAGAAC	2	182-208	FAM	(Tozaki *et al.* 2001a)
ASB2	15	CACTAAGTGTCGTTTCAGAAGG GCACAACTGAGTTCTCTGATAGG	2	216-248	FAM	(Breen *et al.* 1997)
TKY301	23	AATGGTGGCTAATCAATGGG GTGTATGATGCCCTCATCTC	2	149-169	VIC	(Tozaki *et al.* 2001a)
TKY337	4	AGCAGGGTTTAATTACCGAG TAGATGCTAATGCAGCACAG	2	169-189	VIC	(Tozaki *et al.* 2001b)
TKY374	1	CTGGTCCCTCTGGATGGAAG TCCCAAGAGGGAGTACAATC	2	197-225	VIC	(Tozaki *et al.* 2001a)
HTG7	4	CCTGAAGCAGAACATCCCTCCTTG ATAAAGTGTCTGGGCAGAGCTGCT	2	113-123	VIC	(Marklund *et al.* 1994)
UM11	20	TGAAAGTAGAAAGGGATGTGG GTCTCAGAGCAGAAGTCCCTG	2	162-184	NED	(Meyer *et al.* 1997)
TKY394	24	GCATCATCGCCTTGAAGTTG CCTTTCTGGTTGGTATCCTG	2	232-258	NED	(Tozaki *et al.* 2001b)
UM32	14	AAATGGTCAGCCTCTCCTC TGTCTCTCTAGTCCCACTCCTC	2	140-150	PET	(Swinburne *et al.* 2000)
HMS1	15	CATCACTCTTCATGTCTGCTTGG TTGACATAAATGCTTATCCTATGGC	2	170-182	PET	(Guérin *et al.* 1994)
TKY294	27	GATCTATGTGCTAGCAAACAC CTAGTGTTTCAGATAGCCTC	2	216-230	PET	(Tozaki *et al.* 2001a)

ECA: location on *Equus caballus* chromosome

2.1.2 Traditional breeds from Western Eurasia

Geographic scope

Western Eurasia, in this study, is defined as the geographic region stretching from the Iberian Peninsula in the West to Poland in the East, and including Turkmenistan and Iran in the southeast.

Sampling

The dataset consists of 1167 horses from 24 traditional breeds. Samples for this dataset were obtained through different sources: four previously published microsatellite genotyping datasets (Cañón *et al.* 2000; Ząbek *et al.* 2005; Glowatzki-Mullis *et al.* 2006; Luís *et al.* 2007) including reference samples were obtained from the corresponding authors of these four studies. Samples included in these datasets had all been genotyped using the same set of twelve microsatellites (Table 2.3). To increase sample size I genotyped an additional 258 individuals from eight breeds (Camargue, Caspian, Highland Pony, Hucul, Altmark Draught, Noriker, Posavina and Schleswig Draught, Table 2.3, "new data") and aligned the resulting dataset with the four published datasets using between four and seven reference samples for each published dataset.

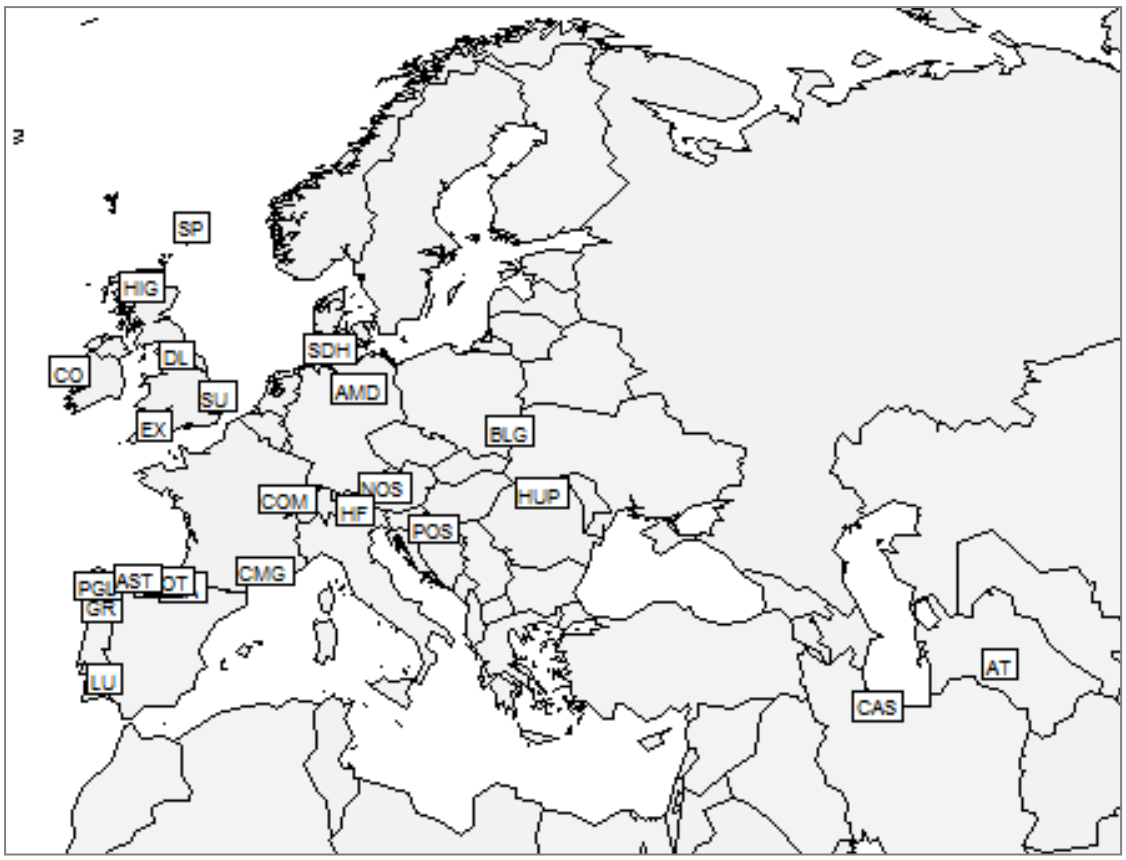

Figure 2.2

Sampling locations in Western Eurasia. AMD – Altmark Draught, AST – Asturcón, AT- Akhal Teke, BLG – Bilgoraj, CAS – Caspian, CMG – Camargue, CO – Connemara, COM – Comtois, DL – Dales, EX – Exmoor Pony, GR – Garrano, HF – Haflinger, HIG – Highland Pony, HUP – Hucul, JNA – Jaca Navarra, LOS – Losino, LU – Lusitano, NOS – Noriker, PGL – Caballo Gallego, POS – Posavina, POT – Pottoka, SDH – Schleswig Draught, SP – Shetland Pony, SU – Suffolk Punch.

Table 2.3

Traditional European horse breeds included in Chapter 6.

ID	Breed	Country of origin	N	Reference
AT	Akhal Teke	Turkmenistan	55	(Luís *et al.* 2007)
CO	Connemara	Ireland (west)	45	(Luís *et al.* 2007)
DL	Dales	England (north)	42	(Luís *et al.* 2007)
EX	Exmoor	England (southwest)	98	(Luís *et al.* 2007)
GR	Garrano	Portugal	37	(Luís *et al.* 2007)
HF	Haflinger	Austria (Tyrol)	45	(Luís *et al.* 2007)
LU	Lusitano	Portugal	52	(Luís *et al.* 2007)
SP	Shetland Pony	Scotland (Shetland Islands)	36	(Luís *et al.* 2007)
SU	Suffolk Punch	England (southeast)	41	(Luís *et al.* 2007)
COM	Comtois	France (east)	33	(Glowatzki-Mullis *et al.* 2006)
AST	Asturcón	Spain (north)	119	(Cañón *et al.* 2000)
JNA	Jaca Navarra	Spain (northwest)	122	(Cañón *et al.* 2000)
LOS	Losino	Spain (north)	66	(Cañón *et al.* 2000)
PGL	Caballo Gallego	Spain (northwest)	72	(Cañón *et al.* 2000)
POT	Pottoka	Basque Country	51	(Cañón *et al.* 2000)
AMD	Altmark Draught	Germany (east)	31	New data
CAS	Caspian Horse	Iran	30	New data
CMG	Camargue	France (south)	22	New data
HIG	Highland Pony	Scotland	25	New data
HUP	Hucul	Carpathian Mountains	17	New data
POS	Posavina	Croatia	24	New data
SDH	Schleswig Draught	Germany (north)	22	New data
NOS	Noriker	Austria	26	New data
BLG	Bilgoraj	Poland	28	(Ząbek *et al.* 2005)

N = sample size

Microsatellite markers

The 258 new samples (indicated by "new data" in Table 2.3) were genotyped at the same set of 12 markers as the samples in the published datasets using three multiplex reactions (Table 2.4).

Table 2.4

Summary of microsatellite markers used in Chapter 6.

Locus	ECA	Primer 5'-3'	Multiplex	Size range	Dye	Reference
AHT4	24	AACCGCCTGAGCAAGGAAGT GCTCCCAGAGAGTTTACCCT	1	151-169	FAM	(Binns *et al.* 1995)
AHT5	8	ACGGACACATCCCTGCCTGC GCAGGCTAAGGAGGCTCAGC	1	123-141	VIC	(Binns *et al.* 1995)
HMS3	9	CCAACTCTTTGTCACATAACAAGA GCCATCCTCACTTTTTCACTTTGTT	1	149-173	PET	(Guérin *et al.* 1994)
HMS6	4	CTCCATCTTGTGAAGTGTAACTCA GAAGCTGCCAGTATTCAACCATTG	1	159-173	VIC	(Guérin *et al.* 1994)
HMS7	1	CAGGAAACTCTCATGTTGATACCATC GTGTTGTTGAAACATACCTTGACTGT	1	170-188	FAM	(Guérin *et al.* 1994)
HTG4	9	CTATCTCAGTCTTGATTGCAGGAC GCTCCCTCCCTCCCTCTGTTCTC	1	121-139	FAM	(Ellegren *et al.* 1992)
VHL20	30	CAAGTCCTCTTACTTGAAGACTAG AACTCAGGGAGAATCTTCCTCA	1	82-102	FAM	(Van Haeringen *et*
ASB2	15	CACTAAGTGTCGTTTCAGAAGG GCACAACTGAGTTCTCTGATAGG	2	214-250	FAM	(Breen *et al.* 1997)
HTG7	4	CCTGAAGCAGAACATCCCTCCTTG ATAAAGTGTCTGGGCAGAGCTGCT	2	113-123	NED	(Marklund *et al.* 1994)
HMS2	10	CTTGCAGTCGAATGTGTATTAAATG ACGGTGGCAACTGCCAAGGAAG	3	213-237	NED	(Guérin *et al.* 1994)
HTG10	21	CAATTCCCGCCCCACCCCCGGCA GTTTTTATTCTGATCTGTCACATTT	3	96-118	NED	(Marklund *et al.* 1994)
HTG6	15	CCTGCTTGGAGGCTGTGATAAGAT GTTCACTGAATGTCAAATTCTGCT	3	85-107	VIC	(Ellegren *et al.* 1992)

ECA: location on the *Equus caballus* chromosome

2.2 DNA extraction

Hair samples were stored in acid-free envelopes until used. Genomic DNA was extracted from hair. Extraction buffer for 100 samples was made up as follows: 7572 µl of ultrapure (MQ) water were added to a 15 ml screw top tube and warmed to 60 ºC. 60µl of Nonidet NP40 and 50µl of Tween were added to the water and the tube was inverted to allow mixing of the reagents. Next, 500µl $MgCl_2$ and 500µl 10x buffer (both BIOTAQ, Bioline Inc.) and 318µl Proteinase K (Sigma Aldrich) were added; the tube was again inverted to allow mixing. Extractions were carried out in 1.5 ml screw top eppendorf tubes containing 15-20 hairs per horse and 90µl extraction buffer. Samples were briefly centrifuged at 1000

rpm before being placed in a water bath at 60ºC, where they were incubated for 45 minutes. Following incubation, samples were placed in a hotblock at 95ºC for 15 minutes to terminate the reaction.

DNA extracts were purified (QIAquick purification kit, Qiagen) and the concentration of the recovered DNA was determined by measuring its OD 260 with a NanoDrop spectrophotometer. Purified extracts were standardised to a concentration of 10ng of DNA/µl.

2.3 Microsatellite analysis

Reverse primers were modified by end-labelling with fluorescent dyes (FAM, VIC, NED, PET) at the 5′-end (for details see Tables 2.2 & 2.4). Microsatellite alleles were separated by electrophoresis on a 5% denaturing polyacrylamide gel in an ABI PRISM 3730 automated DNA Sequencer (Applied Biosystems) and genotypes were scored using GeneMapper Software v37 (Applied Biosystems). The programme FlexiBin v.2 (Amos *et al.* 2007) was used to aid in binning decisions.

2.4 Statistical analysis

Within-population (-breed) genetic diversity was calculated as expected heterozygosity (H_e, unbiased estimator, Nei 1987), observed heterozygosity (H_o), number of private alleles per population (U), and allelic richness (R_S). Estimates of allelic richness were standardised to the smallest sample size in the respective datasets. H_e and R_S were calculated in FSTAT v 2.9.3.2 (Goudet 2001, updated from Goudet 1995), H_o, and U were estimated using GDA (Bohonak 2002).

Genetic differentiation was calculated using two measures, Weir and Cockerham's (Weir & Cockerham 1984) theta, an estimator of Wright's *F*st, (Wright 1943a), and D_{EST}, the unbiased estimator of Jost's D (Jost 2008). The method of Weir & Cockerham (Weir & Cockerham 1984) uses an ANOVA approach to estimate within- and among-population variance components, which are then used to estimate their F_{ST} analogue ϑ.

F-statistics, and in particular F_{ST}, provide important insights into the evolutionary processes that influence the structure of genetic variation within and among populations, which is why it is still one of the most widely used indices of population substructure. F_{ST} and its analogues have been shown to frequently underestimate genetic differentiation at highly variable loci (Hedrick 2005; Jost 2008), I therefore also calculated the harmonic mean of Jost's D_{EST} (Jost 2008), which has been shown to more accurately reflect actual genetic differentiation between populations when highly variable markers are used (Leng & Zhang 2011). The main difference between F_{ST} and D_{EST} is that the former measures deviations from panmixia, whereas the latter measures deviations from total differentiation (Whitlock 2011). The harmonic mean of D_{EST} across all loci was estimated using the online program SMOGD version 1.2.5 (Crawford, 2010; http://www.ngcrawford.com/django/jost/; accessed 10 May 2011)

The statistical significance of the F_{ST} values was assessed using permutation tests; linkage disequilibrium was tested between all pairs of loci over all populations/breeds, and deviations from Hardy-Weinberg equilibrium (HWE) were tested both within (F_{IS} as test statistic) and over all (F_{IT} as test statistic) populations using permutation tests. All permutation tests were performed in FSTAT v 2.9.3.2 using Bonferroni corrections to account for multiple testing. The harmonic mean of Jost's D_{EST} (Jost 2008) between pairs of populations was calculated using the algorithm implemented in the online program SMOGD version 1.2.5 (Crawford 2009; http://www.ngcrawford.com/django/jost/; accessed June 2009). Unbiased estimates of theta were calculated using the algorithm available under http://www.montana.edu/kalinowski, accessed May 2011.

2.5 Datasets used in the various Chapters

In Chapters 3 and 5, analyses are based on all 455 non-breed domestic horses sampled in eastern Eurasia (Figure 2.1). In Chapter 4, where I investigate the origin and spread of domestic horses, I focus on the northern latitudes of Eastern Eurasia because wild horses were absent from the southern latitudes (Figure 4.1). In Chapter 6, I use all the samples from Western Eurasia as defined in section 2.1.2 (Figure 2.2)

3 Genetic diversity in Eastern Eurasia

Abstract[1]

Many events in the history of eastern Eurasia are expected to have affected the genetic structure of domestic horse populations in this area, including the process of domestication itself, the initial spread of domestic horses, and subsequent movements associated with the use of horses for transportation and as objects of trade and prestige. We investigate levels of within- and between population genetic diversity in "non-breed horses" (working horses sampled in remote areas) from 17 locations in Asia and parts of Eastern Europe, using 26 autosomal microsatellite loci. Non-breed horses have not been subject to the same intensity of artificial selection and closed breeding as most breed animals and are thus expected to better reflect the population history of domestic horses. Despite geographic distances between sampling locations of between 300 and 7000 km, pairwise F_{ST} was very low (range: <0.001-0.033), suggesting historically high levels of gene flow. Our analyses of non-breed horses revealed a pattern of isolation by distance and a significant decline in genetic diversity (expected heterozygosity and allelic richness) from East to West, consistent with a westward expansion of horses out of East Asia. While the timing of this putative expansion is unclear, our results highlight the benefit of studying animals that do not belong to particular breeds when investigating aspects of a population's history.

[1] A version of this chapter has been accepted for publication in Animal Genetics

3.1 Introduction

Domestic horses shaped the history of Eastern Eurasia like no other domestic animal, having been the main means of transportation in war and peace, as well as highly prized objects of trade and prestige. At the same time, historically important events in this vast region, such as the domestication process itself, the initial spread of domestic horses and subsequent major population movements, are expected to have left genetic signatures which may have persisted in the genomes of horses from this region.

Previous molecular studies investigating the early history of domestic horses have focused on the use of mitochondrial DNA (mtDNA). mtDNA sequencing of Eurasia-wide populations of domestic horses has revealed exceptionally high levels of matrilineal diversity in the domestic horse gene pool (Lister 2001; Vilà *et al.* 2001; Jansen *et al.* 2002; McGahern *et al.* 2006; Cieslak *et al.* 2010). While suggesting that more than one wild population contributed to the domestic horse gene pool, the generally weak phylogeographic structureof equine mtDNA haplogroups and their more or less homogeneous distribution in Eurasia have made it difficult to investigate more detailed aspects of the genetic history of domestic horses.

In the absence of highly divergent, highly localised mtDNA haplogroups, autosomal markers have proven a powerful alternative to mtDNA in elucidating the genetic history of domestic animals (Boyko *et al.* 2009). Large-scale patterns of autosomal genetic diversity, in particular, have provided important insights into the origins and routes of spread of a number of organisms, including domestic animals (Hanotte *et al.* 2002; Cymbron *et al.* 2005; Laloë *et al.* 2010; Warmuth *et al.* 2011). The geographic distribution of autosomal diversity in European taurine cattle, for instance, suggests that the spread of early domestic cattle in Europe followed two distinct routes (Cymbron *et al.* 2005), consistent with archaeological data.

In this chapter, I investigate the distribution of genetic diversity in horses from a large geographic area covering Eastern and northeastern Europe, Central Asia, East Asia, and parts of South Asia.

In order to minimize the strong genetic signatures of processes associated with breed formation and breed development, such as inbreeding, selective breeding, and/or breed mixing, I focus on local types of horses that are mainly used for everyday work. Following the definition by Clutton-Brock (1999), I understand breeds as being "groups of animals that are selected by humans to possess a uniform appearance that is inheritable and distinguishes it from other groups of animals within the same species". This definition implies that breeds (sensu Clutton-Brock 1999) are subjected to directional selection towards a common breeding goal. While the horse populations investigated here may be subject to directional selection, breeding goals are set at the level of individual owners and breeders and thus vary considerably within local horse populations. I therefore refer to the samples in this, and the two following chapters as "non-breed horses". The dataset presented in this chapter will provide the basis for Chapters 4 and 5, in which I will explore the origin and spread of horse domestication in the Eurasian steppes and post-domestication population movements, respectively.

3.2 Materials and methods

3.2.1 Sampling locations

The sampling locations included in this chapter are shown in Figure 3.1. Further details on the sampling procedure and the sampling locations can be found in section 2.1.1 and in Table 2.1, respectively.

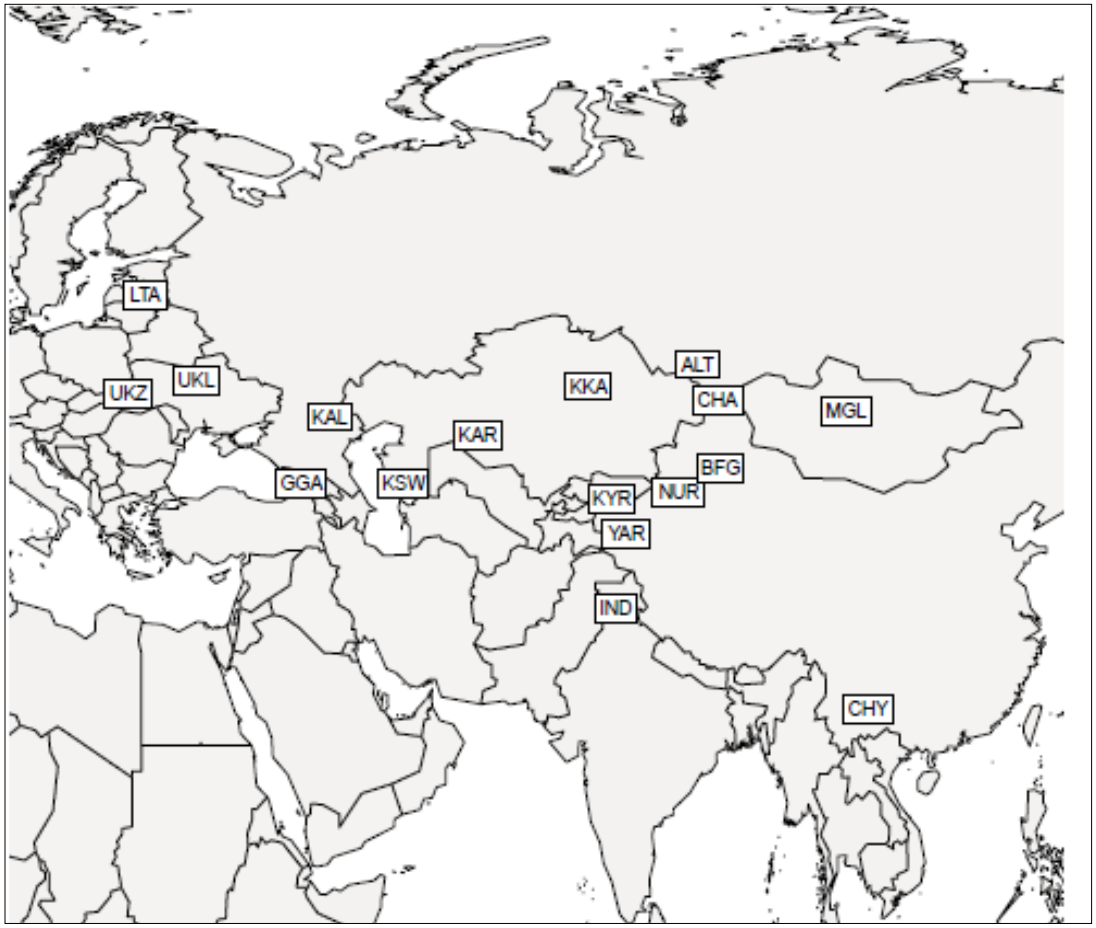

Figure 3.1

Geographic distribution of sampling locations. ALT – Russia, Altai, BFG – China, Xinjiang, CHA, China, Xinjiang, CHY – China, Yunnan, GGA – Georgia, Samegrelo and Zemo Svaneti, IND – India, Jammu and Kashmir, KAL –Russia, Kalmykia, KAR – Kazakhstan, Kyzylorda, KKA – Kazakhstan, Karagandy, KSW – Kazakhstan, Mangystau, KYR – Kyrgyzstan, Naryn, LTA – Lithuania, MGL – Mongolia, Övörkhangai, NUR – China, Xinjiang, UKL – Ukraine – Lviv, UKZ – Ukraine, Zakarpattia, YAR – China, Xinjiang (See also Table 2.1).

3.2.2 DNA extraction and microsatellite analysis

The DNA extraction protocol for hair and the DNA amplification protocol can be found in Chapter 2 (General methods), sections 2.2 and 2.3, respectively. Details of the markers used in this study can be found in Table 2.2

3.2.3 Genetic diversity

Genetic diversity within the 17 populations was calculated as expected heterozygosity (H_e, unbiased estimator, Nei 1987), observed heterozygosity (H_o), number of private alleles (U), and allelic richness (R_S). H_e and R_S were calculated in FSTAT v 2.9.3.2 (Goudet 2001, updated from Goudet 1995), H_o and U were estimated in GDA (Bohonak 2002). Estimates of allelic richness were standardised to the smallest sample size in this dataset, N=15 using the rarefaction algorithm implemented in FSTAT. For more details see Chapter 2 (General Methods).

3.2.4 Pairwise genetic differentiation and isolation by distance

Genetic differentiation between populations was estimated using Weir & Cockerham's (Weir & Cockerham 1984) estimator of F_{ST} (Wright 1943) and the harmonic mean of D_{EST} (Jost 2008) across all loci. Statistical significance of F_{ST} was tested using permutation tests and 10,000 randomisations. The presence of a pattern of isolation by distance (Slatkin 1993) was assessed by regressing geographical distance between populations against pairwise estimates of F_{ST} (linearised through $F_{ST}/(1-F_{ST})$), using the *mantel* function in the R library *vegan* (Oksanen *et al.* 2011) and 10,000 permutations. Geographic distances were measured as great-circle distances [the shortest geographic distance connecting two locations on a sphere], not Euclidean distances. Great-circle distances between sampling locations were computed using the *fields* package (Furrer *et al.* 2010) in R (R Development Core Team 2010). For more details see Chapter 2 (General Methods).

3.2.5 Population genetic structure

The genetic structure of non-breed domestic horses in Eastern Eurasia was investigated using two complementary methods, spatial Bayesian clustering and ordination in reduced space (multivariate analysis). Bayesian clustering methods have proven powerful analytical tools for identifying genetic structure in data sets; however, they assume that populations are in the Hardy–Weinberg equilibrium (HWE) and that loci are in linkage equilibrium (LE), assumptions which are often violated. Multivariate analyses on the other hand are robust to deviations from HWE and LE because they do not rely on underlying population genetic models.

Bayesian clustering analysis was carried out using the spatially explicit LOCPRIOR model (Hubisz *et al.* 2009), the ADMIXTURE model, and the CORRELATED FREQUENCIES model (Falush *et al.* 2003) implemented in the software program STRUCTURE v. 2.3 (Pritchard *et al.* 2000; Falush *et al.* 2003; Hubisz *et al.* 2009). STRUCTURE uses multilocus data to probabilistically assign individuals to *K* genetic groups by minimising Hardy-Weinberg disequilibrium between the groups and linkage disequilibrium between loci within the groups.

I performed preliminary runs to estimate the likelihood for all *K* values from 1 to 17. Likelihood decreased for values of *K* greater than 5 (data not shown); I therefore ran 10 chains for each *K* value 1 to 6, using a burn-in period of 500,000 iterations and 10^6 MCMC iterations. I assumed a uniform prior for α (the parameter describing the degree of admixture), and set alphapropsd (the standard deviation of the proposal distribution) to 0.05 to enable greater admixture. λ, the parameter describing the strength of the correlation of the parental allele frequencies, was set to 0.7, as estimated by the model during preliminary runs. The prior of the F_{ST} as estimated by the algorithm was set to default values. Convergence of likelihood, F_{ST}, and α was monitored throughout the runs. The most likely number of clusters was inferred from the mean estimated log probability of the data (Pritchard *et al.* 2000) and its second-order rate of change (Δ*K*) (Evanno *et al.* 2005).

Given the spatial scale considered in this study, I expected a pattern of isolation by distance. Simulation studies have shown that Bayesian clustering algorithms, including that implemented in STRUCTURE, can overestimate the number of distinct clusters when allele frequencies vary gradually across the study area, i.e. when there is isolation by distance (IBD) (Pritchard *et al.* 2009). To validate the results of the Bayesian clustering analysis, I performed discriminant analysis of principal components [DAPC, (Jombart *et al.* 2010)], which has been shown to recover complex patterns of population subdivision, including clinal patterns (Jombart *et al.* 2010).

Among the large number of different multivariate analyses, the best between-population differentiation is achieved by discriminant analysis (DA). However, DA requires variables in a dataset (alleles) to be uncorrelated and less than the number of observations (Jombart *et al.* 2010), which is usually not the case with genetic data. DAPC involves transforming the dataset through principal component analysis (PCA) before submitting it to discriminant analysis (DA), which ensures that the data fulfils the above mentioned requirements (Jombart *et al.* 2010).

DAPC analyses were performed using the *adegenet* package (Jombart 2008) in R version 2.10.1 (R Development Core Team 2010). The function *optim.a.score* was used to determine the number of PC axes that explain the largest amount of the total genetic variability in the dataset whilst achieving maximum discrimination among populations at the same time. DAPC was run retaining 150 PC axes, which accounted for 97.7 % of the total genetic variability. The number of DA axes retained was set to 16 to capture the maximum amount of variability contained in our dataset. The number of clusters in the dataset was estimated using sequential *K*-means clustering [(Legendre & Legendre 1998); function *find.clusters*] for K_{DAPC}=1 through to K_{DAPC} =50. The Bayesian Information Criterion (BIC) was used to determine the optimal number of clusters as the minimum number of clusters after which the BIC either increased or decreased by a negligible amount. DAPC axes were plotted using the function *scatter.dapc*.

3.3 Results

3.3.1 Genetic diversity

All sampling locations were in Hardy-Weinberg equilibrium at all loci except for locus EB2E8, which significantly deviated from Hardy-Weinberg expectations in three of the 17 sampling locations ($p<0.00011$, the adjusted α-value at the 0.05 level following Bonferroni correction). Microchecker (Van Oosterhout *et al.* 2004) results indicated that the deviations from HWE in locus EB2E8 were most likely due to null alleles [see also (Glowatzki-Mullis *et al.* 2006)], with evidence for null alleles in 14 of the 17 sampling locations. EB2E8 was therefore removed from further analyses. Markers UM11, VHL20, HMS7, and HMS6 showed evidence for null alleles in one, TKY337, HMS3, and LEX33 in two, and UM32 in three sampling locations. Out of 325 pairwise combinations, three pairs of loci showed significant deviations from linkage equilibrium at the adjusted α-value of $p=0.00015$: TKY321 x ASB2, ASB2 x TKY374, and TKY374 x TKY394.

The observed number of alleles per locus ranged between 4 (HMS5) and 17 (TKY343); the observed number of alleles per population ranged between 168 (UKL) and 219 (MGL); the observed number of private alleles per population ranged between 0 (KAR, KAL, UKL, UKZ, and LTA) and 5 (KYR). Mean expected heterozygosity was 0.784, ranging between 0.756 (LTA) and 0.797 (BFG). Mean allelic richness was 6.87, ranging between 6.45 (LTA) and 7.16 (BFG) (Table 3.1).

Table 3.1

Summary statistics for the 17 populations analysed in this chapter

Population origin	ID	N	H_e	H_o	Rs	U	F_{IS}
Mongolia	MGL	44	0.795	0.790	7.12	2	0.007
China	CHA	34	0.793	0.767	7.12	1	0.032
China	BFG	24	0.797	0.787	7.16	1	0.012
China	NUR	23	0.779	0.767	6.92	2	0.016
China	CHY	20	0.793	0.785	6.94	3	0.010
China	YAR	24	0.801	0.789	6.96	1	0.014
Russia	ALT	40	0.784	0.779	6.71	2	0.007
Kyrgyzstan	KYR	20	0.789	0.798	7.00	5	-0.012
Kazakhstan	KKA	25	0.784	0.787	7.00	1	-0.004
Kazakhstan	KAR	35	0.773	0.773	6.69	0	-0.001
Kazakhstan	KSW	24	0.782	0.790	7.10	1	-0.011
Russia	KAL	22	0.782	0.783	6.86	0	-0.002
Georgia	GGA	24	0.783	0.765	6.74	3	0.023
India	IND	24	0.768	0.770	6.71	4	-0.002
Ukraine	UKL	15	0.786	0.741	6.56	0	0.057
Ukraine	UKZ	17	0.775	0.734	6.66	0	0.053
Lithuania	LTA	21	0.756	0.739	6.45	0	0.022

H_e expected heterozygosity (unbiased estimator, Nei 1987), H_o observed heterozygosity, R_s allelic richness, U number of private alleles, F_{IS} fixation index

The highest levels of expected heterozygosity were found in East Asia (MGL, CHA, CHY, KYR, YAR, BFG), whereas the lowest levels were found in Eastern Europe (UKZ), Northern Europe (LTA), and India (IND). The distribution of allelic richness followed a similar pattern (Table 3.1). Linear regression analyses between longitude and both expected heterozygosity and allelic richness revealed a significant decline in genetic diversity from East to West (H_e: R^2= 0.363, F= 8.564, df= 15, p= 0.010; R_S: R^2=0.481, F= 13.91, df= 15, p= 0.002, Figure 3.2 A&B). In contrast, no relationship was found between genetic diversity and latitude, regardless of the diversity measure used (H_e: R^2=0.095, F= 1.578, df=15, p= 0.228, R_S: R^2=0.093, F= 1.541, df= 15, p= 0.234, Figure 3.2 C&D).

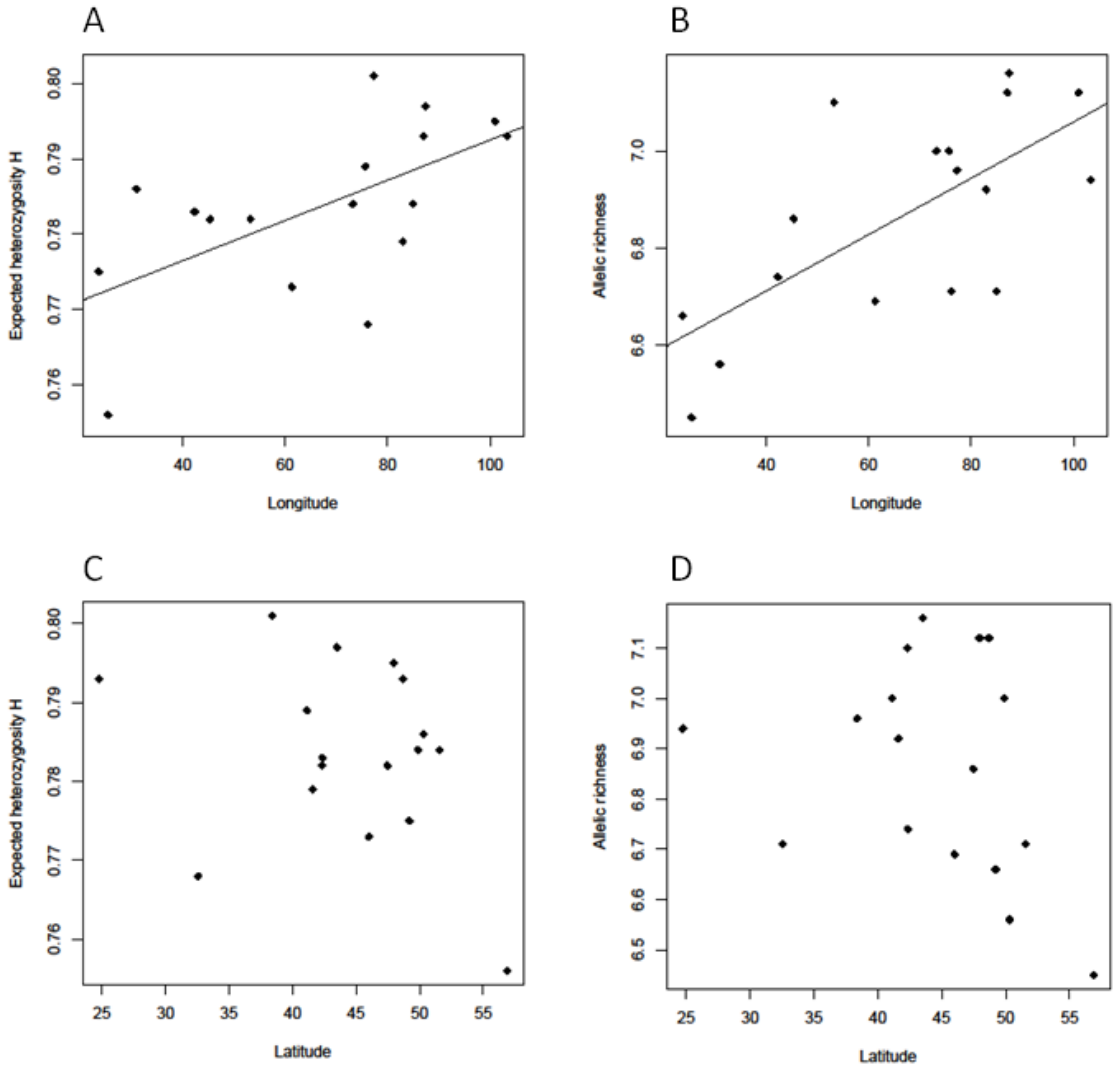

Figure 3.2

Regression of genetic diversity versus latitude and longitude, respectively. Regression of expected heterozygosity, H_e (A) and allelic richness, R_s (B) versus longitude. Regression of expected heterozygosity, H_e (C) and allelic richness, R_s (D) versus latitude.

3.3.2 Pairwise genetic differentiation and isolation by distance

The overall genetic differentiation, as measured by F_{ST}, was 0.013. Pairwise F_{ST} estimates ranged between 0 and 0.032, with only 81 out of the 136 pairwise comparisons (59%) significantly different from zero (Table 3.2). When genetic differentiation between populations was estimated using the harmonic mean of D across loci (Jost 2008), approximately 50% of the pairwise comparisons were between 0 and 0.025 higher than the corresponding pairwise F_{ST} estimates (mostly involving comparisons between CHY, CHA, KYR, KSW, KAR, KKA, ALT, KAL, and LTA), while the other half was between 0.008 and 0.001 lower (mostly involving comparisons between GGA, UKZ, UKL, NUB, BFG, YAR, and IND) (Table 3.2). Both estimators showed the greatest differentiation for pairwise comparisons involving horses from Lithuania (LTA). The relationship between genetic differentiation (linearised through $F_{ST}/(1-F_{ST})$) and geographic distance between sampling locations was weak, albeit significant (Mantel r=0.344, p=0.008; Figure 3.3).

Table 3.2

Pairwise estimates of F_{ST} [ϑ, Weir & Cockerham 1984)] (above diagonal), and pairwise estimates of the harmonic mean of D [$D_{EST,}$ Jost, 2008)] (below diagonal) based on 25 microsatellite markers.

	MGL	CHY	CHA	KYR	KSW	KAR	KKA	ALT	KAL	LTA	GGA	UKZ	UKL	NUB	BFG	YAR	IND
MGL		0.012	0.005^{ns}	0.006	0.012	0.016	0.01	0.017	0.02	0.019	0.011	0.016	0.016	0.009^{ns}	0.002^{ns}	0.002	0.012
CHY	0.016		0.01	0.014	0.013	0.022	0.016	0.022	0.03	0.022	0.014	0.021	0.019	0.017	0.008^{ns}	0.007^{ns}	0.019
CHA	0.008	0.01		0.004^{ns}	0.007^{ns}	0.009	0.009	0.011	0.015	0.019	0.003^{ns}	0.009^{ns}	0.006^{ns}	0.007^{ns}	0.001^{ns}	0.003^{ns}	0.011
KYR	0.006	0.016	0.002		0.007^{ns}	0.015	0.016	0.01	0.015	0.02	0.002^{ns}	0.009^{ns}	0.011^{ns}	0.004^{ns}	0.006^{ns}	0.003^{ns}	0.016
KSW	0.013	0.015	0.006	0.01		0.018	0.01	0.015	0.014^{ns}	0.022	0.004^{ns}	0.013^{ns}	0.006^{ns}	0.010^{ns}	0.012^{ns}	0.003^{ns}	0.012^{ns}
KAR	0.026	0.037	0.016	0.017	0.028		0.016	0.018	0.02	0.025	0.011	0.018	0.014	0.02	0.018	0.022	0.015
KKA	0.011	0.015	0.014	0.031	0.011	0.018		0.017	0.017	0.023	0.009^{ns}	0.017	0.016	0.009	0.009^{ns}	0.007^{ns}	0.015
ALT	0.032	0.047	0.017	0.013	0.016	0.027	0.027		0.011	0.02	0.01	0.011	0.008	0.017	0.016	0.014	0.018
KAL	0.028	0.044	0.025	0.012	0.009	0.028	0.026	0.008		0.032	0.007^{ns}	0.007^{ns}	0.007^{ns}	0.018	0.014	0.011^{ns}	0.012
LTA	0.035	0.031	0.034	0.036	0.041	0.039	0.03	0.033	0.048		0.017	0.022	0.024	0.029	0.016^{ns}	0.016	0.012
GGA	0.015	0.013	0.003	0	0.001	0.009	0.006	0.006	0.003	0.023		0.008^{ns}	0.000^{ns}	0.009^{ns}	0.008^{ns}	0.008^{ns}	0.005^{ns}
UKZ	0.014	0.021	0.007	0.006	0.005	0.025	0.013	0.007	0.004	0.021	0.003		0.000^{ns}	0.013^{ns}	0.012^{ns}	0.009^{ns}	0.012^{ns}
UKL	0.021	0.02	0.006	0.011	0.002	0.021	0.016	0.004	0.005	0.033	0	0		0.015^{ns}	0.014^{ns}	0.014	0.006^{ns}
NUB	0.011	0.019	0.004	0.001	0.008	0.019	0.01	0.031	0.023	0.041	0.004	0.006	0.012		0.002^{ns}	0.004^{ns}	0.016
BFG	0	0.004	0	0.006	0.011	0.027	0.005	0.022	0.014	0.018	0.004	0.008	0.014	0		0.000^{ns}	0.010^{ns}
YAR	0.001	0.007	0.002	0.002	0.001	0.04	0.006	0.022	0.006	0.019	0.007	0.005	0.015	0.001	0		0.010^{ns}
IND	0.012	0.018	0.011	0.016	0.009	0.014	0.02	0.018	0.011	0.011	0.001	0.006	0.004	0.015	0.007	0.007	

All pairwise F_{ST} estimates are significant at p≤ 0.05, except where indicated otherwise; ns=non-significant; p-values obtained after 136,000 permutations.

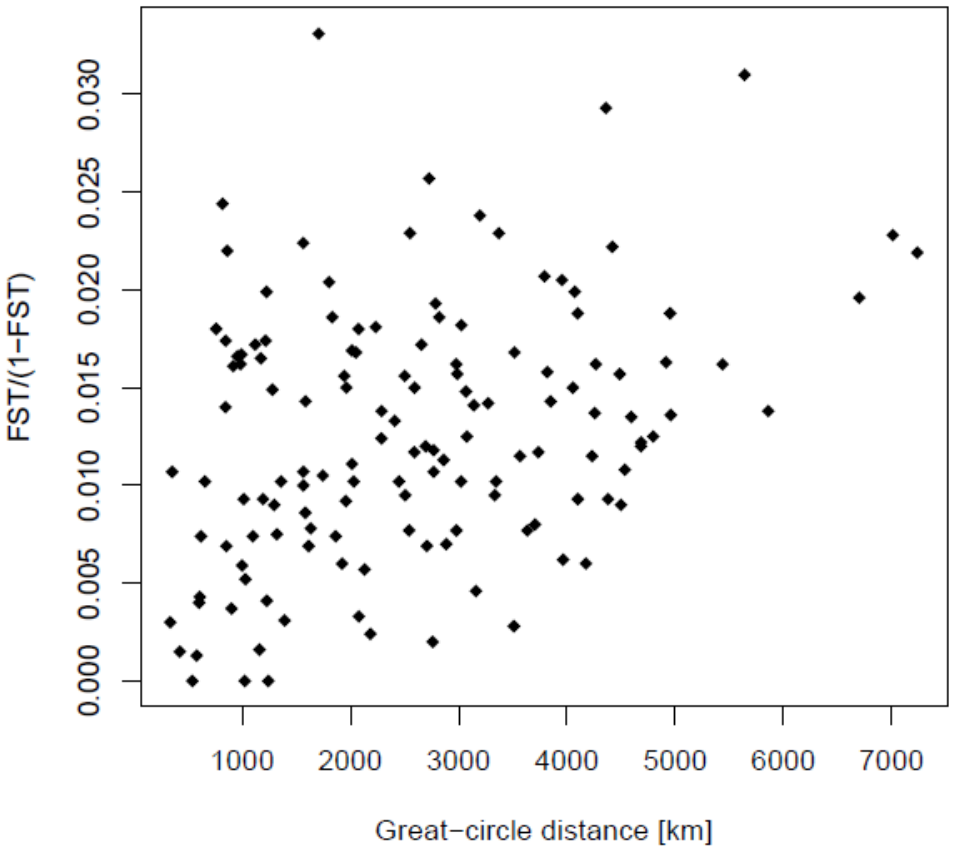

Figure 3.3

Scatterplots of linearised F_{ST} [(Weir & Cockerham 1984) calculated using FSTAT, vers. 1.2; (Goudet 1995)] against great-circle distances (km, calculated in R (R Development Core Team 2010)) between all pairs of populations.

3.3.3 Population genetic structure

STRUCTURE

Log-likelihoods increased steadily from K=1 to K=3, followed by a steep decrease; ΔD values suggested that the most likely number of clusters was three; however, even after more than 950,000 iterations, α values still ranged between 0.7 and 4.2, strongly indicating that there are no genetically distinct populations in the dataset (Pritchard *et al.* 2009) . Visual inspection of the STRUCTURE outputs suggested that allele frequencies varied continuously across the study region (data not shown). I therefore consider the presence of three distinct genetic clusters in this dataset an inaccurate representation of the actual population structuring.

DAPC

The eigenvalues of the analysis (Figure 3.5, inset) showed that most of the genetic structure was captured by the first two principal components. The first principal components axis (eigenvalue = 44.56) roughly aligned populations according to their geographic location along the longitudinal axis with the exception of horses from Lithuania (LTA) and horses from Kyzylorda Oblast in Kazakhstan (KAR), both of which cluster with horses further east than would be expected based on their geographical location (Figure 3.5). The second DAPC axis (eigenvalue = 28.49) slightly separated horses from Altai Krai in eastern Russia (ALT) from the rest (Figure 3.5). BIC values were lowest between K_{DAPC}=2 and K_{DAPC} =4; however, the minimum number of clusters after which BIC values decreased only by a negligible amount was K_{DAPC} =2 (Figure 3.4). At K_{DAPC} =2, all populations had membership proportions of between 17% and 63% in cluster 1, and between 38% and 83% in cluster 2, with the majority of populations from the western parts of our study area having membership proportions > 50% in cluster 1, and all populations from the eastern part of the study area having membership proportions > 50% in cluster 2; at K_{DAPC} =3 and K_{DAPC} =4, no biologically meaningful pattern could be discerned (data not shown).

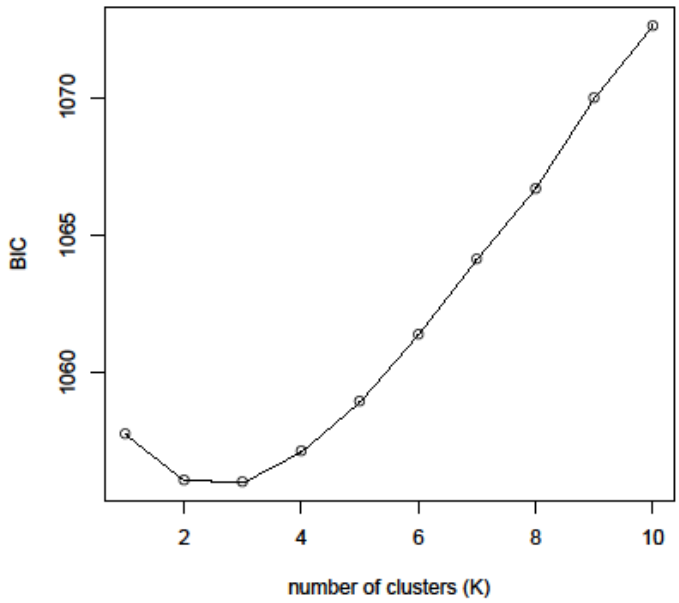

Figure 3.4

Bayesian information criterion (BIC) for K_{DAPC} between one and 10.

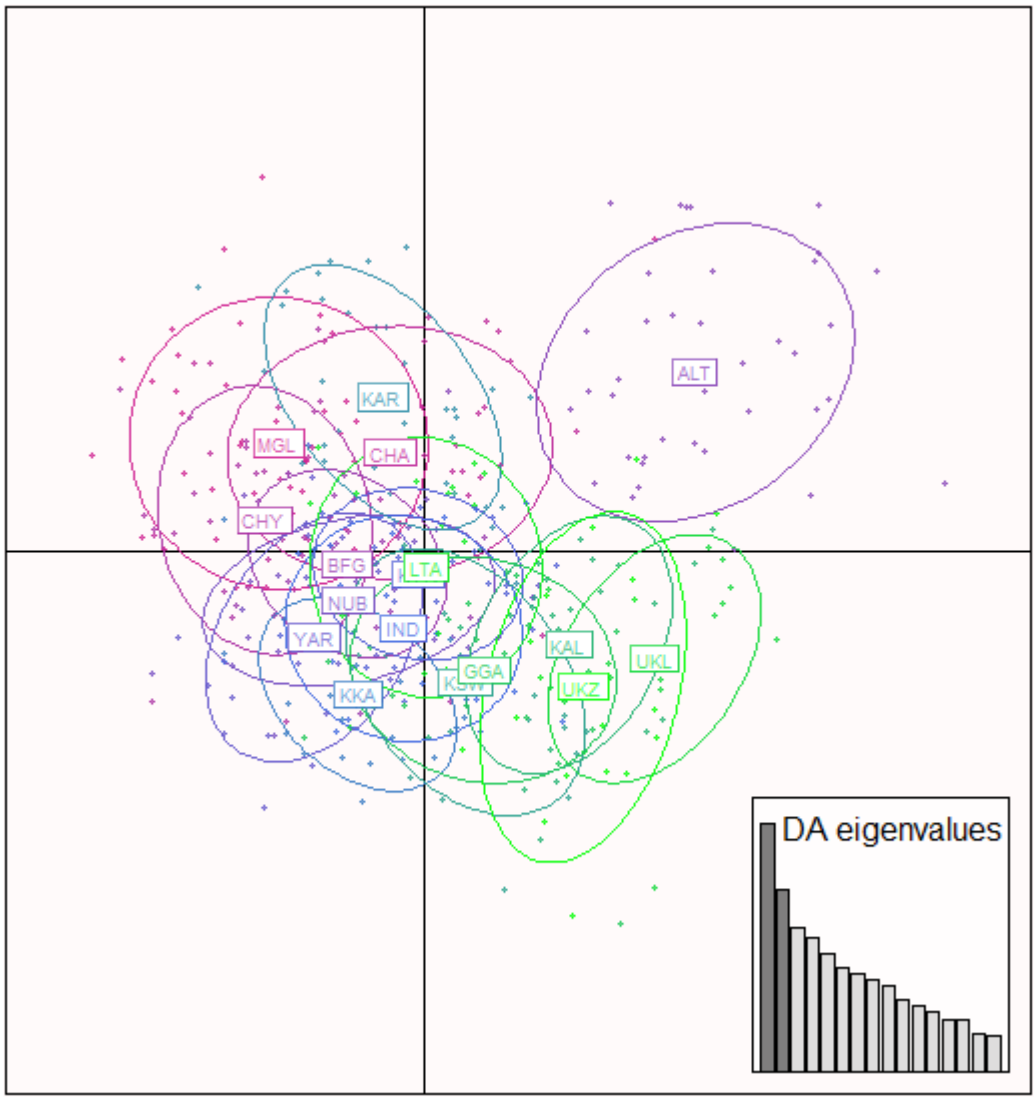

Figure 3.5

Scatterplot of the first two principal components (eigenvalues = 44.56 and 28.49, respectively) of the DAPC analysis using sampling locations as priors for genetic clusters. Populations are represented by coloured inertia ellipses the sizes of which were chosen to roughly include 2/3 of all the data points belonging to a specific sample. The inset shows the DA eigenvalues.

3.4 Discussion

Regression analyses revealed a significant relationship between genetic diversity (measured as both expected heterozygosity and allelic richness) and longitude, with a gradual decline in genetic diversity from East to West. Gradual declines in genetic diversity typically arise during range expansions as a consequence of serial bottlenecks at the expanding range margin (reviewed in Excoffier *et al.* 2009). A number of recent studies have linked observed gradual declines in genetic diversity to the initial colonisation of an area by a species (Prugnolle *et al.* 2005; Ramachandran *et al.* 2005; Linz *et al.* 2007; François *et al.* 2008; Deshpande *et al.* 2009; Tanabe *et al.* 2010). The colonisation of the world by anatomically modern humans has been shown to explain 85% of the observed variance in neutral genetic diversity (Prugnolle *et al.* 2005). Given that humans are known to have moved extensively throughout their history, this suggests that genetic signatures of colonisation are remarkably robust to subsequent movements.

According to the fossil record, the colonisation of Eurasia by horses occurred sometime in the second half of the Pleistocene epoch, when the wild ancestor of domestic horses, *Equus ferus* (Weinstock *et al.* 2005) migrated to Eurasia via the Bering landbridge (Azzaroli 1983). Between 0.117-0.186 million years ago, *Equus ferus* split into two lineages, one leading to modern day Przewalski's horses, *E. f. przewalskii*, and one leading to the extinct ancestor of domestic horses, *E. f. ferus* (Goto *et al*. 2011). Recents research suggests that the lineage leading to domestic horses was widely distributed in northern Eurasia (Lorenzen *et al*. 2011), and that wild horses from throughout this vast area may have been incorporated into the domestic horse gene pool (Cieslak *et al.* 2010). If domestic horses from different areas in Eurasia largely descend from the wild horse populations that were formerly found in these areas (genetic continuity), the observed east-to-west decline in genetic diversity might reflect the colonisation of Eurasia by *Equus ferus*. Alternatively, the observed pattern may have arisen as a consequence of one or several later east-to-west migrations involving large numbers of horses, for example those of the Huns and/or the Mongols. Modelling the parameters most compatible with this pattern will be required to estimate the timing of this putative expansion.

There was no significant correlation between genetic diversity and latitude, despite written accounts documenting annual imports of tens of thousands of steppe horses into India and historic China (Beckwith 1991; Gommans 1994). One explanation for the absence of a correlation between genetic diversity and latitude may lie in the fact that the bulk of horses in China originated from the steppes of central and east Asia [present-day Mongolia, Kyrgyzstan, southeast Kazakhstan, and northeast China, (Deng 1997)], whereas horses in India predominantly originated from west and southwest Asia [present day Kalmykia, southern Kazakhstan, or the Middle East (Chakravarti 1991; Gommans 1994)]. Due to the pronounced east-to-west decline in genetic diversity in the source region (northern Eurasia), horses in China are therefore descended from a genetically much more diverse stock than horses in India. Rather than by latitude, levels of genetic diversity in populations imported into those areas are thus predominantly determined by the diversity of their source populations, and thus, by longitude. However, the majority of populations in our study represent a very limited latitudinal range; more populations at more extreme northern and southern latitudes may be required to assess the relationship between genetic diversity and latitude in more detail.

Genetic differentiation between the populations investigated in this study was extremely low (average F_{ST} = 1.3%), despite geographic distances of between 300 and 7000 km between them. The observed low genetic differentiation probably reflects the combined effects of historically high levels of gene flow and both historically and currently low levels of genetic drift. Genetic drift is expected to be much lower in non-breed populations than in breed populations, owing to the absence of the dramatic population size reductions associated with breed formation, breed development and inbreeding. Pairwise F_{ST} values among 26 Chinese horse breeds separated by between 66 and 2907 km were shown to average 2.4% (Ling *et al.* 2010b)], which is already higher than that of the populations studied here, consistent with the classification of the former populations as breeds. The genetic differentiation among European horse breeds is higher still, with reported average F_{ST} values ranging between 7.8% (Cañón *et al.* 2000) and 10% (Ząbek *et al.* 2005).

The clinal structure of the dataset was confirmed by discriminant analysis of principal components (DAPC) and analysis of isolation by distance. DAPC aligned all populations according to their geographic positions along the longitudinal axis except for horses from the Altai Republic in East Asia (ALT) and horses from Lithuania (LTA). Horses from Altai cluster with horses from the Russian Republic of Kalmykia (KAL) and with horses from Ukraine (UKL and UKZ). Interestingly, all four populations in this cluster are from countries which used to be part of the Soviet Union. The observed genetic homogeneity of horse populations from former Soviet Union countries can be explained by the centralization of stock breeding into large breeding centres and farms as well as the existence of artificial insemination stations (Kosharov *et al.* 1989). While DAPC analysis indicated that horses from Lithuania (LTA) are genetically more related to Eastern horses than to horses from neighbouring populations in Ukraine (Figure 3.5), denser sampling would be required to resolve the genetic relationships between Lithuanian and Asian horse populations in greater detail.

By recovering a significant east-to-west decline in genetic diversity, my analysis of microsatellite markers in non-breed horses from across Eastern Eurasia reveals for the first time a clear genetic pattern in Eurasia-wide populations of domestic horses. The results presented in this chapter thus highlight the great resolving power of microsatellite markers and the potential benefits of focusing on non-breed animals when the history of domestic species prior to breed formation is of concern.

4 Domestication in the Eurasian steppe

Abstract[1]

Despite decades of research across multiple disciplines, the early history of horse domestication remains poorly understood. On the basis of current evidence from archaeology, mitochondrial DNA, and Y-chromosomal sequencing, a number of different domestication scenarios have been proposed, ranging from the spread of domestic horses out of a restricted primary area of domestication to numerous independent domestication events involving distinct wild horse populations, possibly as a consequence of the spread of technique. In this chapter, I investigate the origin and spread of horse domestication in the Eurasian steppes by fitting a spatially explicit stepping stone model to genotype data from more than 300 horses sampled across northern Eurasia. Because the wild progenitor of domestic horses, *Equus ferus*, is no longer extant, I first reconstruct the population genetic structure of *E. ferus*, and then infer the origin and mode of spread of horse domestication by testing explicit scenarios. I found strong evidence for an expansion of *E. ferus* out of East Asia about 160 kya, likely reflecting the colonisation of Eurasia by this species. The best-fitting scenario further suggests that horse domestication originated in the western part of the Eurasian steppe, and that domestic herds were repeatedly restocked with local wild horses as they spread out of this area. By showing that horse domestication was initiated in the western Eurasian steppe, and that the spread of domestic herds across Eurasia involved extensive introgression from the wild, the model of horse domestication proposed here integrates evidence from archaeology, mtDNA, and Y-chromosomal DNA.

[1] A version of this chapter has been submitted for publication.

4.1 Introduction

Investigating the origin and spread of horse domestication constitutes a major area of research in multiple disciplines, yet the key question of whether horse domestication occurred in a small number of geographically restricted areas or whether wild horse populations from across Eurasia were domesticated more or less independently, remains poorly understood (Cieslak *et al.* 2010). An increasing body of evidence from archaeology suggests that horses were first domesticated in the steppes of modern-day Ukraine and Kazakhstan (the western steppes, (Brown & Anthony 1998; for review see Olsen 2006a; Outram *et al.* 2009); however, there is so far no corroborating molecular evidence for a geographically restricted origin of horse domestication anywhere in the Eurasian steppe (Cieslak *et al.* 2010), a vast belt of grassland stretching from Hungary in the West to Mongolia in the East. Thus far, the only geographically restricted region that has been identified as a primary area of horse domestication is the Iberian Peninsula (Lira *et al.* 2010; Warmuth *et al.* 2011).

A related question concerns the mode of spread of horse domestication in the Eurasian steppe: did the spread of horse domestication involve a movement of domestic herds ("demic diffusion") (Anthony *et al.* 1986; Bökönyi 1993), as appears to have been the case in most other domestic animals (Laloë *et al.* 2010)? Or did pastoral communities throughout the steppes domesticate locally available wild populations after having acquired the knowledge of how to do so (Vilà *et al.* 2001; Jansen *et al.* 2002)?

Low levels of genetic variability in the paternally inherited Y-chromosome could reflect a demic spread of domestic herds out of a geographically restricted domestication origin (Lindgren *et al.* 2004). However, levels of mitochondrial DNA variability in domestic horses have been found to be exceptionally high (Lister 2001; Vilà *et al.* 2001; Jansen *et al.* 2002; McGahern *et al.* 2006; Cieslak *et al.* 2010), a pattern which has been widely interpreted as resulting from the domestication of multiple wild populations (Lister *et al.* 1998; Vilà *et al.* 2001; Jansen *et al.* 2002; Bruford *et al.* 2003; Olsen 2006a).

In this chapter, I parameterise a spatially and demographically explicit model with autosomal genetic data from more than 300 horses to reconstruct the origin and mode of spread of horse domestication in the Eurasian steppe. Assuming domestication occurred in a small number of geographically defined areas, the model used here allows me to trace the geographic origin of horse domestication, and to determine the relative roles of demic diffusion versus recruitment of local wild stock in the spread of horse domestication.

4.2 Materials and Methods

4.2.1 Sampling

Due to unsuitable climate, wild horses were likely absent in the Indian subcontinent and present-day China, except for the far north (Olsen 1988). For this study, I therefore excluded the Indian (IND) and all Chinese samples, except for CHA, which is from the far northwest of China. The dataset used here consists of 322 randomly sampled non-breed horses from 12 sampling areas spanning 8 countries (Figure 4.1). For more details on the samples see Table 4.1.

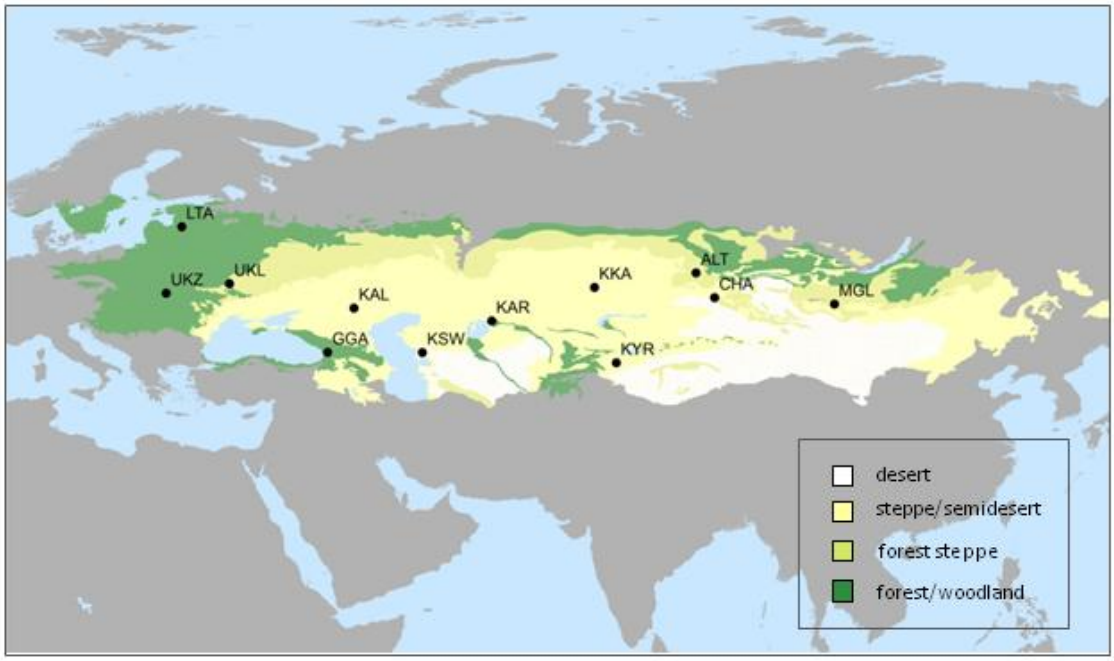

Figure 4.1

Geographic distribution of sampling location and the model. A) LTA-Lithuania, UKZ-Ukraine, Zakarpattia Oblast, UKL – Ukraine, Lviv Oblast, KAL – Russia, Republic of Kalmykia, GGA – Georgia, KSW – Kazakhstan, Mangystau Oblast, KAR – Kazakhstan, Kyzylorda Oblast, KKA – Kazakhstan, Karagandy Oblast, KYR – Kyrgyzstan, Naryn Oblast, ALT – Russia, Altai Republic, CHA – China, Xinjiang Province, MGL – Mongolia, Övörkhangai Aimag

4.2.2 DNA extraction and microsatellite analysis

The DNA extraction protocol for hair and the DNA amplification protocol can be found in Chapter 2 (General Methods), sections 2.2 and 2.3, respectively. Details on the markers used in this study can be found in Table 2.2.

4.2.3 Stepping stone model

Since the wild progenitor of domestic horses (*Equus ferus*) is extinct, we first reconstructed the population genetic structure of Eurasian populations of *E. ferus,* assuming that the distribution of its genetic variability survives in its domestic descendants. Eurasian populations of *E. ferus* were represented by a linear chain of 80 demes, with 20 demes appended on both sides to avoid boundary artefacts (see Figure 4.2 for a schematic representation of the model and section 4.2.4 for details on the geographic placement of the demes). Each deme was 100 km in diameter; the string of 80 demes thus corresponded to 8000 km.

We investigated three putative population origins of *E. ferus* in Eurasia, one in Far East Asia (represented by deme zero), one in Central Asia (represented by deme 40), and one in Europe (represented by deme 80). In each scenario, the deme corresponding to the population origin of *E. ferus* in Eurasia was populated by randomly sampling K_0 diploid individuals from a hypothetical ancestral population. The initial founder population grows linearly at rate rK horses per generation until it reaches carrying capacity K. Demes at carrying capacity send out a fraction c of colonisers to neighbouring empty demes. In addition, occupied neighbouring demes exchange mN_{min} migrants per generation, where N_{min} represents the smaller of the two population sizes. The general stepping-stone dynamic used here can deal with a wide variety of demographic scenarios, including populations at migration-drift equilibrium and populations having undergone range expansions, depending on the choice of parameters.

The domestication process was initiated 450 generations before the present (t-450), the number of generations that have elapsed since domestication started, assuming an average generation time of 12 years for horses (Sokolov & Orlov 1986) and a start date for horse

domestication 6000 years ago (Ludwig *et al.* 2009; Outram *et al.* 2009). Populations of domestic horses were represented by a linear chain of demes parallel to the one representing populations of wild horses and with the same spatial structuring. The sampled horse populations (Figure 4.1) were placed on this chain according to their geographic distance from the easternmost deme (see section 4.2.4 for details). We considered four putative origins of horse domestication, one in western and central Europe (centred around deme 75 and stretching from western France to central Poland), one in the western steppe (centred around deme 50 and stretching from central Poland to central Kazakhstan), one in the central steppe (centred around deme 25 and stretching from central Kazakhstan to western Mongolia), and one in the eastern steppe (centred around deme zero and stretching from western Mongolia to the Bering Sea).

In each scenario, domestication was initiated by randomly sampling $c_{d0}K_d$ individuals from the deme representing the wild population closest to the domestication origin in that particular scenario. The initial founder population grows within one generation to size K_d, the carrying capacity of the domestic demes. Subsequent, empty demes (i.e. domestic herds) are established from c_dK_d effective individuals representing a mixture of already domesticated horses from the previous deme and local wild horses in proportions q and 1-q, respectively. Once established, neighbouring domestic populations exchange migrants at rate $m_{d.}$

The parameter q describes the proportion of already domesticated horses in the founding stock of subsequent domestic populations. The extreme case whereby horse domestication spreads solely through the independent domestication of numerous wild horse populations (i.e. no domestic horse movement) is represented by q=0, while q=1 describes the other extreme, where horse domestication spreads solely through the movement of domestic animals, without any introgression from the wild (pure demic diffusion). In the former case (q=0), models with different domestication origins would be expected to fit the data equally well.

To find the parameter combinations that best fit the observed data, we ran parameter sweeps for each of the 12 scenarios, simultaneously comparing expected within- and between-population heterozygosities of the model with that of the dataset (12 within-

population estimates and 66 between-population estimates). The parameter space of each scenario was at first adaptively investigating using a Markov chain Monte Carlo approach (MCMC), followed by uniform sampling.

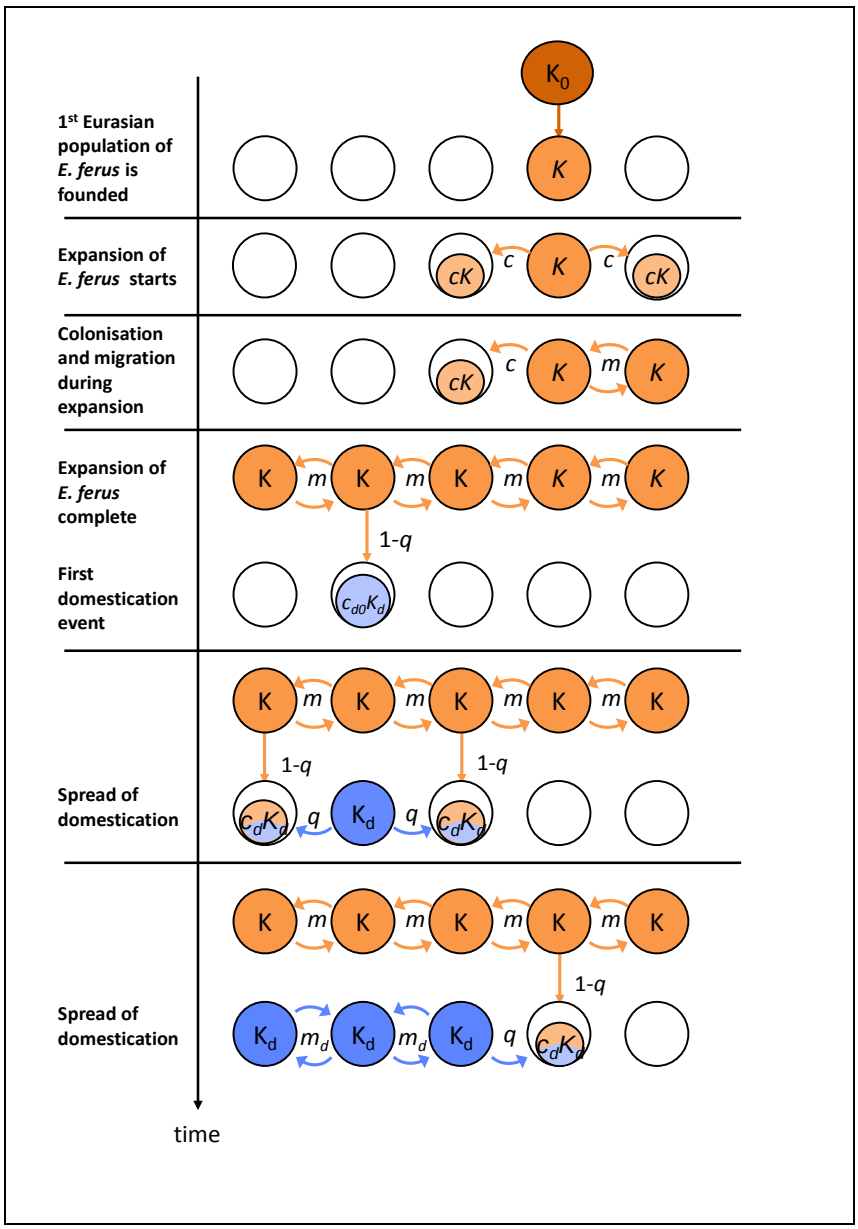

Figure 4.2

Schematic representation of the linear stepping stone model used in this study. Note: for illustrative purposes only 5 demes are shown.

4.2.4 Assigning sample locations to demes in the model

The easternmost of the central 80 demes in the stepping stone model was chosen as the anchor point (latitude 50° and longitude 125°). Sample populations were assigned to demes such that their distances to the anchor deme within the stepping stone model were as close as possible to the shortest distances on land between the corresponding geographic locations (Table 4.1, calculated using the method in Prugnolle *et al.* 2005).

Table 4.1

Details of the samples include in this chapter

Country	Region	ID	Latitude	Longitude	distance from anchor deme [km]	*H*	N
Mongolia	Övörkhangai	MGL	48.0	101.0	1951	0.790	44
China	Xinjiang	CHA	48.7	87.0	3035	0.767	34
Russia	Altai	ALT	51.6	85.0	3041	0.779	40
Russia	Kalmykia	KAL	47.5	45.3	5833	0.783	22
Kyrgyzstan	Naryn	KYR	41.1	75.7	4359	0.798	20
Kazakhstan	Mangystau	KSW	42.3	53.2	5687	0.790	24
Kazakhstan	Kyzylorda	KAR	46.0	61.3	4997	0.773	35
Kazakhstan	Karagandy	KKA	50.0	73.0	3981	0.787	25
Ukraine	Lviv	UKL	50.3	30.9	6359	0.741	21
Ukraine	Zakarpattia	UKZ	49.2	23.6	6739	0.734	18
Lithuania	Vilnius	LTA	56.9	25.4	6223	0.739	21
Georgia	Samegrelo	GGA	42.3	42.3	6361	0.765	24

4.2.5 Predicted heterozygosities of wild and domesticated horses

Heterozygosity was calculated as 1-*F*, where *F* is the corresponding homozygosity. Let $F_{ij}^{ww}(t,\mu)$ denote the expected homozygosity of a pair of alleles drawn randomly from two wild horses in demes i and *j* in generation *t*, under the infinite allele model with mutation probability μ per locus and generation (i.e. when each mutation gives a new allelic variant). Given the migration matrix $M_{ij}^{w}(t)$ (the probability that, in generation *t*, an individual in deme *j* emigrates to deme i) and the population size $N_i^w(t)$ of deme *i* in generation *t*, we can write a recursion for $F_{ij}^{ww}(t,\mu)$:

$$F_{ij}^{ww}(t+1) = (1-\mu)^2 \sum_{i'j'} M_{ii'}^{w} M_{jj'}^{w} \left[\frac{\delta_{i'j'}}{2N_{i'}^{w}} + \left(1 - \frac{\delta_{i'j'}}{2N_{i'}^{w}}\right) F_{i'j'}^{ww} \right]$$

where the right-hand side is evaluated in generation *t*.

When comparing two domesticated horses, or one wild and one domesticated horse, we describe the effect of migration and colonisation in two stages. The homozygosity after migration in generation *t* is

$$\tilde{F}_{ij}^{dd}(t) = (1-\mu)^2 \sum_{i'j'} M_{ii'}^{d} M_{jj'}^{d} \left[\frac{\delta_{i'j'}}{2N_{i'}^{d}} + \left(1 - \frac{\delta_{i'j'}}{2N_{i'}^{d}}\right) F_{i'j'}^{dd} \right]$$

$$\tilde{F}_{ij}^{wd}(t) = (1-\mu)^2 \sum_{i'j'} M_{ii'}^{w} M_{jj'}^{d} F_{i'j'}^{wd}$$

Here N_i^d is the effective population size in the domesticated horse population of deme i, and M_{ij}^{d} is the migration rate between the domesticated populations in demes i and j.

The effect of establishment of newly colonised demes on the homozygosity of two domesticated horses is described by the following relations:

When $i = j$ and i is being colonised,

$$F_{ij}^{dd}(t+1) = \frac{1}{2c_d K_d} + \left(1 - \frac{1}{2c_d K_d}\right)\left[q^2 \tilde{F}_{k_i k_i}^{dd} + 2q(1-q)\tilde{F}_{ik_i}^{wd} + (1-q)^2 \tilde{F}_{ii}^{ww}\right].$$

When i is being colonised, and j is already colonised,

$$F_{ij}^{dd}(t+1) = q\tilde{F}_{k_i j}^{dd} + (1-q)\tilde{F}_{ij}^{wd}.$$

When $i \neq j$ and both demes are being colonised,

$$F_{ij}^{dd}(t+1) = q^2 \tilde{F}_{k_i k_j}^{dd} + q(1-q)\left(\tilde{F}_{ik_j}^{wd} + \tilde{F}_{jk_i}^{wd}\right) + (1-q)^2 \tilde{F}_{ij}^{ww}.$$

otherwise $\quad F_{ij}^{dd}(t+1) = \tilde{F}_{ij}^{dd}(t+1).$

The corresponding relation for the homozygosity of a wild and a domesticated horse is

$$F_{ij}^{wd}(t+1) = \begin{cases} q\tilde{F}_{ik_j}^{wd} + (1-q)\tilde{F}_{ij}^{ww} & \text{when } j \text{ is being colonised} \\ \tilde{F}_{ij}^{wd} & \text{otherwise} \end{cases}$$

When the first domestic deme is colonised, we take $q = 0$, since in this case all horses must come from the local wild population.

To calculate the corresponding recursions for the SMM model, let t_{ij} be the number of generations to the most recent common ancestor of a pair of individuals from demes i and j. Under the SMM model the difference Δ in repeat count of two alleles is then the sum of $2t_{ij}$ independent identically distributed random variables, each of which is -1, 0, or 1 with probabilities $\mu/2$, $1-\mu$, and $\mu/2$, respectively. Hence, the characteristic function for the difference in repeat number, $\langle e^{i\omega\Delta}\rangle$, is $(1-\mu+\mu\cos\omega)^{2t_{ij}}$. It follows that the homozygosity under the SMM model is

$$F_{ij}^{\mathrm{SMM}}(t,\mu) = \frac{1}{2\pi}\int_0^{2\pi} \left\langle (1-\mu+\mu\cos\omega)^{2t_{ij}} \right\rangle d\omega,$$

where the angular brackets denote expectation over gene genealogies. Thus, it follows that the SMM homozygosity is related to the infinite-alleles homozygosity as

$$F_{ij}^{\mathrm{SMM}}(t,\mu) = \frac{1}{2\pi}\int_0^{2\pi} F_{ij}\big(t,\mu(1-\cos\omega)\big)\, d\omega$$

where F_{ij} is any of F_{ij}^{ww}, F_{ij}^{wd} or F_{ij}^{dd}. We used the following numerical approximation to evaluate the integral:

$$F_{ij}^{\mathrm{SMM}}(t,\mu) \approx \sum_{k=1}^{n} F_{ij}\left(t,\mu\left(1-\cos\frac{\pi(k-1/2)}{n}\right)\right)$$

This approximation is very accurate when n is large enough that the probability of observing a difference of more than n repeat units can be ignored. Using $n = 50$ was enough to obtain machine precision for the parameters used in this study.

4.2.6 Model fitting

We fitted the data by finding parameter combinations yielding high values of R^2 comparing predicted versus observed heterozygosity. Because some parameters affect mainly the within-population heterozygosity (e.g. *c*), whereas others are more important for the between-population heterozygosity values (e.g. *m*), we calculated R^2 for within- and between-population heterozygosity values separately and used the average of the two for fitting the model.

In order to explore which parameter combinations in the model best explain the data, we sampled parameter values randomly according to a uniform distribution of the log of parameter (for all parameters except *t* and *q*, which were drawn randomly from the ranges [500, 30 000] and [0, 1], respectively). Because some parameters are highly constrained by the data, straightforward implementation of this scheme is very inefficient. To overcome this problem, we employed a tiered strategy where we first used a simple Monte-Carlo method to identify the good areas of the parameter space (using 50 independent chains running for 60,000 steps each, following a 40,000 steps burn-in), followed by a uniform parameter sweep restricted to these areas. Each MC chain was started from a randomly chosen starting point in the parameter space. We then added a randomly chosen change to each parameter (within ±1% of the parameter's range), and calculated the R^2 of the new parameter combination. This combination was accepted with probability $\min(1, \exp(\gamma(R^2_{\text{new}} - R^2_{\text{old}})))$. Using $\gamma = 1$ yielded satisfactory convergence to good parameter combinations, while allowing the chains to avoid getting stuck at local optima.

Based on the MC sweeps, we restricted the parameter sweep to the regions $t \in [1000,14500]$, $m \in [10^{-6}, 10^{-3}]$, $cK \in [1, 10^4]$, $r \in [0.005,1]$, $K \in [4000, 10^5]$, $K_0 \in [1, 10^5]$, $m_d \in [10^{-4}, 0.5]$, $K_d \in [500, 10^4]$, $c_d K_d \in [1, 10^3]$, $q \in [0,1]$, and $c_d K_{d0} \in [1, 10^3]$. In addition, the MC sweeps revealed a strong connection between K_0 and *t*, where either $t \in [12000,14500]$ or the expected heterozygosity of pairs of individuals with MRCA dating back to the ancestral population is in the range [0.79, 0.81].

We explore this connection further in the next section. We drew 20 million uncorrelated samples from this region, for comparing how often each domestication origin provides the best explanation of that data. Finally, we drew an additional 20 million uncorrelated samples for the best-fitting combination (expansion of wild horses from the east and domestication in the western steppe), giving a total of 40 million samples for generating the distribution of parameters in Figure 4.6.

4.2.7 Connection between K_0 and t

In this section we analyse in detail the connection between K_0 and t, Consider two horses sampled from deme n, in the first generation after the deme reached carrying capacity. Let P_a denote the probability that the horses' MRCA dates back to the ancestral population. Ignoring the effect of migration, and using that the number of generations to colonise a deme is $\tau = \lfloor (1-c)/r \rfloor$, we obtain

$$\begin{aligned} P_a &= \left[\prod_{t=0}^{\tau} \left(1 - \frac{1}{2K(c+rt)} \right) \right]^n \\ &= \left[\frac{\Gamma(\tau+1+(c-1/2K)/r)\,\Gamma(c/r)}{\Gamma(\tau+1+c/r)\,\Gamma(\tau+1+(c-1/2K)/r)} \right]^n \\ &\approx \left[(1-1/2Kc)\,(c+r/2)^{1/2Kr} \right]^n \end{aligned}$$

With $n = 50$, corresponding to the centre of domestication, we find that $\log_{10}(1-P_a)$ is highly correlated with the arrival time of the expansion to this deme ($R^2 = 96\%$). (Using the starting time instead gives $R^2 = 82\%$.) Hence, recent expansions correspond to weak bottlenecks (P_a close to unity), and old expansions corresponds to strong bottlenecks (P_a close to zero).

Figure 4.3 shows how P_a is related to arrival time t_a of the expansion to the domestication centre, the expected heterozygosity H_a of horses with MRCA in the ancestral population, and the carrying capacity of the ancestral population (K_0). The heterozygosity is calculated using

$$H_a(t, K_0) = 1 - \frac{1}{2\pi}\int_0^{2\pi} d\omega \frac{[1-\mu+\mu\cos\omega]^{2(t+1)}}{2K_0-(2K_0-1)[1-\mu+\mu\cos\omega]^2}.$$

From the left panel we see that the points fitting the data fall into two distinct categories, based on whether bottlenecks during the expansion where strong or weak (P_a small or close to unity, respectively). When bottlenecks are strong ($P_a < 0.2$) the age of the expansion (t) is confined to between 150kya and 170kya. In contrast, if bottlenecks were weak or intermediate ($P_a > 0.2$), good parameter combinations are characterised by H_a around 0.8, and the age of the expansion can be anything from 10kya to 170kya.

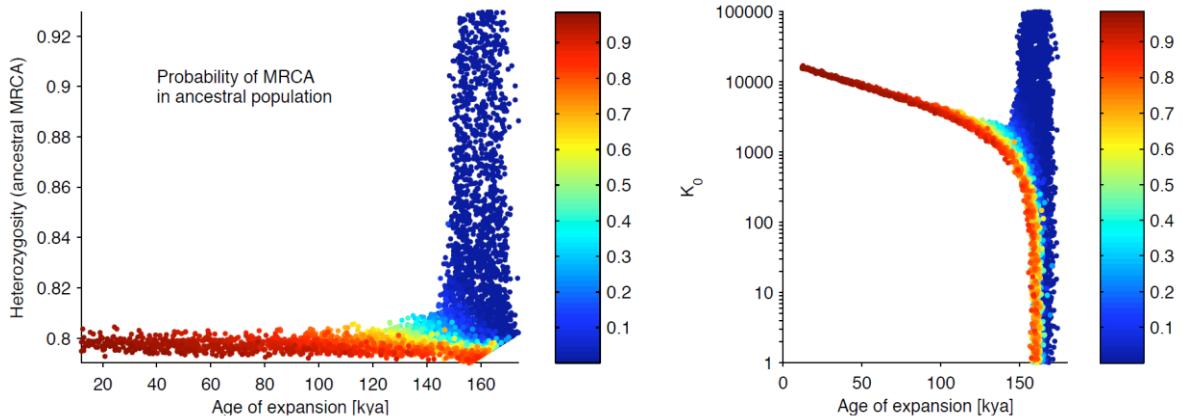

Figure 4.3

Relation between carrying capacity, heterozygosity and time of expansion. A) The probability that two wild horses sampled from the domestication centre have their Most Recent Common Ancestor (MRCA) in the ancestral population (colour coded) as a function of the age of the expansion and of the expected heterozygosity of individuals with MRCA in the ancestral population. B) As in A), but as a function of the age of the expansion and ancestral population size (K_0).

4.3 Results and Discussion

Initial investigation of the parameter space using a Markov chain Monte Carlo approach (MCMC) led to the exclusion of scenarios postulating an origin of wild horses in Central Asia and Europe, leaving only the scenario where wild horses originated in Far East Asia (Figure 4.4A). This result is consistent with paleontological evidence whereby caballine horses entered the Eurasian continent via the Bering land bridge in the Far East (Forsten 1992; Prothero & Schoch 2002).

Next, I formally investigated the four domestication scenarios postulating an origin of wild horses in Far East Asia by uniformly sampling their respective parameter spaces. The scenario in which horse domestication originated in the western steppes received by far the most support, providing the best fit to the data in over 60% of the parameter combinations with R^2>30% (Figure 4.4B). For this scenario, several parameter combinations fitted the data well, explaining up to 55% of the total variance in heterozygosity (R^2=49% for within-population, and R^2=34% for between-population estimates only, Figure 4.5A). By supporting an origin of horse domestication in the western Eurasian steppe, my results provide further evidence for the earliest horse domestication in the western Eurasian steppe as suggested by archaeological data (Brown & Anthony 1998; Clutton-Brock 1999; Anthony *et al.* 2006; Olsen 2006a; Outram *et al.* 2009).

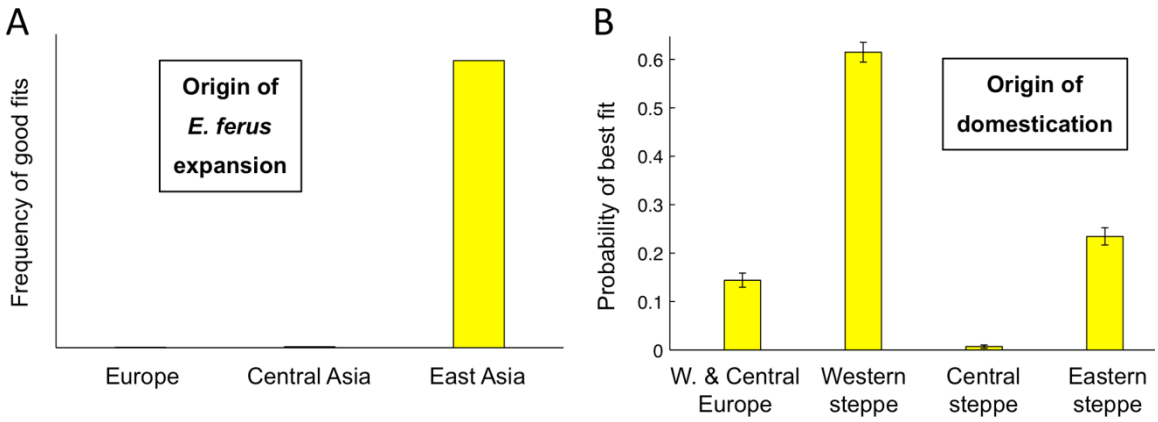

Figure 4.4

Support for scenarios describing the origin of wild and domestic horses. A) Frequency of good fits (R^2>30%) for the three potential origins of the wild ancestor of domestic horses B) The probability of producing the best fit to the data for each of the four potential domestication origins.

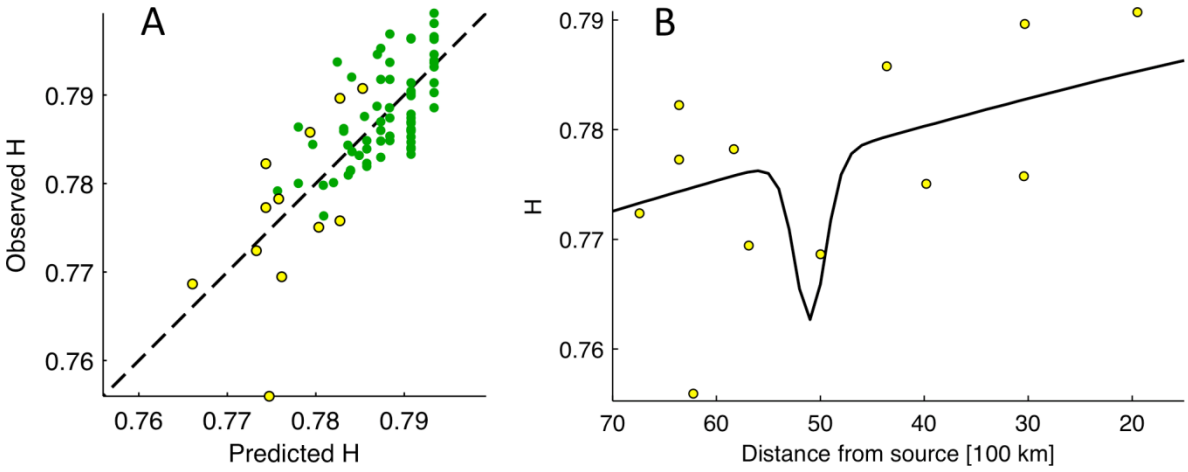

Figure 4.5

Model fit. A) Observed versus predicted heterozygosity (within-population estimates in yellow, between-population estimates in green) for a combination of parameter values that provided a very good fit to the data (Expansion of wild horses out of the Far East and spread of horse domestication out of the western steppes, R^2=55%). B) Within-population heterozygosity (*H*) as observed in 12 horse populations from northern Eurasia (red dots), and as predicted from the model with the best fit (black line).

Next, I explored the demographic details of wild horse expansion out of Far East Asia and their subsequent domestication in the western steppe by inspecting the distribution of parameter combinations that fitted the data well (R^2>30%). The best-fitting model strongly suggests that the wild progenitor of domestic horses expanded out of Far East Asia around 160 kya (Figure 4.6A). According to the fossil record, *Equus ferus*, the progenitor of both domestic horses (*E. f. caballus*) and the Przewalski's horse (*E. f. przewalskii*) (Weinstock *et al.* 2005), first appeared in East Asia around 200 kya (Prothero & Schoch 2002). The expansion captured by the model therefore likely reflects the colonisation of Eurasia by this species. The best-fitting model further suggests that this expansion was characterised by relatively strong founder effects (small cK, Figure 4.6B), large effective population sizes (K, Figure 4.6C), and a rate of spread in the order of ca. 100 km in 300 years (growth rate r=0.04, Figure 4.6D). Due to the early date of this expansion, it is not possible to accurately reconstruct migration rates in the ancestral wild horse populations.

The best-supported model further suggests that the spread of domestication out of the western steppe involved both a movement of domestic horses and the recruitment of local wild horses en route, with new domestic herds founded by between 30-70% domestic horses from the previous deme (q=0.3-0.7 in Figure 4.6F). The demography of horses changed markedly following domestication. Compared to the expansion of *E. ferus*, the spread of domestic horses was characterised by weaker founder effects (larger c_dK_d, Figure 4.6B), and smaller effective sizes of established populations (K_d, Figure 4.6C). We obtained strong support for high migration rates between domestic populations (m_dK_d, Figure 4.6E), consistent with the increased mobility of human societies following horse domestication.

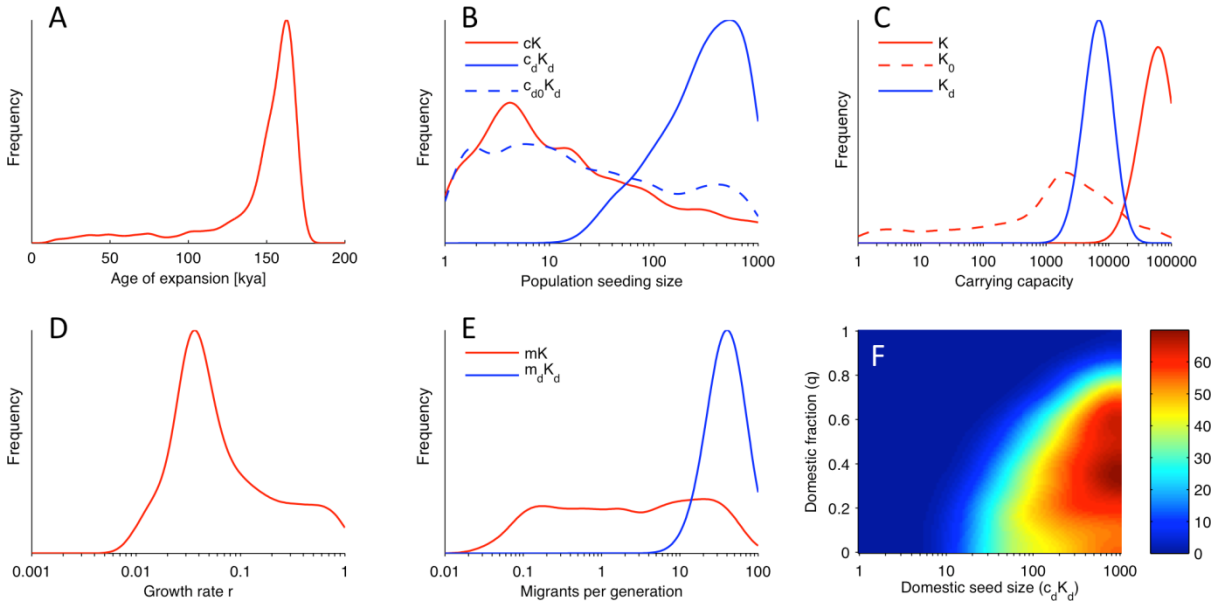

Figure 4.6

Description of selected demographic parameters in both wild and domestic populations. A) Timing (t) of the expansion of wild horses out of Far East Asia, as predicted by the best-fitting model. B) Effective founder population sizes of wild horse (cK, red line), the first ($c_{d0}K_d$, blue dashed line), and subsequent domestic populations (c_dK_d, blue solid line). C) Effective population sizes of wild (K, red line), and domestic populations (K_d, blue line). D) Growth rate, r, of wild horse populations. E) Migration between populations of wild (mK, red line), and domestic horses (m_dK_d, blue line). F) The proportion of domestic (q) and wild (1-q) horses in domestic horse founder populations as a function of their effective size, c_dK_d. At c_dK_d =500 (see panel B), q values between 0.3 and 0.7 receive the most support.

The results of this study suggest that the geographic pattern of within- and between-population heterozygosity observed in contemporary horses from the Eurasian steppes is a consequence of both the east-to-west expansion of *E. ferus*, and a movement of domestic horses (demic diffusion) out of the western steppe. The expansion of *E. ferus* out of Far East Asia set up an isolation-by-distance (IBD) pattern (off-diagonal elements, Figure 4.7A) which has been preserved in Eurasian steppe horses due to extensive and widespread backcrossing with their wild ancestor (off-diagonal elements, Figure 4.7B). The demic component of the spread of horse domestication out of the western steppe, on the other hand, accentuated the east-to-west decline in within-population genetic diversity (on-diagonal elements, Figure 4.7B), which had been much weaker in the ancestral wild populations (on-diagonal elements, Figure 4.7B). While migration rates in domestic horses are high (Figure 4.4E), the relatively recent occurrence of horse domestication (from ca 5.5 kya (Ludwig *et al.* 2009; Outram *et al.* 2009)) means that there has not been enough time for increased population movements to obscure this ancient pattern over large geographic scales [an observation that also holds true for humans (Prugnolle *et al.* 2005; Ramachandran *et al.* 2005; Manica *et al.* 2007)].

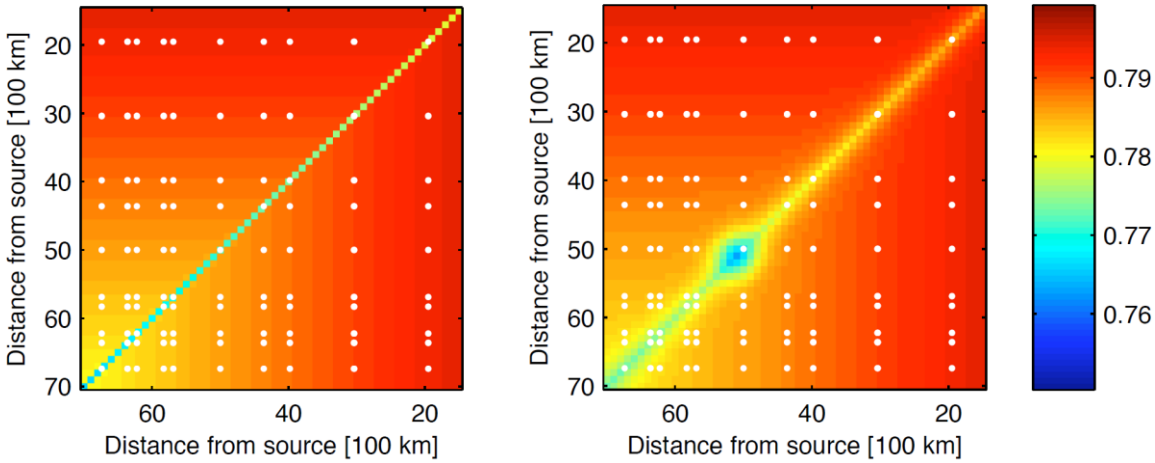

Figure 4.7

Within and between population genetic diversity in wild and domestic horses as a function of distance from the expansion origin in East Asia. A) Between-population heterozygosity in wild horses (off-diagonal elements) corresponds to a pattern of isolation-by-distance (IBD). The decline in within-population heterozygosity (on-diagonal elements) is relatively weak. B) The original pattern of IBD (off-diagonal) has been preserved in modern domestic horses from the steppes owing to the extensive incorporation of wild horses into domestic stock, and the east-to-west decline in within-population diversity (on-diagonal) has been accentuated owing to the combined effects of a demic spread out of the western steppe and continued introgression from the wild. The dip in within-population heterozygosity around 5000 km reflects the strong bottleneck associated with the initial domestication of horses in the western steppe.

The model of horse domestication suggested here might also help understand the differential patterns of diversity observed in mitochondrial (maternal lines) and Y-chromosomal (paternal lines) DNA, which have previously led to conflicting conclusions regarding the number of origins and the mode of spread of horse domestication. Provided there was a sex bias in the recruitment and breeding of horses, these two uniparental markers probably reflect different aspects of the domestication process: while the observed low levels of Y-chromosome variability (Lindgren *et al.* 2004; Wallner *et al.* 2004; Ling *et al.* 2010a) might be a consequence of the strong bottleneck associated with the domestication of wild stock from a geographically restricted region in the western steppe (low $c_{d0}K_d$, Figure 4.6B) (Lindgren *et al.* 2004), the high diversity and limited geographic structure in the horse mitochondrial genome (Lister *et al.* 1998; Vilà *et al.* 2001; Jansen *et al.* 2002; McGahern *et al.* 2006; Cieslak *et al.* 2010) might reflect the continued augmentation of domestic herds with local wild females ($1-q>0.3$, Figure 4.6F).

The repeated capture of wild female horses for the purpose of maintaining or growing domestic herd sizes may seem counterintuitive, given that in other domestic animal species, introgression from the wild usually involves domestic females being impregnated by wild males (Luikart *et al.* 2001). However, given the initial difficulties in successfully breeding the most closely related wild relative of domestic horses the Przewalski's horse in captivity (Boyd & Houpt 1994), it can be speculated that it might initially have been too difficult to maintain herd sizes solely through breeding existing stock (see also Levine 1999a). Since stallions are inherently more difficult to handle than mares, a sex-bias towards females in capturing wild horses may thus explain the large number of female lineages in the domestic horse gene pool, as suggested by mtDNA studies (Lister 2001; Vilà *et al.* 2001; Jansen *et al.* 2002; McGahern *et al.* 2006; Cieslak *et al.* 2010).

In this study, the combined use of a spatially explicit model and autosomal markers has led me to uncover aspects of the origin and spread of horse domestication at a level of detail on par with that seen in other species (Linz *et al.* 2007; François *et al.* 2008; Tanabe *et al.* 2010), including humans (Ramachandran *et al.* 2005; Liu *et al.* 2006). In horses, I was able to reconstruct the population genetic structure of the extinct ancestral wild species, and to reconstruct the complex process of horse domestication, providing a scenario that integrates previously seemingly conflicting lines of evidence from archaeology, mtDNA and Y-chromosomal sequence data.

5 Post-domestication gene flow

Abstract

Gene flow among populations of domestic animals is primarily determined by humans. However, the factors influencing human-mediated gene flow among domestic animals, especially non-breed animals, are poorly understood. In a previous study (Chapter 3), I showed that gene flow among non-breed domestic horses in Eastern Eurasia is a function of geographic distance; however, the correlation between pairwise genetic and simple geographic distances was shown to be weak. Here, I explore the effect of three additional factors on gene flow in non-breed domestic horses, namely inland water bodies, altitude, and the ancient network of trade routes known as the Silk Roads. Using a least-cost path algorithm, I computed matrices of effective distances among 17 populations sampled across Eastern Eurasia and correlated each of them with a pairwise genetic distance matrix. Mantel tests revealed that the genetic structure of the populations investigated here is best explained by elevated levels of gene flow along the Silk Roads, whereas high altitude appeared to have played no major role in shaping the genetic structure of Eastern Eurasian horses. This suggests that the Silk Roads facilitated gene flow even across the highest and most inaccessible mountain ranges, such as the Himalayas and the Tibetan Plateau.

5.1 Introduction

Determining routes of gene flow can provide important insight into the demographic history of species and populations. In natural populations, gene flow is a function of species attributes, such as dispersal ability, and the cost associated with travelling across the landscape. In homogeneous landscapes, gene flow among populations is primarily determined by geographic distance, with an increase in genetic differentiation with increasing geographic distance (isolation by distance, IBD, Wright 1943). At larger spatial scales most if not all species are affected by IBD (Guillot *et al.* 2009).

In heterogeneous landscapes, certain environmental or landscape elements are expected to enhance or impede gene flow between subpopulations. Factors that have been shown to influence the connectivity of natural populations in addition to geographic distances include natural landscape elements such as topography (Funk *et al.* 2005), vegetation cover (Sacks *et al.* 2008), and waterways, climate-related factors such as snow cover (Stenseth *et al.* 2004) and temperature (Pilot *et al.* 2006), and anthropogenic features such as roads (Gerlach & Musolf 2000; Coulon *et al.* 2006; Riley *et al.* 2006) and dams (Yamamoto *et al.* 2004). In most natural populations, geographic distances that take landscape elements into account (effective or cost distances) have been shown to explain genetic structure better than simple geographic distances.

In domestic animals, gene flow among subpopulations, and thus their genetic structure, is expected to be determined by the connectivity between owners and traders of breeding stock ("farmer connectivity", Berthouly *et al.* 2009). The connectivity of owners and traders of animals belonging to specific registered breeds is primarily determined by breed membership, a breed being defined as „a group of animals that has been selected by humans to possess a uniform appearance that is inheritable and distinguishes it from other groups of animals within the same species" (Clutton-Brock 1999). Thus, the genetic structure of breed animals typically coincides with breed boundaries. The genetic variability of most European populations of cattle (Gautier *et al.* 2010), sheep (Handley *et al.* 2007a), pigs (Megens *et al.* 2008), and horses (Bjornstad & Roed 2001; Glowatzki-Mullis *et al.* 2006), for example, is clearly partitioned into breeds.

In many parts of the world, breeding for a narrowly defined phenotype is either not desirable, or not practicable. In domestic animals which are mainly used for work-related tasks (and where the intensity of artificial selection for specific phenotypes is therefore relaxed), the connectivity among owners and traders of breeding animals, and thus gene flow among them, is expected to be determined by factors other than breed membership. In a recent study, Berthouly *et al.* (2009) showed that the connectivity between goat farmers in a small Vietnamese province is a function of their ethnicity and their husbandry style, whereas natural landscape features such as landscape topography appear to play no role in this system.

However, while social relationships among owners and traders of breeding stock may play an important role in facilitating gene flow between populations of such "non-breed" animals, other factors are expected to play a role in structuring non-breed domestic populations at larger spatial scales. The Yangtse River in China, for instance, has been suggested to act as a barrier to gene flow between neighbouring populations of swamp buffalo (Zhang *et al.* 2007); similarly, high altitudes have been suggested to impede gene flow between yak populations in Mongolia (Xuebin *et al.* 2005). While the results of these studies suggest that landscape topography may play a role in structuring populations of non-breed animals over a large spatial scale, the factors that determine gene flow among populations of domestic animals that do not belong to any particular breed have never been investigated formally.

In Chapter 3, I showed geographic distance played a significant role in structuring the genetic variability of non-breed horses in Eastern Eurasia (Figure 3.3). In this chapter, I explore the effect of three additional factors on the genetic structure of domestic horses: inland water bodies, altitude, and the network of ancient trade routes spanning Eastern Eurasia, known as the Silk Roads.

5.2 Methods

5.2.1 Samples

The dataset analysed here consists of 455 non-breed horses from 17 populations distributed throughout Eastern Eurasia. The distribution of the sampling locations is shown in Figure 5.1. Further information on the samples can be found in Table 2.1.

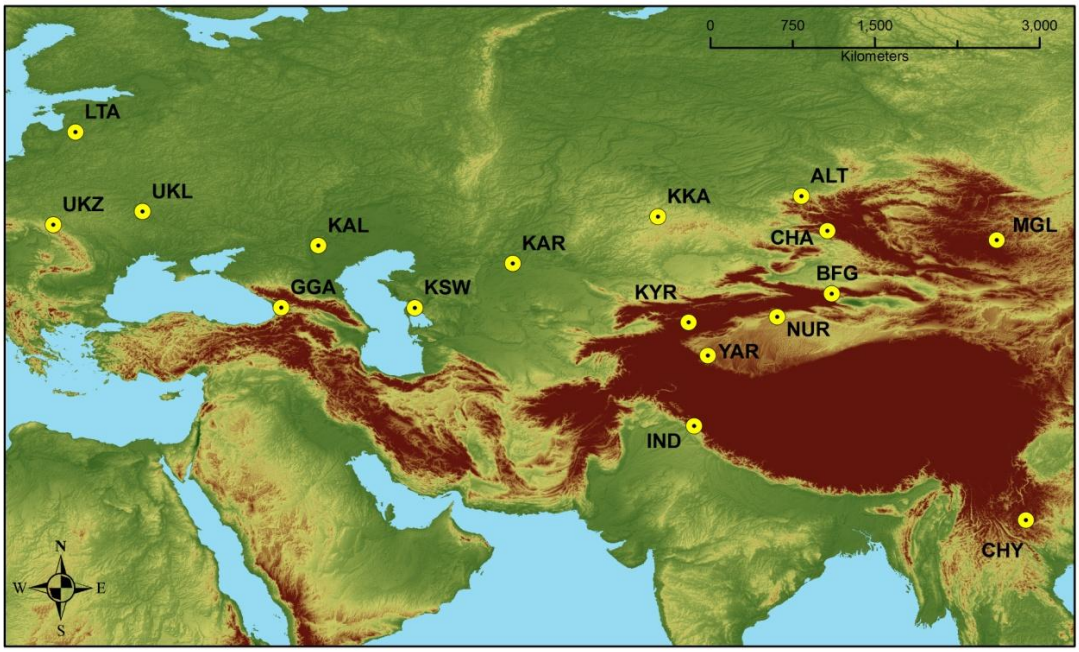

Figure 5.1

Geographic distribution of the 17 populations analysed in this chapter. ALT –Russia, Altai; BFG – China, Xinjiang; CHA – China, Xinjiang; CHY – China, Yunnan, GGA – Georgia; IND – India, Jammu & Kashmir; KAL – Russia, Kalmykia; KAR – Kazakhstan, Kyzylorda; KKA – Kazakhstan, Karagandy; KSW – Kazakhstan, Mangystau; KYR – Kyrgyzstan, Naryn; LTA-Lithuania; MGL – Mongolia, Övörkhangai; NUR – China, Xinjiang; UKL – Ukraine, Lviv; UKZ- Ukraine, Zakarpattia; YAR – China, Xinjiang.

5.2.2 DNA extraction and genotyping

The DNA extraction protocol for hair and the DNA amplification protocol can be found in Chapter 2 (General Methods), sections 2.2 and 2.3, respectively. Details on the markers used in this study can be found in Table 2.2.

5.2.3 Genetic diversity

Levels of genetic variation were calculated for the whole dataset and within each of the 17 populations as expected heterozygosity (H_e, unbiased estimator, Nei 1987), observed heterozygosity (H_o), and allelic richness (R_S). Estimates of allelic richness were standardised to the smallest sample size in this dataset, N=15 using the rarefaction algorithm implemented in FSTAT v 2.9.3.2 (Goudet 2001, updated from Goudet 1995). Genetic differentiation between populations was estimated using pairwise $F_{ST,}$ and the statistical significance of the F_{ST} values was tested using permutation tests. Linkage disequilibrium was tested between all pairs of loci over all populations, and deviations from Hardy-Weinberg equilibrium (HWE) were tested both within (F_{IS} as test statistic) and overall (F_{IT} as test statistic) populations using permutation tests. All permutation tests were performed in FSTAT using Bonferroni corrections to account for multiple testing. For more details see Chapter 2 (General Methods).

5.2.4 Spatial genetic structure

In Chapter 3, I used discriminant analysis of principal components (DAPC) to determine the degree of differentiation between the 17 populations in my dataset. DAPC, while achieving maximal differentiation between groups, may fail to detect spatial structuring if the structuring is not associated with strong genetic differentiation (Jombart *et al.* 2008). I therefore investigated the spatial genetic structure of Eastern Eurasian horse populations using spatial principal components analysis (sPCA, Jombart *et al.* 2008), a multivariate method that explicitly incorporates spatial autocorrelation between populations into the clustering procedure.

I used spatial principal components analysis (sPCA, Jombart *et al.* 2008) to determine the spatial genetic structure in our dataset. sPCA is a spatially explicit multivariate method that identifies spatial patterns using allele frequency data of individuals or populations (entities). Briefly, sPCA uses a matrix of allele frequencies **X**, and a spatial weighting matrix L, which contains information on the spatial proximity of the entities. The spatial proximity of the entities is derived from a connection network that is established using one of several available algorithms (e.g., Legendre & Legendre 1998). The spatial weighting matrix **L** is then used to compute the spatial autocorrelation *I* (Moran 1948) of the variables x (i.e. the allele frequencies) using

$$I(\mathrm{x}) = \frac{\mathbf{x}^{\mathrm{T}}\mathbf{L}\mathbf{x}}{x^{T}x} \qquad (1)$$

When spatial neighbours (as defined by the connection network) have similar values of x, the spatial autocorrelation *I* will be positive. Conversely, when spatial neighbours have very different values of x, the spatial autocorrelation will be negative. In the former case, the spatial structure in the dataset is said to be global, whereas it is said to be local in the latter.

To measure the genetic variability in addition to the degree of spatial autocorrelation in x, sPCA defines

$$\mathrm{C(x)} = \mathrm{var(x)} * I(\mathrm{x}) = \frac{1}{n}\mathrm{x}^{T}\mathbf{L}\mathrm{x} \qquad (2)$$

C (x) (the so called component scores) are positive when the variance in allele frequencies is high and spatial autocorrelation is positive (i.e. if global structuring is present); conversely, C(x) are negative when the variance in allele frequencies is high and spatial autocorrelation is negative (i.e. if local structuring is present). All sPCA analyses were performed using the *adegenet* package version 1.2.8 (Jombart 2008) in the statistical software R, version 2.13.0 (R Development Core Team 2010). The spatial connectivity network was defined using the Delaunay triangulation and the significance of the global and local structuring was tested using a Monte Carlo approach (n=10,000).

5.2.5 Landscape genetics

To identify the factors that have influenced the genetic structure of non-breed domestic horses in Eastern Eurasia, I used a landscape genetics approach (Manel *et al.* 2003). In landscape genetics, features that are thought to increase or decrease the cost of movement of the study organism through the landscape are assigned different cost values and distances that minimise the overall cost of travelling across the resulting resistance surface (so called least-cost distances) are computed between all pairs of sampling units (individuals or populations). Least-cost distance matrices are then correlated with genetic distance matrices and the correlation between the matrices is evaluated using Mantel and partial Mantel tests. The factors that most influence the genetic structure of the study organism can then be identified by comparing the strength of the correlation produced by the different resistance matrices.

5.2.6 Potential drivers of genetic structure

I investigated three factors potentially influencing the genetic structure in non-breed horse populations: water bodies, altitude, and the main trade routes that connected East, South and West Asia from as early as 4000 BP (Curtin 1985). To compute distances between populations, I used a graph theory-based approach (Manica *et al.* 2005; Prugnolle *et al.* 2005). I represented the world as a graph of 40,962 equally spaced nodes on a spherical referential, with every node being connected to its six closest neighbours (approximately 100 km from each other). Shortest distances between two locations (i.e. two nodes) on the graph were determined using Dijkstra's algorithm (Dijkstra 1959). In unweighted graphs, the shortest distance between two locations approximates to the great circle distance between them, whereas in weighted graphs, the shortest distance between locations represents the path that yields the lowest overall cost of travelling between two locations.

To assess the effect of water bodies on domestic horse genetic structure, I assigned a uniform cost of one to all edges connecting nodes representing land ("land nodes"), and removed all edges that involved nodes representing water ("water nodes"), thus preventing crossing of water bodies (model [W1]). The effect of altitude was modelled by assigning a weight w to all edges connecting pairs of land nodes using the following equation

$$w_{ab}=1+\tau\,(\text{altitude}_a + \text{altitude}_b)/2 \quad (5.1)$$

where w_{ab} is the weight of the edge connecting nodes *a* and *b*, and τ a scaling factor. I considered three values for τ, τ =0.002 (model [A1]), τ= 0.001 (model [A2]), and τ =0.0006 (model [A3]), corresponding to a 100%, 50%, and 30% increase in cost for every1000 meters of altitude ascended respectively. The geographic coordinates of 86 locations along the main Silk Roads (obtained from www.ciolek.com/OWTRAD/DATA/oddda.html) were used to define edges that represent trade routes. Because the relative cost of travelling along trade routes versus non-trade routes is unknown, several travel cost ratios were tested: 0.8:1 (model [T1]), 0.6:1 (model [T2]), 0.4:1 (model [T3]), 0.2:1 (model [T4]), 0.1:1 (model [T5]), and 0.05:1 (model [T6]). The reverse case, i.e. a higher cost of travelling along trade routes versus non-trade routes was also investigated.

Least cost paths were estimated using Dijkstra's algorithm using the RBGL package in R. The relationship between the genetic distance matrix [G] and all cost distance matrices was evaluated using simple Mantel (Mantel 1967) and partial Mantel tests (Smouse *et al.* 1986). Mantel tests and partial Mantel tests were carried out in R using the package *vegan* (Dixon 2003) and 10,000 permutations.

5.3 Results

5.3.1 Genetic Diversity

All sampling locations were in Hardy-Weinberg equilibrium at all loci except for locus EB2E8, which significantly deviated from Hardy-Weinberg expectations in three of the 17 sampling locations ($p<0.00011$, the adjusted α-value at the 0.05 level following Bonferroni correction). Out of 325 pairwise combinations, three pairs of loci showed significant deviations from linkage equilibrium at the adjusted α-value of p=0.00015: TKY321 x ASB2, ASB2 x TKY374, and TKY374 x TKY394. For more details, see Chapter 3, section 3.3.1.

The observed number of alleles per locus ranged between 4 (HMS5) and 17 (TKY343); the observed number of alleles per population ranged between 168 (UKL) and 219 (MGL); the

observed number of private alleles per population ranged between 0 (KAR, KAL, UKL, UKZ, and LTA) and 5 (KYR). Mean expected heterozygosity was 0.784, ranging between 0.756 (LTA) and 0.797 (BFG). Mean allelic richness was 6.87, ranging between 6.45 (LTA) and 7.16 (BFG). For more details, see Chapter 3, section 3.3.1 and Table 3.1.

5.3.2 Spatial genetic structure

The spatial connectivity of the populations is shown in Figure 5.1A. Visual inspection of the screeplot showed that the first global component captured most of the structure in the dataset (Figure 5.1B). However, Moran's *I* values were greater than 0.5 for the first (*I*=0.69) and the second (*I*=0.53) global component (Figure 5.1C&D), indicating considerable spatial autocorrelation in both these components. A global test confirmed the existence of global pattern (max(**t**) = 0.139, p=0.019, 10,000 permutations), but a lack of local structure (max(**t**) = 0.063, NS, 10,000 permutations). Visual representation of the scores of the first global component shows that populations in the middle of the connection network have less extreme scores (smaller squares in Figure 5.1C) than locations on the periphery, corresponding to a longitudinal cline in allele frequencies, rather than distinct genetic clusters; this is consistent with previous results (Chapter 3). The second global component shows a cluster in Central Asia (Figure 5.1D) which is characterised by high scores in southwest Kazakhstan and successively smaller scores with increasing distance from that area, possibly reflecting the spread of domestic horses out of the western steppe (Chapter 4).

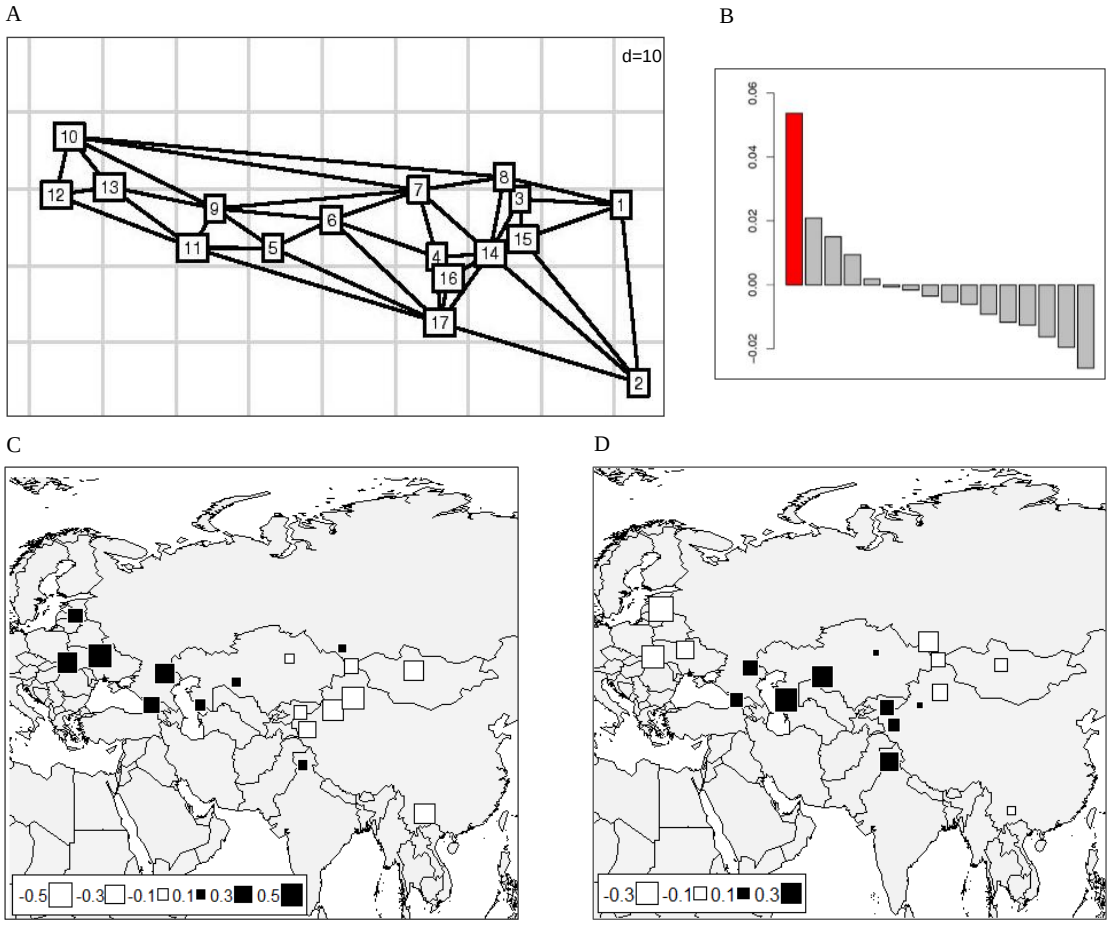

Figure 5.2
A) Connection network (Delaunay triangulation) used to define spatial weightings. 1 MGL, 2 CHY, 3 CHA, 4 KYR, 5 KSW, 6KAR, 7 KKA, 8 ALT, 9 KAL, 10 LTA, 11 GGA, 12 UKZ, 13 ULK, 14 NUR, 15 BFG, 16 YAR, 17 IND. B) Screeplot of eigenvalues. C) Representation of the first component (Moran's *I*=0.69). D) Representation of the second global component (Moran's *I*=0.53). Black and white symbols represent the values of the scores, with different sizes representing different absolute values.

5.3.3 Landscape genetics

Simple Mantel tests suggested that the genetic distance matrix [G] was significantly correlated with the great-circle distance matrix [D] (see also Chapter 3) and all cost-distance matrices except for [A1] (Figure 5.3). The highest correlation coefficient was obtained for the trade route model [T4] with a trade route : non-trade-route cost ratio of 0.2 : 1 (r=0.518, p= 0.001, 10,000 permutations, Figures 5.3 and 5.4). When great-circle distances were partialled out using partial Mantel tests, neither shortest distances on land [W] nor the three least-cost distances accounting for altitude ([A1] - [A3]) were significantly correlated with genetic distance [D] (p>0.2 for all models; Table 5.1).

Similarly, trade route distances with a cost ratio of trade route : non-trade route larger than 0.4 : 1 were non-significant. In contrast, trade route distances with a cost ratio of trade route : non-trade route smaller than 0.4 : 1 were significant even after controlling for great-circle distances (Table 5.1 and Figure 5.4; see Figure 5.5 for an illustrative representation of the shortest distances along trade routes). Thus, distances accounting for low-cost trade routes, but not for inland water bodies and altitude explained additional variation not explained by great-circle distance.

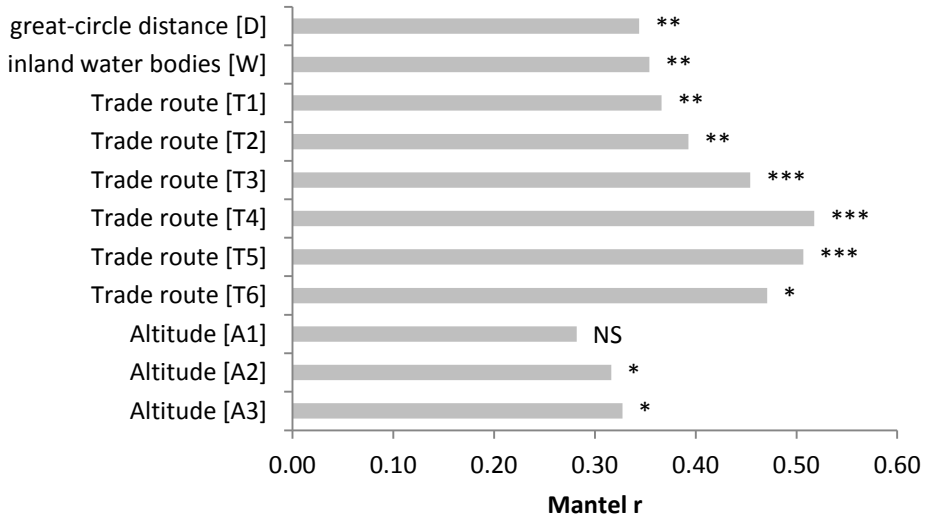

Figure 5.3

Results of simple Mantel tests. *p < 0.05, **p < 0.01, ***p < 0.001, NS=non-significant.

Table 5.1
Evaluation of the least-cost distance models

	Partial Mantel test	
Least cost distance model	r	p-value
WG.D	0.118	0.222
A1G.D	-0.033	0.549
A2G.D	-0.018	0.506
A3G.D	-0.009	0.486
T1G.D	0.223	0.092
T2G.D	0.293	0.057
T3G.D	**0.391**	**0.024**
T4G.D	**0.417**	**0.012**
T5G.D	**0.416**	**0.025**
T6G.D	**0.417**	**0.026**

W=water bodies, G=genetic distance, D=great-circle distance, A=altitude, T=trade route, AG.D is a partial Mantel test between the altitude [A] and the genetic distance [G] matrices with the great-circle distance matrix [D] partialled out. p-values are shown in bold typeface for significant correlations.

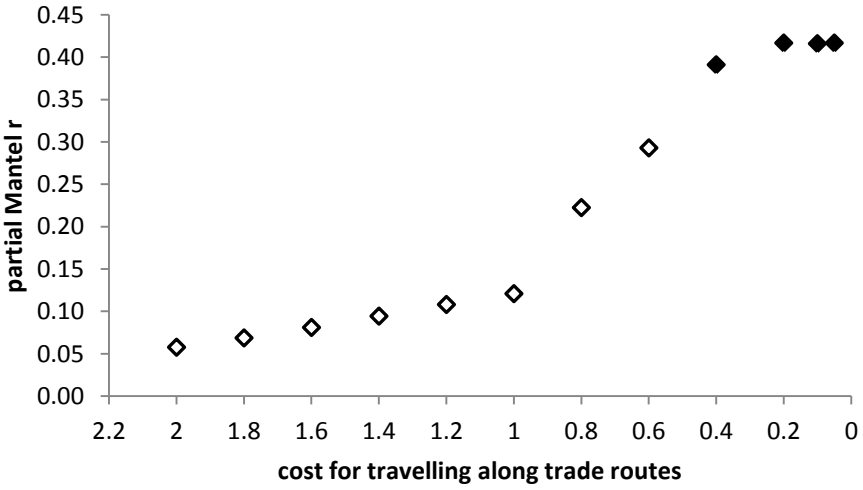

Figure 5.4

Partial Mantel coefficient between pairwise genetic and trade route distances, partialling out great-circle distances. Correlations significant at the 0.05 level are indicated by closed symbols, and those that are not significant are indicated by open symbols.

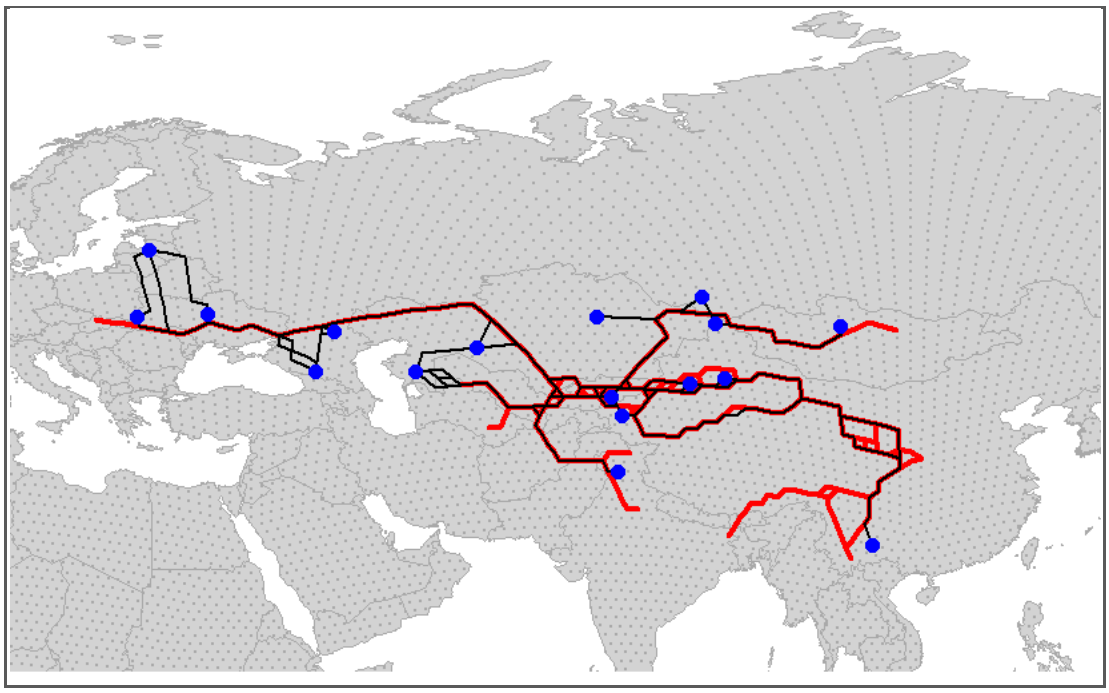

Figure 5.5
Least cost paths (black lines) among 17 Eastern Eurasian horse populations (blue dots) for the scenario with a 0.2:1 trade route (red lines) : non-trade route (grey nodes) cost ratio.

5.4 Discussion

In this chapter, investigated the effect of inland water bodies, altitude, and the main trade routes connecting West, East, and South Asia ("the Silk Roads") on the genetic structure of non-breed horses from across eastern Eurasia. I showed that the trade routes considered here were the most important factor shaping the genetic structure of our study populations, whereas altitude was the least important. The combined results suggest that the Silk Roads facilitated high levels of gene flow among horse populations from across Eurasia, regardless of landscape topography.

The most challenging mountain ranges for horse traders to cross would have been the Himalayan Mountains and the Tibetan Plateau (Figure 5.1). Incidentally, these two mountain ranges separated the steppe-empires in northern Eurasia from the agriculture-based states in India and ancient China. Unlike the empires from the northern steppes, both India and historic China suffered from a lack of suitable horse pastures and expertise in raising horses (Beckwith 1991; Deng 1997; Yang 2004, Gommans, 1994). As a consequence, both regions relied on regular imports of horses right until the advent of modern transport (Beckwith 1991). Judging from historic records, the number of horses that both India and historic China regularly obtained from the nomadic steppe empires in the north must have been considerable.During both the T'ang (618–907 AD) and the Song Dynasties (960–1279 AD), China imported tens of thousands of horses per year (Beckwith 1991). Similarly, up until the 18th century, the number of horses imported to India per year ranged between 25,000 and 100,000 (Levi 1999).

Written accounts of horse trade in these regions indicate that the bulk of horses imported to India and historic China came via overland trade routes connecting them with the horse producing areas further north (Beckwith 1991; Gommans 1994; Deng 1997). Horses entering India via overland trade routes had to cross the Himalayan mountain range, with the two most important routes leading through the Sulaiman Mountains in present-day Pakistan and Afghanistan (Gommans 1994). Similarly, southwest China (Yunnan and Sichuan Provinces) obtained large numbers of horses through trade with Tibet via the "tea-horse-route" (Deng 1997). The observation that trade routes explained a considerably higher proportion of the genetic structure in our dataset than least-cost paths minimizing the cost of ascending to high altitudes suggests that trade routes reduced the overall cost of travelling with respect to many factors. Additional samples from southerly latitudes of Eastern Eurasia, particularly from trans-Himalayan regions, would prove very valuable in furthering our understanding of the effect of trade routes on gene flow across high altitudes.

According to historical records, horse trade also occurred in an east-to-west direction. For instance, the Nogai, a nomadic confederacy extending from the Volga River in southwest Russia to the Irtush River in northeast Russia, supplied Muscovite Russia with as many as 30,000–40,000 horses per year (Gommans 2007). In the steppes, benefits of travelling along trade routes versus non-trade routes may have included safer travel, especially during periods when strong rulers controlled considerable stretches of the trade routes, as well as opportunities to restock on supplies and to engage in trade in the numerous cities and trading posts that were strung along the Silk Roads.

While the least-cost models used here are well-suited for testing the relative performance of a set of pre-determined pathways in explaining the genetic structure of a study organism, an approach which can incorporate all possible pathways between populations, such as isolation by resistance models (McRae 2006), may reveal additional important routes of gene flow.

My study has shown that, over large spatial scales, interregional and -continental trade networks may play an important role in structuring populations of domestic animals. The trade network that is widely known as the Silk Roads has been suggested to have been part of a Eurasia-wide exchange system as early as 6000 BP (Christian 2000). Unsurprisingly, large-scale patterns of human populations in Central and East Asia also appear to have been shaped by genetic exchange along the Silk Roads (Zhao & Lee 1989; Comas *et al.* 1998; Yao *et al.* 2000; Yao *et al.* 2004; Yang *et al.* 2008). It is to be expected that the Silk Roads have shaped the genetic structure of many more plants and animals that were traded along them in high volumes, e.g. apples, which are thought to have travelled along the Silk Roads together with horses, either in saddle bags or in horse's guts (Harris *et al.* 2002).

6 Domestication in Europe

Abstract[1]

The role of European wild horses in horse domestication is poorly understood. While the fossil record for wild horses in Europe prior to horse domestication is scarce, there have been suggestions that wild populations in various European regions might have contributed to the gene pool of domestic horses. To distinguish between regions where domestic populations are mainly descended from local wild stock and those where horses were largely imported, we investigated patterns of genetic diversity in 24 European horse breeds typed at 12 microsatellite loci. The distribution of high levels of genetic diversity in Europe coincides with the distribution of predominantly open landscapes prior to domestication, as suggested by simulation-based vegetation reconstructions, with breeds from Iberia and the Caspian Sea region having significantly higher genetic diversity than breeds from previously forested regions in central Europe and the UK. My results suggest that not only the Eastern steppes, but also the Iberian Peninsula provided refugia for wild horses in the Holocene, and that the genetic contribution of these wild populations to local domestic stock may have been considerable. In contrast, the consistently low levels of diversity in central Europe and the UK suggest that domestic horses in these regions largely derive from horses that were imported from the Eastern refugium, the Iberian refugium, or both.

[1] A version of this chapter has been published in PLoS ONE

6.1 Introduction

The domestication of horses was a fundamental step in the history of humankind, providing horse-centred societies with enormous advantages over agricultural societies with regard to long-distance travel, warfare and trade. Consistent with the preference of horses for predominantly open landscapes, the earliest evidence for horse domestication (morphometric data, horse milk residues in pots, and tooth wear resembling that of frequently bitted horses) appears in the Eurasian steppes around 3500 BCE (Anthony *et al.* 1986; Outram *et al.* 2009). In a recent study, Ludwig *et al.* (2009) provide further evidence for the importance of the Eurasian steppe in horse domestication by showing that coat colours other than the wild type first arose in Siberia and Eastern Europe, probably reflecting human selection.

Around the time when the first domesticated horses appeared in the Eurasian steppes, large parts of Europe were still covered by vast expanses of dense forest (Huntley 1988), a habitat that horses avoid (Linklater *et al.* 2000). Accordingly, the fossil record for wild horses at that time is extremely scarce (Von Koenigswald 2002; Boyle 2006), suggesting that European domestic horses largely descend from stock that was imported from elsewhere in a process known as demic diffusion (Ammerman & Cavalli-Sforza 1973) (colonisation of an area through population movement, Childe 1925). On the other hand, recent mitochondrial DNA (mtDNA) sequence data from a large number of both pre-domestic and domestic horses has shown that European wild populations also contributed to the gene pool of domestic horses (Cieslak *et al.* 2010; Lira *et al.* 2010). Unfortunately, it is currently difficult to distinguish between regions in Europe where the genetic contribution of local wild horses to domestic stock was substantial, and regions where domestic stock was largely introduced, and backcrossing with local wild horses played only a minor role.

To identify primary areas of horse domestication in Europe, we investigate spatial patterns of genetic diversity in horse breeds for which empirical evidence demonstrates a historic origin in a distinct region of mainland Europe or the UK (henceforth referred to as "traditional breeds").

For the purpose of this paper, we define primary areas of horse domestication as regions where local domestic populations largely descend from local wild stock, be it through their initial recruitment to found domestic populations ("independent" domestication), their extensive introgression into local domestic populations, or both.

If there were only a few, geographically restricted regions in Europe where the genetic contribution of local wild horses to domestic stock was substantial, and if domestic populations from such areas were imported into regions where local wild stock was scarce, we would expect the former areas to have retained high levels of genetic diversity, and the latter areas to be characterised by low levels of diversity. The rationale behind this reasoning is that, as populations expand out of origins, genetic diversity will be lost as a consequence of the (usually) small population sizes involved in such expansions ("founder effect") (Handley *et al.* 2007b). Clear declines in autosomal genetic diversity (allelic richness, heterozygosity) with increasing distance from primary areas of domestication have been found in a number of livestock species, such as cattle (Loftus *et al.* 1999; Cymbron *et al.* 2005; Freeman *et al.* 2006a; Medugorac *et al.* 2009), sheep (Handley *et al.* 2007a; Peter *et al.* 2007), and goats (Cañón *et al.* 2006).

To investigate spatial patterns of autosomal genetic diversity in European horses, I assembled a unique dataset consisting of more than 1100 horses typed at 12 autosomal microsatellite loci, using both new and previously published data. The combined dataset represents the largest and most comprehensive microsatellite dataset on traditional European horse breeds to date.

6.2 Methods

6.2.1 Datasets

In this chapter, I present new genotyping data, supplemented by microsatellite genotyping data published in (Cañón *et al.* 2000; Ząbek *et al.* 2005; Glowatzki-Mullis *et al.* 2006; Luís *et al.* 2007). Individual datasets were aligned using a minimum of four reference samples each. The dataset from (Glowatzki-Mullis *et al.* 2006) had been standardised to reference samples from the ISAG Horse Comparison Test and could therefore be aligned directly. Owing to a lack of reference samples, genotypes of the Bilgoraj breed (Ząbek *et al.* 2005) could not be aligned with the rest. The Bilgoraj breed was therefore only included in comparisons of within-population diversity.

6.2.2 Choice of samples

For the final dataset, I excluded all non-European breeds as well as breeds that are known to have been introduced to various European islands. In order to maximise the chance to detect signals of domestication, I furthermore excluded modern "warmblood" breeds which, by definition, are composite breeds with varying contributions of "heavy" draft horses and "light" riding horses (Hendricks 1995). My a-priori rules for the inclusion of breeds therefore focused on pony and draft horse breeds from mainland Europe and Great Britain for which a historic founding date can be demonstrated, including breeds which are known to have been crossbred with Middle Eastern breeds and/or the English Thoroughbred. The final dataset (Table 2.3) includes 1167 individuals from 24 traditional breeds from mainland Europe and the UK.

6.2.3 DNA extraction and PCR amplification

Previously unpublished data

Genomic DNA, extracted from blood, purified and at a standard concentration of 100ng/µl was available for the samples provided by Ottmar Distl. The DNA extraction protocol for blood is published in (Druml *et al.* 2006). The DNA extraction protocol for hair and the DNA amplification protocol can be found in Chapter 2 (General methods), sections 2.2 and 2.3, respectively. Details of the markers used in this study can be found in Table 2.4.

Published datasets

The DNA extraction and PCR amplification protocols used by my collaborators can be found in the original publications (Cañón *et al.* 2000; Ząbek *et al.* 2005; Glowatzki-Mullis *et al.* 2006; Luís *et al.* 2007).

6.2.4 Data analysis

Genetic Diversity

Nei's gene diversity *H* (Nei 1987) and the inbreeding coefficient F_{IS} (Wright 1965) were estimated using FSTAT v 2.9.3.2 (Goudet 2001, updated from Goudet 1995). Allelic richness was estimated using the rarefaction algorithm implemented in the programme ADZE (Szpiech *et al.* 2008). The estimates of allelic richness were standardised to the smallest sample size in our dataset, N=17. Private alleles were determined using GDA (Lewis & Zaykin 2001). Permutation tests were carried out in FSTAT and Wilcoxon tests were carried out in R (R Development Core Team 2010). For more details see Chapter 2 (General methods), section 2.4.

Spatial interpolation of genetic diversity

Because of the uneven sampling of populations across Europe, I used an approach based on Gaussian kernel interpolation that allows for an adaptive kernel width (developed by Anders Eriksson). Using a hexagonal grid representation of Eurasia (grid points spaced approximately 110 km apart, each land grid point is connected to up to six neighbours as in (Manica *et al.* 2005; Prugnolle *et al.* 2005)). The shortest distance on land from each grid point *i* to each sample location *j* was calculated as d_{ij}. The United Kingdom, Ireland, and Shetland were connected to the rest of the graph by creating suitable "landbridges". Using Gaussian kernel interpolation, the value of genetic diversity ($\hat{H}_i$) was then calculated for each grid point *i* as

$$\hat{H}_i = \sum_{j=1}^{n} e^{-d_{ij}^2/\sigma_i^2} H_j \Big/ \sum_{j=1}^{n} e^{-d_{ij}^2/\sigma_i^2} \qquad (6.1)$$

where *n* is the number of sample locations, H_j is the genetic diversity for location *j*, and σ_i is the kernel width for the grid point *i*. Because the distribution of the sample points was rather inhomogeneous, with dense sampling in western Europe and very sparse sampling in the East, the width σ_i of the kernel in grid point *i* was chosen to be proportional to the harmonic average of the distance to the sample locations (in order to avoid artefacts from the finite resolution of the grid, distances are forced to be at least 100 km, the typical distance between neighbouring grid points):

$$\sigma_i = a \Big/ \left[\frac{1}{n} \sum_{j=1}^{n} 1/\max\left(100, d_{ij}\right) \right]. \qquad (6.2)$$

The scale factor *a* was chosen to *a*=23, such that the kernel $\exp\left(-d_{ij}^2/\sigma_i^2\right) \approx 0.5$ when the distance d_{ij} is twice the distance to the closest sample point. I used Arcview v32 (ESRI) to produce the figures from the grid point estimates.

Potential confounding effects from recent demography

The observed genetic pattern could be a consequence of recent demographic processes rather than a signal of domestication. I considered three major confounding factors that would invalidate the interpretation of genetic hotspots as primary areas of domestication: admixture, recent population declines, and population substructure.

Admixture

In the recent past, breeds from the Middle East have been widely used to "improve" horse breeds throughout Europe (Hendricks 1995). Since admixture can affect patterns of genetic diversity, we estimated the contribution of three Middle Eastern horse breeds (Arab, Akhal Teke, Caspian) to all other breeds in our dataset. I used two measures of admixture: the admixture coefficient m_Y (Bertorelle & Excoffier 1998) and expected homozygosity F_S. m_Y coefficients and standard deviations were computed as averages of 1,000 random bootstrap samples using the programme ADMIX (Bertorelle & Excoffier 1998; Dupanloup & Bertorelle 2001). The calculation of m_Y is based on the assumption that allele frequencies in the admixed populations are linear combinations of those in the parental populations; contrary to other admixture coefficients, m_Y takes into account allele frequency differences as well as the degree of molecular divergence between alleles and has been shown to be appropriate for use with microsatellite data (Bertorelle & Excoffier 1998). Since the true parental populations (i.e. European populations of wild horses) are not available, I chose the Hucul, an old breed from the area where wild horses survived the longest (Hendricks 1995), to represent the genetic component of non-Middle Eastern breeds. The relative genetic contribution of the Middle Eastern breeds to central European/UK breeds was established by individual comparison of each of the three Middle Eastern breeds with the Hucul breed. Since the surrogate parental populations chosen here are unlikely to represent the genetic variability present in the true parental populations, the resulting m_Y values merely describe the *relative* contribution of the surrogate parental populations to the admixed populations, not their absolute contributions.

Effect of population substructure on within-population heterozygosity

If mating is non-random, substructure within breeds may arise, causing a reduction in overall heterozygosity (Wahlund effect). This reduction can be measured using F_{IS} (Wright 1965). If the decreased diversity in central Europe/the UK arose because breeding practices in this area have promoted stronger population substructure than those in the proposed refugia, we would expect to see a higher proportion of positive F_{IS} values in the former.

Recent declines in population sizes

Recent bottlenecks might have contributed to the low diversity observed in central Europe and Great Britain (cE/UK), as compared to Iberia and western Asia. I explored the magnitude of the bottlenecks that would have been necessary to produce the lower median diversity found in cE/UK using the recursion

$$H_{t+1}=H_t*(1-1/2N_t), \quad (6.3)$$

where H_t is the within-population heterozygosity and N_t the effective population size in generation t. I set the initial diversity H_t equal to the median diversity observed in the putative refugial populations. This is a very conservative estimate, since it (incorrectly) assumes that the latter did not experience recent declines in population sizes.

I considered scenarios in which central European and British populations were reduced to minimum effective population sizes of N=10, 20, 30, 40, or 50 either six or three generations ago, and then recovered at an annual growth rate r equal to 1.1. Using a generation time of 12 years, the bottlenecks coincide with the 1940s and the 1970s, two periods in which many native horse breeds in Europe experienced dramatic declines in population sizes (Aberle *et al.* 2004).

6.3 Results

6.3.1 Spatial patterns of genetic diversity in traditional European horse breeds

Geographic variation in gene diversity (*H*) revealed two hotspots of diversity, one in the Caspian region of western Asia, the easternmost sampling location in this dataset, and one in the Iberian Peninsula (Figure 6.1A). A very similar pattern is obtained for allelic richness (R_S, Figure 6.1B). The Iberian hotspot coincides with the only region in central and western Europe that was characterised by appreciable expanses of open landscape in the mid-Holocene (Figure 6.1C, adapted from (Gallimore *et al.* 2005)), suggesting that not only the Eurasian steppes but also the Iberian Peninsula served as a refugium for wild horses in the early and mid-Holocene, when vast expanses of forest would have rendered most of Europe unsuitable for this steppe-adapted species.

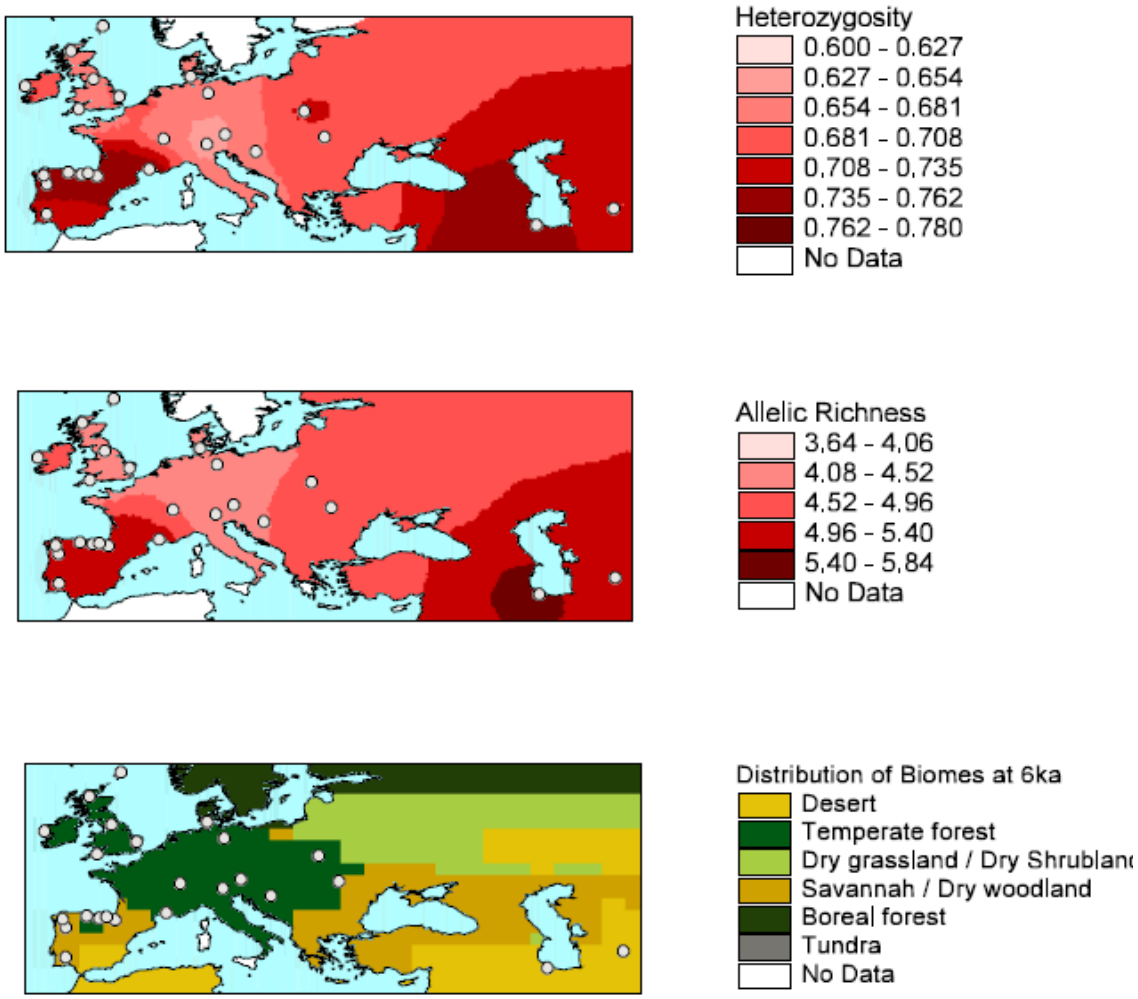

Figure 6.1

High diversity in European horses mirrors the distribution of open landscape in the mid-Holocene. A) Interpolation of expected heterozygosity *H* in 24 traditional European horse breeds. High levels of genetic diversity, as indicated by dark shading, are found in the Caspian region of western Asia and the Iberian Peninsula. White circles indicate the approximate location of origin for each breed. B) Interpolation of allelic richness R_S in 24 native European breeds using a minimum sample size of N=17. C) Spatial distribution of biomes in Europe and western Asia 6000 years ago (6ka) as inferred from model simulations. [Map adapted from Gallimore *et al*. 2005].

In a comparison of diversity between breeds from regions that were predominantly open versus those that were predominantly forested at 6 ka, I find that the latter group has significantly lower diversity (median H= 0.687, median R_S=4.42) than the former (median H=0.733, median R_S=5.09; two-sided permutation tests with 10,000 runs; H: p=0.006, Fig. 6.2A; R_S: p=0.002, Fig. 6.2B). Low levels of diversity in breeds from previously forested areas are consistent with a loss of diversity as small herds of domestic horses were imported into these areas, following their domestication in Iberia or the Eastern steppes. Estimating the relative contribution of the two refugial populations to individual breeds is not possible here due to the limited number of markers used.

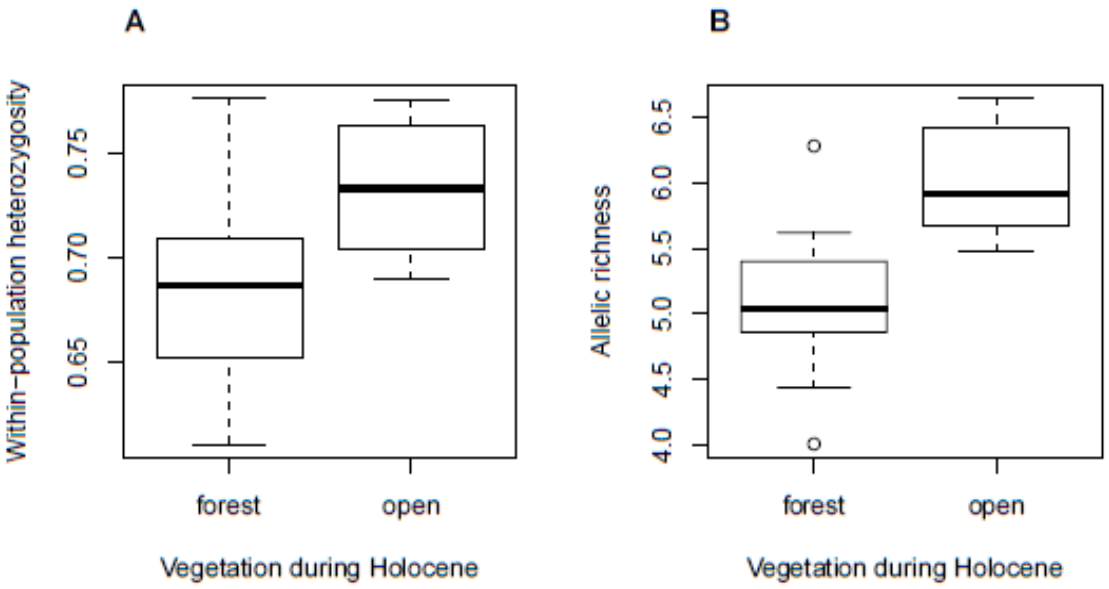

Figure 6.2

Levels of genetic diversity in Iberia (group: "open") and central Europe/Britain (group: "forested"). Average gene diversity H (A) and average allelic richness R_S (B) per group ("open": N=9; "forested": N=15). Statistical significance was determined using a two-sided permutation test (* p<0.05) and 10,000 randomisations.

6.3.2 Ancient history versus recent demography

The observed genetic pattern could be a consequence of recent demographic processes: high diversity in Iberia might reflect disproportionally high levels of admixture from non-Iberian breeds. Similarly, low diversity in central Europe and the UK (cE/UK) might reflect disproportionally severe recent bottlenecks or higher levels of inbreeding in breeds from these areas. Since domestication, horses from the Middle East have been among the most widely used to "improve" horse breeds across Europe (Hendricks 1995). I estimated the genetic component of three Middle Eastern breeds (Arab, Akhal Teke and Caspian) in Iberian and cE/UK breeds using two different measures of admixture, the admixture coefficient m_Y (Table 6.1 A-C) and expected homozygosity F_S (Table 6.2). I found no significant difference in the level of admixture from Middle Eastern breeds between Iberian and cE/UK horses (Wilcoxon tests, admixture with Arab: m_Y: W=43, p=0.877; F_S: W=72.5, p=0.086; admixture with Akhal Teke: m_Y: W=28, p=0.183; F_S: W=63.5, p=0.296; admixture with Caspian: m_Y: W=23, p=0.081; F_S: W=37, p=0.389; Fig. 3). Similarly, there is no significant difference in F_{IS} between Iberian and cE/UK breeds (Wilcoxon test, W=70.5, p=0.217; median (IQR) Iberia: 0.035 (0.007-0.052); cE/UK: -0.009 (-0.037-0.028)), implying that breeding practices are unlikely to explain the observed pattern in diversity. Furthermore, cE/UK breeds, but not Iberian breeds, would have had to undergo extreme recent contraction, to average effective population sizes (N_e) of between ten and 20 individuals, to generate the observed pattern (equation (6.3)). While a few individual breeds are known to have undergone such severe bottlenecks in the recent past, these include breeds from the proposed refugia (Firouz 1972; Cañón *et al.* 2000; Royo *et al.* 2005). Based on the evidence presented here, I infer that the observed pattern of genetic diversity is unlikely to be the result of recent demographic processes.

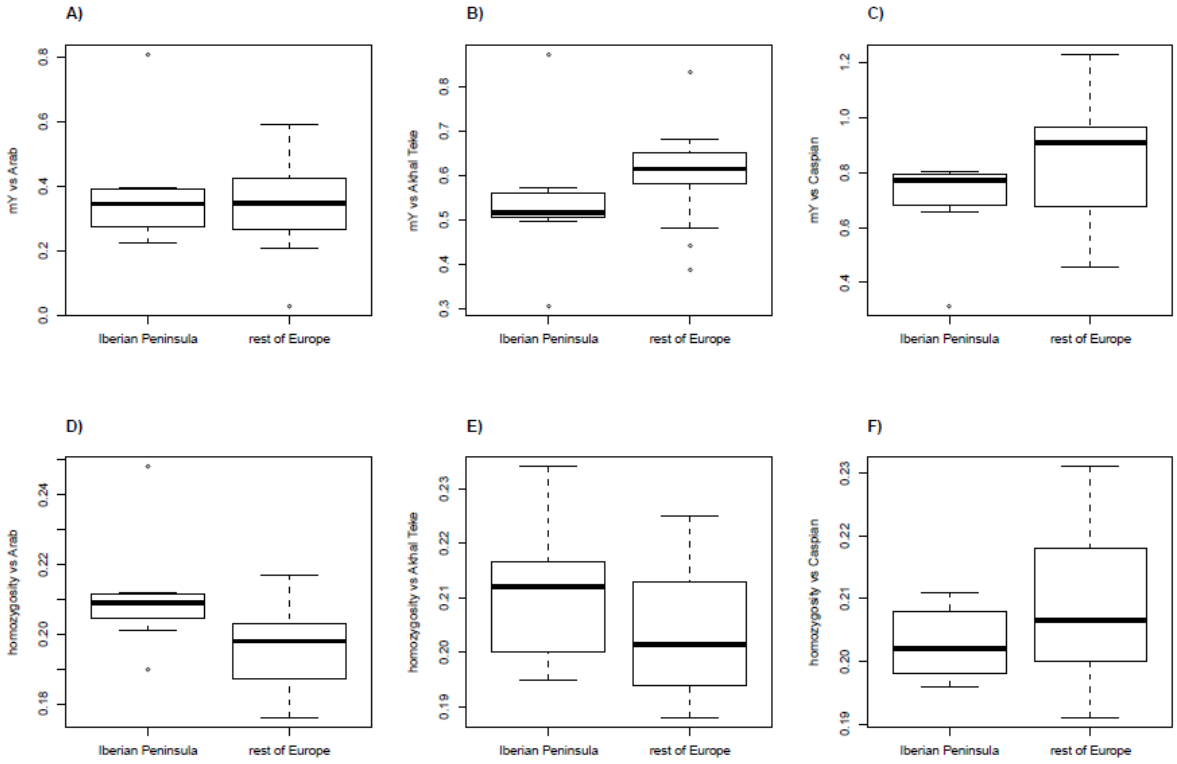

Figure 6.3

Levels of admixture from three Middle Eastern breeds in Iberia and central Europe/Britain. A-C: relative genetic contribution of the Arab (A), Akhal Teke (B) and Caspian (C) breed to Iberian and cE/UK breeds based on the admixture coefficient m_Y. D-F: relative genetic contribution of the Arab (D), Akhal Teke (E), and Caspian (F) breed to Iberian and cE/UK breeds based on expected homozygosity F_S.

Table 6.1 A)

Admixture coefficients m_Y for all breeds using Arab and Hucul as parental populations.

Breed	Group	Arab		Hucul	
		Bootstrap* average	**Bootstrap standard deviation**	**Bootstrap average**	**Bootstrap standard deviation**
Altmark Draught	Central Europe/Britain	0.4006	0.1795	0.5994	0.1795
Camargue	Central Europe/Britain	0.4849	0.0942	0.5151	0.0942
Comtois	Central Europe/Britain	0.3497	0.1845	0.6503	0.1845
Connemara	Central Europe/Britain	0.4237	0.1306	0.5763	0.1306
Dale	Central Europe/Britain	0.2307	0.1235	0.7693	0.1235
Exmoor	Central Europe/Britain	0.3232	0.1178	0.6768	0.1178
Haflinger	Central Europe/Britain	0.3242	0.1693	0.6758	0.1693
Highland	Central Europe/Britain	0.5179	0.1105	0.4821	0.1105
Noriker	Central Europe/Britain	0.2098	0.1549	0.7902	0.1549
Posavina	Central Europe/Britain	0.2673	0.1314	0.7327	0.1314
Schleswig Draught	Central Europe/Britain	0.3980	0.1278	0.6020	0.1278
Shetland	Central Europe/Britain	0.0311	0.1389	0.9689	0.1389
Suffolk Punch	Central Europe/Britain	0.5926	0.0857	0.4074	0.0857
Asturcon	Iberian Peninsula	0.2824	0.1199	0.7176	0.1199
Caballo Gallego	Iberian Peninsula	0.3472	0.0900	0.6528	0.0900
Garrano	Iberian Peninsula	0.3934	0.0987	0.6066	0.0987
Jaca Navarra	Iberian Peninsula	0.3947	0.1050	0.6053	0.1050
Losino	Iberian Peninsula	0.2244	0.0969	0.7756	0.0969
Lusitano	Iberian Peninsula	0.8069	0.0934	0.1931	0.0934
Pottoka	Iberian Peninsula	0.2722	0.0865	0.7278	0.0865

* m_Y estimates and standard deviations were determined by averaging over 1,000 random bootstrap samples

Table 6.1 B)

Admixture coefficients m_Y for all breeds using Akhal Teke and Hucul as parental populations.

Breed	Group	Akhal Teke		Hucul	
		Bootstrap* average	**Bootstrap standard deviation**	**Bootstrap average**	**Bootstrap standard deviation**
Altmark Draught	Central Europe/Britain	0.6061	0.1329	0.3939	0.1329
Camargue	Central Europe/Britain	0.6185	0.0927	0.3815	0.0927
Comtois	Central Europe/Britain	0.5890	0.1440	0.4110	0.1440
Connemara	Central Europe/Britain	0.6154	0.1030	0.3846	0.1030
Dale	Central Europe/Britain	0.6303	0.0864	0.3697	0.0864
Exmoor	Central Europe/Britain	0.4822	0.0934	0.5178	0.0934
Haflinger	Central Europe/Britain	0.6499	0.1156	0.3501	0.1156
Highland	Central Europe/Britain	0.5825	0.1038	0.4175	0.1038
Noriker	Central Europe/Britain	0.6658	0.1055	0.3342	0.1055
Posavina	Central Europe/Britain	0.4409	0.1110	0.5591	0.1110
Schleswig Draught	Central Europe/Britain	0.6823	0.1033	0.3177	0.1033
Shetland	Central Europe/Britain	0.3874	0.0947	0.6126	0.0947
Suffolk Punch	Central Europe/Britain	0.8321	0.0756	0.1679	0.0756
Asturcon	IberianPeninsula	0.4954	0.0935	0.5046	0.0935
Gallego	IberianPeninsula	0.5160	0.0710	0.4840	0.0710
Garrano	IberianPeninsula	0.5721	0.0854	0.4279	0.0854
JacaNavarra	IberianPeninsula	0.5511	0.0853	0.4489	0.0853
Losino	IberianPeninsula	0.3049	0.0904	0.6951	0.0904
Lusitano	IberianPeninsula	0.8725	0.0775	0.1275	0.0775
Pottoka	IberianPeninsula	0.5140	0.0801	0.4860	0.0801

* m_Y estimates and standard deviations were determined by averaging over 1,000 random bootstrap samples

Table 6.1 C)

Admixture coefficients m_Y for all breeds using Caspian and Hucul as parental populations.

		Caspian		Hucul	
Breed	**Group**	**Bootstrap* average**	**Bootstrap standard deviation**	**Bootstrap average**	**Bootstrap standard deviation**
Altmark Draught	Central Europe/Britain	1.1340	0.2697	-0.1340	0.2697
Camargue	Central Europe/Britain	0.6755	0.1745	0.3245	0.1745
Comtois	Central Europe/Britain	0.9095	0.3237	0.0905	0.3237
Connemara	Central Europe/Britain	0.5169	0.3521	0.4831	0.3521
Dale	Central Europe/Britain	0.8336	0.1571	0.1664	0.1571
Exmoor	Central Europe/Britain	0.6042	0.2289	0.3958	0.2289
Haflinger	Central Europe/Britain	1.2294	0.2151	-0.2294	0.2151
Highland	Central Europe/Britain	0.9483	0.9483	0.0517	0.1865
Noriker	Central Europe/Britain	0.9568	0.2146	0.0432	0.2146
Posavina	Central Europe/Britain	0.9654	0.1592	0.0346	0.1592
Schleswig Draught	Central Europe/Britain	0.8452	0.2011	0.1548	0.2011
Shetland	Central Europe/Britain	0.4581	0.2047	0.5419	0.2047
Suffolk Punch	Central Europe/Britain	1.0487	0.1838	-0.0487	0.1838
Asturcon	IberianPeninsula	0.7073	0.2201	0.2927	0.2201
Gallego	IberianPeninsula	0.8043	0.1135	0.1957	0.1135
Garrano	IberianPeninsula	0.7874	0.1425	0.2126	0.1425
JacaNavarra	IberianPeninsula	0.7716	0.1774	0.2284	0.1774
Losino	IberianPeninsula	0.3138	0.2069	0.6862	0.2069
Lusitano	IberianPeninsula	0.8019	0.3376	0.1981	0.3376
Pottoka	IberianPeninsula	0.6589	0.1213	0.3411	0.1213

* m_Y estimates and standard deviations were determined by averaging over 1,000 random bootstrap samples

Table 6.2

Expected homozygosity (F_S) of different European horse breeds with the Arab, Akhal Teke, and Caspian, respectively.

		Expected homozygosity F_S		
Breed	**Group**	**Arab**	**AkhalTeke**	**Caspian**
AltmarkDraught	Central Europe/Britain	0.18	0.20	0.21
Camargue	Central Europe/Britain	0.22	0.22	0.21
Comtois	Central Europe/Britain	0.20	0.19	0.19
Connemara	Central Europe/Britain	0.21	0.22	0.20
Dale	Central Europe/Britain	0.19	0.19	0.21
Exmoor	Central Europe/Britain	0.19	0.21	0.21
Haflinger	Central Europe/Britain	0.18	0.19	0.22
Highland	Central Europe/Britain	0.20	0.21	0.21
Hucul	Central Europe/Britain	0.20	0.19	0.22
Noriker	Central Europe/Britain	0.18	0.20	0.22
Posavina	Central Europe/Britain	0.19	0.20	0.22
SchleswigDraught	Central Europe/Britain	0.20	0.20	0.20
Shetland	Central Europe/Britain	0.21	0.23	0.23
SuffolkPunch	Central Europe/Britain	0.19	0.21	0.20
Asturcon	IberianPeninsula	0.20	0.20	0.21
Gallego	IberianPeninsula	0.21	0.21	0.20
Garrano	IberianPeninsula	0.21	0.21	0.20
JacaNavarra	IberianPeninsula	0.21	0.22	0.21
Losino	IberianPeninsula	0.21	0.20	0.20
Lusitano	IberianPeninsula	0.25	0.23	0.21
Pottoka	IberianPeninsula	0.19	0.20	0.20

6.4 Discussion

My investigation of genetic diversity in traditional European horse breeds revealed two hotspots of genetic diversity, one in the Caspian region of western Asia and one in the Iberian Peninsula. The distribution of high genetic diversity in European horses coincides with the distribution of open vegetation in the mid-Holocene, suggesting that these areas acted as refugia for wild horses at a time when most of Europe was covered by dense forest (Mitchell 2005).

A hotspot of genetic diversity in the Iberian Peninsula indicates that *E. ferus* may have also survived in Iberia. The Iberian Peninsula was the only region in central and western Europe in which appreciable expanses of open habitat persisted throughout the Holocene (Preece 1991; Pantaléon-Cano *et al.* 2003). The presence of wild horses in the Iberian Peninsula prior to domestication is supported by findings of horse remains in Neolithic and Copper Age sites (sixth to fourth millennium B.P., Chapman 1990; Uerpmann 1990). More recently, it has been shown that several pre-domestic Iberian maternal lineages survive in modern horses of Iberian descent (Cieslak *et al.* 2010; Lira *et al.* 2010), thus documenting a genetic contribution of Iberian wild stock to local domestic horses. Here, I go on to show that the genetic contribution of Iberian wild stock to local domestic horses may have been substantial: a hotspot of genetic diversity in the Iberian Peninsula is consistent with the persistence of *E. ferus* in this region from the Pleistocene through the Holocene, and the subsequent extensive use of local Iberian wild horses in establishing and/or restocking local domestic populations.

Hypotheses of local domestication in other parts of Europe could not be confirmed in this study. Levels of genetic diversity in breeds from previously forested areas were consistently low, suggesting a scenario whereby these areas primarily relied on an import of horses from either the Iberian or the Asian or both refugia (i.e. demic diffusion). This hypothesis is consistent with the fossil record for horses, which, in turn, reflects the ecology of this large, group-living animal. While my results do not imply that wild horses were entirely absent from forested parts of Holocene Europe, I suggest that their presence in these regions was spatially and temporally discontinuous, with local extinctions and re-colonisations occurring in response to natural forest gap dynamics (see also Sommer *et al.* 2011).

In this chapter, I confirm previous claims whereby populations of *E. ferus* persisted in refugial steppe habitat in the East (Stewart 2007), and provide further evidence for a second Holocene refugium for wild horses in the Iberian Peninsula. My results suggest that primary areas of horse domestication were confined to regions where considerable expanses of open landscape persisted throughout the Holocene, and that previously forested regions in Europe primarily relied on an import of domestic horses. Whether the knowledge of how to successfully capture, tame and breed horses reached Iberia through cultural transmission, or whether this knowledge was acquired independently, is an open question that cannot be answered with genetic data.

7 General Discussion

Despite decades of research into the origin and spread of horses, archaeological, mitochondrial, and Y chromosomal DNA data have allowed multiple interpretations. In this thesis, large-scale autosomal genetic data, analysed using population genetics approaches, have revealed a scenario of horse domestication which integrates key aspects of previously suggested scenarios, thus providing for the first time a coherent picture of the origin and spread of horse domestication both in the Eurasian steppes and in Europe.

7.1 The origin and spread of horse domestication

In this thesis, I showed that horse domestication originated in the western part of the Eurasian steppe, and that the eastward spread of horse domestication involved both population movement and extensive recruitment of wild horses. In Europe, the only region where local wild populations contributed considerably to local domestic stock was the Iberian Peninsula, where open landscapes have been shown to have persisted throughout the Holocene period. These findings highlight the importance of taking the ecology of the ancestral wild species in combination with palaeovegetation reconstructions into consideration when searching for primary areas of domestication.

In Europe, high levels of diversity were shown to correspond to primary areas of horse domestication, consistent with expectations whereby primary areas of domestication harbour higher levels of genetic diversity than areas into which domestic stock was introduced. In contrast, in the Eurasian steppe region, high levels of genetic diversity were shown to correspond to the entry point, and therefore the population origin, of *Equus ferus* in Eurasia.

In the Eurasian steppe region, extensive and widespread introgression of wild horses from throughout the steppe into domestic stock has led to the genetic signature of the spread of *E. ferus* across Eurasia to be preserved in modern domestic horses. In contrast, the genetic contribution of wild horses from Central Europe and Britain (if there were any) to the domestic herds that spread out of the Iberian Peninsula was much lower, strongly suggesting that the decline in genetic diversity away from the Iberian Peninsula reflects the gradual loss of genetic diversity as horses re-colonised Central Europe and Great Britain from this primary area of domestication. Together, these results suggest that, when introgression of wild animals into domestic stock is extensive and widespread, hotspots of genetic diversity may not always correspond to primary areas of domestication.

It has previously been argued by some that the observed large number of female lineages reflects "numerous successful efforts at horse domestication in different regions" (Olsen 2006a); (see also Vilà *et al.* 2001; Jansen *et al.* 2002; Bruford *et al.* 2003). The scenario of horse domestication with the greatest support (Chapter 4) suggests that a nucleus of domestic horses spread out of the western Eurasian steppe and that wild horses were incorporated into this founder stock *en route*. While it is no longer necessary to invoke numerous independent domestication events facilitated by cultural transmission to explain the high matrilineal diversity in horses, my results highlight the need to clearly define domestication. In my view, domestication is defined as an, often prolonged, process whereby domestic populations are founded *de novo* by capturing, taming, and, eventually breeding wild animals from a defined geographic area, and not the incorporation of individual wild animals from a given region into already existing domestic stock. In light of my findings, I therefore suggest that the observed high matrilineal diversity in domestic horses reflects widespread introgression of wild female horses into existing domestic stock, rather than multiple domestication events. However, I do acknowledge that, in cases where a small number of descendants of the original founder stock is restocked with a large number of wild animals from a particular geographic area, the distinction between domestication proposed above and introgression becomes difficult.

7.2 The importance of using species-specific markers of domestication

The domestication of the five major Eurasian livestock species (cattle, pigs, goats, sheep, and horses) is likely to have differed in at least some respects, owing to considerable differences in the ecology, social system, body size, tractability, and other life history traits of the wild species that gave rise to them. Markers that have been successfully used to locate primary areas of domestication in one species may therefore not be informative in others. For instance, unlike many other domestic animals, horses do not display marked size or morphological changes until rather late in the domestication process (Olsen 2006a); similarly, owing to the more or less homogeneous distribution of equine mtDNA haplogroups in Eurasia, it has been difficult to identify primary areas of horse domestication and to distinguish them from areas where introgression occurred. In cases such as the domestic horse, where it is suspected that not every ancestral haplgroup represents a domestication event, autosomal genetic data may help distinguish primary areas of domestication from zones of introgression.

7.3 Post-domestication movement

In Chapter 5, I showed that the Silk Roads played a considerable role in shaping the genetic structure of Eastern Eurasian horses, channeling gene flow along routes through deserts and across high mountain chains. The fact that I was able to trace these ancient population movements based on genetic data from 17 modern populations demonstrates once more the high resolving power of microsatellite markers and highlights the potential benefit of studying non-breed animals. Most modern breeds are subjected to intense artificial selection for narrow phenotypic traits, a process which often involves close inbreeding and the use of genetically divergent animals for breed "improvement" purposes. Since the genetic signatures associated with such practices can be very strong, the genetic signatures of demographic processes predating breed formation may be obscured in modern breeds. While I have no genetic data from breed animals in Eastern Eurasia I could compare my data with, I suspect that the genetic signature of horse movement along the Silk Roads, which is already comparatively weak in non-breed horses, would hardly be

discernible in breed horses. Using samples that carry the genetic signatures characteristic for their geographic origin is a prerequisite for accurately inferring the history of domestic populations prior to breed establishment.

7.4 Further avenues

In this thesis, I used microsatellite markers to investigate the demographic history of domestic horses. While microsatellite markers have great potential to continue to providing insights into the genetic history of domestic animals, genetic data from modern populations alone may fail to capture certain historic events. In pigs, for example, genetic data from ancient specimen revealed that the original domestic stock in Europe was of southwest Asian descent, and that this stock was later entirely replaced by the descendants of pigs that were domesticated from local European stock (Larson *et al.* 2007). Furthermore, while the presence of genetic continuity in a specific area can be inferred from the genetic signature of events known to have occurred in this area, as in the genetic signature of increased gene flow in horses along the Silk Roads, only analysis of ancient as well as modern specimens from the same geographic area will provide proof of genetic continuity. So far, mitochondrial DNA has been the marker of choice for use with ancient DNA. However, inference of genetic continuity based on uniparental markers is difficult in horses because of the broad geographic distribution of most mtDNA haplotypes. In addition, due to their small effective size, uniparental markers are strongly affected by genetic drift, so that wild and domestic animals might not carry the same haplotypes simply because of stochastic losses.

Single nucleotide polymorphisms (SNPs), if carefully selected, would provide statistically more robust estimates of the haplotype distribution in both ancient and modern samples, allowing more accurate inference of genetic continuity in addition to opening up a variety of other avenues. Unfortunately, the currently available SNP chip for horses was developed based on genomic data from seven horse breeds which, in addition, are expected to be genetically rather similar, owing to the frequent use of two of them to “upgrade” the majority of the others. Crucially, Eastern Eurasian horses were not involved in the development of equine SNPs.

Ascertainment bias is a well-known problem in studies of human demography, and the same issue is likely to apply to the current generation of equine SNPs chips. The development of SNPs, ascertained from a broad panel of horse populations, including Eastern Eurasian populations, would provide researchers with unbiased markers, enabling accurate inference of demographic processes in horses.

It would, for example, be interesting to assess the degree of admixture from foreign horse breeds in non-breed horses from different parts of Eastern Eurasia to assess in how far these horses represent distinct, indigenous populations and not admixtures of various horse breeds. The distinction between truly indigenous and breed-admixed individuals is important because the former, but not the latter, are expected to be genetically more closely related to early domestic horses, and thus more informative regarding horse population history (Boyko *et al.* 2009). In the absence of non-admixed animals, knowledge of admixture levels in local populations would enable researchers to at least account for the effect of increased admixture, thus providing more accurate estimates of population history. In cases of admixture from several populations with an unknown origin, SNP data from horses sampled across a large geographic area and representing several time periods may help resolve the origin of parental populations and the temporal sequence of admixture.

SNP typing of mid-Holocene wild, and modern indigenous domestic horses from Iberia, Central Europe, the Western and the Eastern steppes, could be used to test the scenario of horse domestication proposed here. Ideally, this approach would show that wild horses from the western steppe were the primary source of genetic diversity for modern horses from throughout the Eurasian steppes, and that Iberian wild populations were the primary source of the diversity found in modern Iberian horses. Dense genetic data from ancestral wild and domestic horses could also be searched for regions that might contain adaptive substitutions due to positive selection during the initial phase of horse domestication, thus reveal potential "domestication genes". In domestic dogs, such an approach has recently revealed signals near genes that have been implicated in memory formation and/or behavioural sensitisation in mouse or human studies (ryanodine receptor 3, adenylate cyclase 8, and a gene responsible for Williams–Beuren syndrome in humans which is characterised by social traits such as exceptional gregariousness (VonHoldt *et al.* 2010).

References

Aberle K.S., Hamann H., Drögemüller C. & Distl O. (2004) Genetic diversity in German draught horse breeds compared with a group of primitive, riding and wild horses by means of microsatellite DNA markers. *Animal Genetics* **35**, 270-7.

Achilli A., Bonfiglio S., Olivieri A., Malusa A., Pala M., Kashani B.H., Perego U.A., Ajmone-Marsan P., Liotta L., Semino O., Bandelt H.J., Ferretti L. & Torroni A. (2009) The Multifaceted Origin of Taurine Cattle Reflected by the Mitochondrial Genome. *PLoS ONE* **4**.

Achilli A., Olivieri A., Pellecchia M., Uboldi C., Colli L., Al-Zahery N., Accetturo M., Pala M., Kashani B.H., Perego U.A., Battaglia V., Fornarino S., Kalamati J., Houshmand M., Negrini R., Semino O., Richards M., Macaulay V., Ferretti L., Bandelt H.-J., Ajmone-Marsan P. & Torroni A. (2008) Mitochondrial genomes of extinct aurochs survive in domestic cattle. *Current Biology* **18**, R157-R8.

Albarella U., Dobney K. & Rowley-Conwy P. (2006) The domestication of the pig (*Sus scrofa*): new challenges and approaches In: *Documenting Domestication: New Genetic and Archaeological Paradigms* (ed. by Zeder M.A., Bradley D.G., Emshwiller E. & Smith B.D.). University of California Press, Berkeley and Los Angeles.

Ammerman A.J. & Cavalli-Sforza L.L. (1973) A Population Model for the Diffusion of Early Farming in Europe. In: *The Explanation of Culture Change* (ed. by Renfrew C.), pp. 343–358, Duckworth, London.

Amos W., Hoffman J.I., Frodsham A., Zhang L., Best S. & Hill A.V.S. (2007) Automated binning of microsatellite alleles: problems and solutions. *Molecular Ecology Notes* **7**, 10-4.

Anthony D.W. (1995) Horse, wagon and chariot: Indo-European languages and archaeology. *Antiquity* **69**, 554-65.

Anthony D.W., Bogucki P., Comşa E., Gimbutas M., Jovanović B., Mallory J.P. & Milisaukas S. (1986) The "Kurgan Culture," Indo-European origins, and the domestication of the horse: a reconsideration [and comments and replies]. *Current Anthropology* **27**, 291-313.

Anthony D.W. & Brown D.R. (1991) The origins of horseback riding. *Antiquity* **65**, 22-38.

Anthony D.W. & Brown D.R. (2000) Eneolithic horse exploitation in the Eurasian steppes: diet, ritual and riding. *Antiquity* **74**, 75-86.

Anthony D.W. & Brown D.R. (2003) Eneolithic horse rituals and riding in the steppes: new evidence. In: *Prehistoric Steppe Adaptation and the Horse* (ed. by Levine M., Renfrew C. & Boyle K.), pp. 55–68, McDonald Institute for Archaeological Research, Cambridge.

Anthony D.W., Brown D.R. & George C. (2006) Early horseback riding and warfare: the importance of the magpie around the neck. In: *Horses and Humans: the evolution of human-equine relationships* (ed. by Olsen S.L., Grant S., Choyke A.M. & Bartosiewicz L.), pp. 137-57. Archaeopress, Oxford.

Azzaroli A. (1983) Quaternary mammals and the "end-Villafranchian" dispersal event - a turning point in the history of Eurasia. *Palaeogeography, Palaeoclimatology, Palaeoecology* **44**, 117-39.

Beckwith C.I. (1991) The impact of the horse and silk trade on the economies of T'ang China and the Uighur Empire: on the importance of international commerce in the early Middle Ages. *Journal of the Economic and Social History of the Orient* **34**, 183-98.

Beja-Pereira A., Caramelli D., Lalueza-Fox C., Vernesi C., Ferrand N., Casoli A., Goyache F., Royo L.J., Conti S., Lari M., Martini A., Ouragh L., Magid A., Atash A., Zsolnai A., Boscato P., Triantaphylidis C., Ploumi K., Sineo L., Mallegni F., Taberlet P., Erhardt G., Sampietro L., Bertranpetit J., Barbujani G., Luikart G. & Bertorelle G. (2006) The origin of European cattle: Evidence from modern and ancient DNA. *Proceedings of the National Academy of Sciences* **103**, 8113-8.

Benecke N. (1993) Tierdomestikationen in Europa in vor- und frühgeschichtlicher Zeit - neue Daten zu einem alten Thema. In: *Jahressitzung der Römisch-Germanischen Kommission*, pp. 5-47. Verlag Phillip von Zabern Mainz am Rhein.

Benecke N. (1994a) *Der Mensch und seine Haustiere: Die Geschichte einer jahrtausendealten Beziehung*. Konrad Theiss Verlag, Stuttgart.

Benecke N. (1994b) Zur Domestikation des Pferdes in Mittel-und Osteuropa. Einige neue archäozoologische Befunde. In: *Die Indogermanen und das Pferd* (ed. by Hänsel B. & Zimmer S.), pp. 123-44. Archaeolingua Alapítvány, Budapest.

Benecke N. (2006) Late prehistoric exploitation of horses in central Germany and neighbouring areas: the archaeozoological record. In: *Horses and humans: the evolution of human-equine relationships* (ed. by Olsen S.L., Grant S., Choyke A.M. & Bartosiewicz L.), pp. 195-208. BAR International Series, Oxford.

Berthouly C., Do Ngoc D., Thévenon S., Bouchel D., Nhu Van T., Danes C., Grosbois V., Hoang Thanh H., Vu Chi C. & Maillard J.C. (2009) How does farmer connectivity influence livestock genetic structure? A case-study in a Vietnamese goat population. *Molecular Ecology* **18**, 3980-91.

Bertorelle G. & Excoffier L. (1998) Inferring admixture proportions from molecular data. *Molecular Biology and Evolution* **15**, 1298-311.

Bibikova V.I. (1975) Formen der Viehzucht bei den äneolithischen Stämmen Südosteuropas. In: *Moderne Probleme der Archäologie* (ed. by Otto K.-H. & Brachmann H.-J.), pp. 237-45. Akademie Verlag, Berlin.

Bibikova V.I. (1986a) A study of the earliest domestic horses of Eastern Europe. In: *Dereivka - a settlement and cemetery of Copper Age horse keepers on the Middle Dnieper* (ed. by Telegin DY), pp. 135-62. BAR International Series.

Bibikova V.I. (1986b) On the history of horse domestication in south-east Europe. In: *Dereivka - a settlement and cemetery of Copper Age horse keepers on the Middle Dnieper* (ed. by Telegin D.Y.), pp. 163-82. BAR International Series.

Binns M.M., Holmes N.G., Rolliman A. & Scott A.M. (1995) The identification of polymorphic microsatellite loci in the horse and their use in thoroughbred parentage testing. *British Veterinary Journal* **151**, 9-15.

Bjørnstad G. & Røed K.H. (2001) Breed demarcation and potential for breed allocation of horses assessed by microsatellite markers. *Animal Genetics* **32**, 59-65.

Bogucki P. (1996) The spread of early farming in Europe. *American Scientist* **84**, 242–53.

Bohonak A.J. (2002) IBD (Isolation By Distance): a program for analyses of isolation by distance. *Journal of Heredity* **93**, 153-4.

Bökönyi S. (1974) *History of domestic mammals in Central and Eastern Europe*. Akadémiai Kiadó, Budapest.

Bökönyi S. (1993) *Pferdedomestikation, Haustierhaltung und Ernährung*. Archaeolingua, Budapest.

Bollongino R., Edwards C.J., Alt K.W., Burger J. & Bradley D.G. (2006) Early history of European domestic cattle as revealed by ancient DNA. *Biology Letters* **2**, 155-9.

Boyd L. & Houpt K.A. (1994) Activity patterns. In: *Przewalski's Horse, the History and Biology of an Endangered Species* (ed. by Boyd L. & Houpt K.A.), pp. 195-228. State University of New York Press, Albany.

Boyko A.R., Boyko R.H., Boyko C.M., Parker H.G., Castelhano M., Corey L., Degenhardt J.D., Auton A., Hedimbi M., Kityo R., Ostrander E.A., Schoenebeck J., Todhunter R.J., Jones P. & Bustamante C.D. (2009) Complex population structure in African village dogs and its implications for inferring dog domestication history. *Proceedings of the National Academy of Sciences* **106**, 13903-8.

Boyle K.V. (2006) Neolithic wild game animals in Western Europe: the question of hunting. In: *Animals in the Neolithic of Britain and Europe* (ed. by Serjeantson D. & Field D.), pp. 10-23. Oxbow Books, Oxford.

Bradley D.G., MacHugh D.E., Cunningham P. & Loftus R.T. (1996) Mitochondrial diversity and the origins of African and European cattle. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 5131-5.

Bradley D.G., MacHugh D.E., Loftus R.T., Sow R.S., Hoste C.H. & Cunningham E.P. (1994) Zebu-taurine variation in Y chromosomal DNA: a sensitive assay for genetic introgression in West African trypanotolerant cattle populations. *Animal Genetics* **25**, 7-12.

Bradley D.G. & Magee D.A. (2006) Genetics and the origins of cattle. In: *Documenting Domestication: New Genetic and Archaeological Paradigms* (ed. by Zeder M.A., Bradley D.G., Emshwiller E. & Smith B.D.), pp. 317-28. University of California Press.

Breen M., Lindgren G., Binns M.M., Norman J., Irvin Z., Bell K., Sandberg K. & Ellegren H. (1997) Genetical and physical assignments of equine microsatellites - first integration of anchored markers in horse genome mapping. *Mammalian Genome* **8**, 267-73.

Brown D. & Anthony D. (1998) Bit wear, horseback riding and the Botai site in Kazakhstan. *Journal of Archaeological Science* **25**, 331-47.

Bruford M.W., Bradley D.G. & Luikart G. (2003) DNA markers reveal the complexity of livestock domestication. *Nature Reviews Genetics* **4**, 900-10.

Cañón J., Checa M.L., Carleos C., Vega-Pla J.L., Vallejo M. & Dunner S. (2000) The genetic structure of Spanish Celtic horse breeds inferred from microsatellite data. *Animal Genetics* **31**, 39-48.

Cañón J., García D., García-Atance M.A., Obexer-Ruff G., Lenstra J.A., Ajmone-Marsan P. & Dunner S. (2006) Geographical partitioning of goat diversity in Europe and the Middle East. *Animal Genetics* **37**, 327-34.

Carrión J.S., Munuera M., Dupré M. & Andrade A. (2001) Abrupt vegetation changes in the Segura Mountains of southern Spain throughout the Holocene. *Journal of Ecology* **89**, 783-97.

Chakravarti R. (1991) Horse trade and piracy at Tana (Thana, Maharashtra, India): gleanings from Marco Polo. *Journal of the Economic and Social History of the Orient* **34**, 159-82.

Chapman R. (1990) *Emerging Complexity: the Later Prehistory of South-East Spain, Iberia and the West Mediterranean*. Cambridge University Press, Cambridge.

Chessa B., Pereira F., Arnaud F., Amorim A., Goyache F., Mainland I., Kao R.R., Pemberton J.M., Beraldi D., Stear M.J., Alberti A., Pittau M., Iannuzzi L., Banabazi M.H., Kazwala R.R., Zhang Y.P., Arranz J.J., Ali B.A., Wang Z.L., Uzun M., Dione M.M., Olsaker I., Holm L.E., Saarma U., Ahmad S., Marzanov N., Eythorsdottir E., Holland M.J., Ajmone-Marsan P., Bruford M.W., Kantanen J., Spencer T.E. & Palmarini M. (2009) Revealing the history of sheep domestication using retrovirus integrations. *Science* **324**, 532-6.

Childe V.G. (1925) *The dawn of European civilization*. Routledge & Kegan Paul, London.

Christian D. (2000) Silk Roads or Steppe Roads? The Silk Roads in world history. *Journal of World History* **11**, 1-26.

Cieslak M., Pruvost M., Benecke N., Hofreiter M., Morales A., Reissmann M. & Ludwig A. (2010) Origin and history of mitochondrial DNA lineages in domestic horses. *PLoS ONE* **5**, e15311.

Clutton-Brock J. (1999) *A Natural History of Domesticated Mammals*. Cambridge University Press, Cambridge.

Comas D., Calafell F., Mateu E., Pérez-Lezaun A., Bosch E., Martínez-Arias R., Clarimon J., Facchini F., Fiori G., Luiselli D., Pettener D. & Bertranpetit J. (1998) Trading

genes along the Silk Road: mtDNA sequences and the origin of Central Asian populations. *The American Journal of Human Genetics* **63**, 1824-38.

Coogle L., Reid R. & Bailey E. (1996) Equine dinucleotide repeat loci from LEX025 to LEX033. *Animal Genetics* **27**, 289.

Coulon A., Guillot G., Cosson J.F., Angibault J.M.A., Aulagnier S., Cargnelutti B., Galan M. & Hewison A.J.M. (2006) Genetic structure is influenced by landscape features: empirical evidence from a roe deer population. *Molecular Ecology* **15**, 1669-79.

Cramer B. (2002) Morphometrische Untersuchungen an Quartären Pferden in Mitteleuropa. In: *Geowissenschaftliche Fakultät*, p. 250. Eberhard-Karl Universität, Tübingen.

Crawford N.G. (2009) smogd: software for the measurement of genetic diversity. *Molecular Ecology Resources* **10**, 556-7.

Curtin P.D. (1985) *Cross-Cultural Trade in World History*. Cambridge University Press, Cambridge.

Cymbron T., Freeman A.R., Malheiro M.I., Vigne J.D. & Bradley D.G. (2005) Microsatellite diversity suggests different histories for Mediterranean and Northern European cattle populations. *Proceedings of the Royal Society B-Biological Sciences* **272**, 1837-43.

De Cupere B., Lentacker A., Van Neer W., Waelkens M. & Verslype L. (2000) Osteological evidence for the draught exploitation of cattle: first applications of a new methodology. *International Journal of Osteoarchaeology* **10**, 254-67.

Deng G. (1997) The foreign staple trade of China in the pre-modern era. *The International History Review* **19**, 253-85.

Deshpande O., Batzoglou S., Feldman M.W. & Luca Cavalli-Sforza L. (2009) A serial founder effect model for human settlement out of Africa. *Proceedings of the Royal Society B: Biological Sciences* **276**, 291-300.

Dijkstra E.W. (1959) A note on two problems in connexion with graphs. *Numerical Mathematics* **1**, 269-71.

Dixon P. (2003) VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science* **14**, 927-30.

Dobney K. & Larson G. (2006) Genetics and animal domestication: new windows on an elusive process. *Journal of Zoology* **269**, 261-71.

Druml T., Curik I., Baumung R., Aberle K., Distl O. & Sölkner J. (2006) Individual-based assessment of population structure and admixture in Austrian, Croatian and German draught horses. *Heredity* **98**, 114-22.

Dupanloup I. & Bertorelle G. (2001) Inferring admixture proportions from molecular data: extension to any number of parental populations. *Molecular Biology and Evolution* **18**, 672-5.

Edwards C.J., Bollongino R., Scheu A., Chamberlain A., Tresset A., Vigne J.D., Baird J.F., Larson G., Ho S.Y.W., Heupink T.H., Shapiro B., Freeman A.R., Thomas M.G., Arbogast R.M., Arndt B., Bartosiewicz L., Benecke N., Budja M., Chaix L., Choyke A.M., Coqueugniot E., Döhle H.J., Göldner H., Hartz S., Helmer D., Herzig B., Hongo H., Mashkour M., Özdogan M., Pucher E., Roth G., Schade-Lindig S., Schmölcke U., Schulting R.J., Stephan E., Uerpmann H.P., Vörös I., Voytek B., Bradley D.G. & Burger J. (2007) Mitochondrial DNA analysis shows a Near Eastern Neolithic origin for domestic cattle and no indication of domestication of European aurochs. *Proceedings of the Royal Society B-Biological Sciences* **274**, 1377-85.

Eisenmann V. (1996) Quarternary horses: possible candidates to domestication. In: *Proceedings of the XIII Congress Forlí,* vol. 6, no. 1, Italy.

Ellegren H., Johansson M., Sandberg K. & Andersson L. (1992) Cloning of highly polymorphic microsatellites in the horse. *Animal Genetics* **23**, 133-42.

Epstein H. (1957) The sanga cattle of East Africa. *East African Agricultural* **22**, 149-64.

Epstein H. & Mason I.L. (1984) Cattle. In: *Evolution of domesticated animals* (ed. by Mason I.L.), pp. 6-27. Longman, London.

Evanno G., Regnaut S. & Goudet J. (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611-20.

Excoffier L., Foll M. & Petit R.J. (2009) Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics* **40**, 481-501.

Falush D., Stephens M. & Pritchard J.K. (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**, 1567-87.

Fang M., Larson G., Soares Ribeiro H., Li N. & Andersson L. (2009) Contrasting mode of evolution at a coat color locus in wild and domestic Pigs. *PLoS Genetics* **5**, e1000341.

Firouz L. (1972) *The Caspian Miniature Horse of Iran*. privately published.

Forsten A. (1988) Middle Pleistocene replacement of stenonid horses by caballoid horses - ecological implications. *Palaeogeography, Palaeoclimatology, Palaeoecology* **65**, 23-33.

Forsten A. (1992) Mitochondrial-DNA time-table and the evolution of *Equus*: comparison of molecular and paleontological evidence *Annales Zoologici Fennici* **28**, 301-9.

François O., Blum M.G.B., Jakobsson M. & Rosenberg N.A. (2008) Demographic history of European populations of *Arabidopsis thaliana*. *PLoS Genet* **4**, e1000075.

Freeman A.R., Bradley D.G., Nagda S., Gibson J.P. & Hanotte O. (2006a) Combination of multiple microsatellite data sets to investigate genetic diversity and admixture of domestic cattle. *Animal Genetics* **37**, 1-9.

Froehlich D.J. (2002) Quo vadis eohippus? The systematics and taxonomy of the early Eocene equids (Perissodactyla). *Zoological Journal of the Linnean Society* **134**, 141-256.

Fu Y.X. (1997) Statistical Tests of Neutrality of Mutations Against Population Growth, Hitchhiking and Background Selection. *Genetics* **147**, 915-25.

Fu Y.X. & Li W.H. (1993) Statistical Tests of Neutrality of Mutations. *Genetics* **133**, 693-709.

Funk W.C., Blouin M.S., Corn P.S., Maxell B.A., Pilliod D.S., Amish S. & Allendorf F.W. (2005) Population structure of Columbia spotted frogs (*Rana luteiventris*) is strongly affected by the landscape. *Molecular Ecology* **14**, 483-96.

Furrer R., Nychka D. & Sain S. (2010) fields: Tools for spatial data. R package version 6.3.

Gallimore R., Jacob R. & Kutzbach J. (2005) Coupled atmosphere-ocean-vegetation simulations for modern and mid-Holocene climates: role of extratropical vegetation cover feedbacks. *Climate Dynamics* **25**, 755-76.

Gautier M., Laloë D. & Moazami-Goudarzi K. (2010) Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds. *PLoS ONE* **5**, e13038.

Gerlach G. & Musolf K. (2000) Fragmentation of landscape as a cause for genetic subdivision in bank voles. *Conservation Biology* **14**, 1066-74.

Glowatzki-Mullis M.L., Muntwyler J., Pfister W., Marti E., Rieder S., Poncet P.A. & Gaillard C. (2006) Genetic diversity among horse populations with a special focus on the Franches-Montagnes breed. *Animal Genetics* **37**, 33-9.

Gommans J. (1994) The horse trade in eighteenth-century South Asia. *Journal of the Economic and Social History of the Orient* **37**, 228-50.

Gommans J. (2007) Warhorse and post-nomadic empire in Asia, c. 1000-1800. *Journal of Global History* **2**, 1-21.

Goto H., Ryder O.A., Fisher A.R., Schultz B., Kosakovsky Pond S.L., Nekrutenko A. & Makova K.D. (2011) A Massively Parallel Sequencing Approach Uncovers Ancient Origins and High Genetic Variability of Endangered Przewalski's Horses. *Genome Biology and Evolution* **3**, 1096-106.

Goudet J. (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity* **86**, 485-6.

Goudet J. (2001) fstat, a program to estimate and test gene diversities and fixation indices, version 2.9.3. Available from http://www.unil.ch/izea/softwares/fstat.html [updated from Goudet (1995)].

Gralak B., Coppieters W. & Van de Weghe A. (1994) Two new equine dinucleotide repeat microsatellites at the EA2C4 and the EB2E8 loci. *Animal Genetics* **25**, 285.

Grigson C. (1978) The craniology and relationships of four species of *Bos*. *Journal of Archaeological Science* **5**, 123-52.

Grigson C. (1991) An African origin for African cattle? — some archaeological evidence. *African Archaeological Review* **9**, 119-44.

Groeneveld L.F., Lenstra J.A., Eding H., Toro M.A., Scherf B., Pilling D., Negrini R., Finlay E.K., Jianlin H., Groeneveld E. & Weigend S. (2010) Genetic diversity in farm animals - a review. *Animal Genetics* **41**, 6-31.

Guérin G., Bertaud M. & Amigues Y. (1994) Characterisation of seven new horse microsatellites: HMS1, HMS2, HMS3, HMS5, HMS6, HMS7, HMS8. *Animal Genetics* **25**, 62.

Guillot G., Leblois R., Coulon A. & Frantz A.C. (2009) Statistical methods in spatial genetics. *Molecular Ecology* **18**, 4734-56.

Haile J., Froese D.G., MacPhee R.D.E., Roberts R.G., Arnold L.J., Reyes A.V., Rasmussen M., Nielsen R., Brook B.W., Robinson S., Demuro M., Gilbert M.T.P., Munch K., Austin J.J., Cooper A., Barnes I., Möller P. & Willerslev E. (2009) Ancient DNA reveals late survival of mammoth and horse in interior Alaska. *Proceedings of the National Academy of Sciences* **106**, 22352-7.

Handley L.J.L., Byrne K., Santucci F., Townsend S., Taylor M., Bruford M.W. & Hewitt G.M. (2007a) Genetic structure of European sheep breeds. *Heredity* **99**, 620-31.

Handley L.J.L., Manica A., Goudet J. & Balloux F. (2007b) Going the distance: human population genetics in a clinal world. *Trends in Genetics* **23**, 432-9.

Hanotte O., Bradley D.G., Ochieng J.W., Verjee Y., Hill E.W. & Rege J.E.O. (2002) African pastoralism: Genetic imprints of origins and migrations. *Science* **296**, 336-9.

Hanotte O., Tawah C.L., Bradley D.G., Okomo M., Verjee Y., Ochieng J. & Rege J.E.O. (2000) Geographic distribution and frequency of a taurine *Bos taurus* and an indicine *Bos indicus* Y specific allele amongst sub-Saharan African cattle breeds. *Molecular Ecology* **9**, 387-96.

Harris S.A., Robinson J.P. & Juniper B.E. (2002) Genetic clues to the origin of the apple. *Trends in Genetics* **18**, 426-30.

Hedrick P.W. (2005) A standardized genetic differentiation measure. *Evolution* **59**, 1633-8.

Hendricks B.L. (1995) *International Encyclopedia of Horse Breeds*. University of Oklahoma Press, Norman.

Hiendleder S., Kaupe B., Wassmuth R. & Janke A. (2002) Molecular analysis of wild and domestic sheep questions current nomenclature and provides evidence for domestication from two different subspecies. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **269**, 893-904.

Hofreiter M., Serre D., Poinar H.N., Kuch M. & Pääbo S. (2001) Ancient DNA. *Nature Reviews Genetics* **2**, 353-9.

Hofreiter M. & Stewart J. (2009) Ecological Change, Range Fluctuations and Population Dynamics during the Pleistocene. *Current Biology* **19**, R584-R94.

Hopman T.J., Han E.B., Story M.R., Schug M.D., Aquadro C.F., Bowling A.T., Murray J.D., Caetano A.R. & Antczak D.F. (1999) Equine dinucleotide repeat loci COR001-COR020. *Animal Genetics* **30**, 225-6.

Hubisz M.J., Falush D., Stephens M. & Pritchard J.K. (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* **9**, 1322-32.

Huntley B. (1988) Europe. In: *Vegetation history* (ed. by Huntley B, Webb III, T.). Kluwer Academic Publishers, Dordrecht.

Hutchison C.A., Newbold J.E., Potter S.S. & Edgell M.H. (1974) Maternal inheritance of mammalian mitochondrial DNA. *Nature* **251**, 536-8.

Jankovich M. (1971) *They rode into Europe*. The Long Rider's Guild Press.

Jansen T., Forster P., Levine M.A., Oelke H., Hurles M., Renfrew C., Weber J. & Olek K. (2002) Mitochondrial DNA and the origins of the domestic horse. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 10905-10.

Jombart T. (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403-5.

Jombart T., Devillard S. & Balloux F. (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genetics* **11**, 94.

Jombart T., Devillard S., Dufour A.B. & Pontier D. (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. *Heredity* **101**, 92-103.

Jost L. (2008) G(ST) and its relatives do not measure differentiation. *Molecular Ecology* **17**, 4015-26.

Jost L. (2009) D vs. G(ST): Response to Heller and Siegismund (2009) and Ryman and Leimar (2009). *Molecular Ecology* **18**, 2088-91.

Kalis A.J., Merkt J. & Wunderlich J. (2003) Environmental changes during the Holocene climatic optimum in central Europe - human impact and natural causes. *Quaternary Science Reviews* **22**, 33-79.

Kantanen J., Edwards C.J., Bradley D.G., Viinalass H., Thessler S., Ivanova Z., Kiselyova T., Cinkulov M., Popov R., Stojanovic S., Ammosov I. & Vilkki J. (2009) Maternal and paternal genealogy of Eurasian taurine cattle (*Bos taurus*). *Heredity* **103**, 404-15.

Kavar T. & Dovc P. (2008) Domestication of the horse: genetic relationships between domestic and wild horses. *Livestock Science* **116**, 1-14.

Kijas J.M.H. & Andersson L. (2001) A phylogenetic study of the origin of the domestic pig estimated from the near-complete mtDNA genome. *Journal of Molecular Evolution* **52**, 302-8.

Kijas J.W., Townley D., Dalrymple B.P., Heaton M.P., Maddox J.F., McGrath A., Wilson P., Ingersoll R.G., McCulloch R., McWilliam S., Tang D., McEwan J., Cockett N., Oddy V.H., Nicholas F.W., & Raadsma H. for the International Sheep Genomics C. (2009) A genome-wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS ONE* **4**, e4668.

Koch P.L. & Barnosky A.D. (2006) Late Quaternary Extinctions: State of the Debate. *Annual Review of Ecology, Evolution, and Systematics* **37**, 215-50.

Kosharov A.N., Pern E.M. & Rozhdestvenskaya G.A. (1989) Horses. In: *Animal genetic resources of the USSR* (ed. by Dmitriev N.G. & Ernst L.K.), pp. 272-343. Food and Agriculture Organization (FAO) of the United Nations, Rome.

Kosintsev P.A. (2006) The human-horse relationship on the Eropean-Asian border in the Neolithic and early Iron Age. In: *Horses and Humans: the evolution of human-equine relationships* (ed. by Olsen S.L., Grant S., Choyke A.M. & Bartosiewicz L.), pp. 127-37. Archaeopress, Oxford.

Kurtén B. & Anderson E. (1980) *Pleistocene mammals of North America*. Columbia University Press, New York.

Kuz'mina I.E. (1993) The horse of Botai [Loshadi Botaya]. In: *Problems in the Reconstruction of Economy and Technology from Archaeological Data [Problemi Rekonstruktsii Khozyaistba i Tekhnologii po Dannim Arkheologii]* (ed. by Zaibert V.F., Aleksashenko H.A. & Myaksheva O.V.), pp. 178–88, Petropavlovsk.

Laloë D., Moazami-Goudarzi K., Lenstra J.A., Marsan P.A., Azor P., Baumung R., Bradley D.G., Bruford M.W., Cañón J., Dolf G., Dunner S., Erhardt G., Hewitt G., Kantanen J., Obexer-Ruff G., Olsaker I., Rodellar C., Valentini A., Wiener P., European Cattle Genetic Diversity C. & Econogene C. (2010) Spatial trends of genetic variation of domestic ruminants in Europe. *Diversity* **2**, 932-45.

Lang G. (1994) *Quarternäre Vegetationsgeschichte Europas*. Spektrum Akademischer Verlag, Stuttgart.

Larson G., Albarella U., Dobney K., Rowley-Conwy P., Schibler J., Tresset A., Vigne J.-D., Edwards C.J., Schlumbaum A., Dinu A., Bălăçsescu A., Dolman G., Tagliacozzo A., Manaseryan N., Miracle P., Van Wijngaarden-Bakker L., Masseti M., Bradley D.G. & Cooper A. (2007) Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proceedings of the National Academy of Sciences* **104**, 15276-81.

Larson G., Dobney K., Albarella U., Fang M.Y., Matisoo-Smith E., Robins J., Lowden S., Finlayson H., Brand T., Willerslev E., Rowley-Conwy P., Andersson L. & Cooper A. (2005) Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* **307**, 1618-21.

Larson G., Liu R., Zhao X., Yuan J., Fuller D., Barton L., Dobney K., Fan Q., Gu Z., Liu X.-H., Luo Y., Lv P., Andersson L. & Li N. (2010) Patterns of East Asian pig domestication, migration, and turnover revealed by modern and ancient DNA. *Proceedings of the National Academy of Sciences* **107**, 7686-91.

Legendre P. & Legendre L. (1998) *Numerical Ecology*. Elsevier.

Lei C.Z., Su R., Bower M.A., Edwards C.J., Wang X.B., Weining S., Liu L., Xie W.M., Li F., Liu R.Y., Zhang Y.S., Zhang C.M. & Chen H. (2009) Multiple maternal origins of native modern and ancient horse populations in China. *Animal Genetics* **40**, 933-44.

Leng L. & Zhang D.-X. (2011) Measuring population differentiation using G_{ST} or D? A simulation study with microsatellite DNA markers under a finite island model and nonequilibrium conditions. *Molecular Ecology*, 2494-509.

Levine M. (1990) Dereivka and the problem of horse domestication. *Antiquity* **64**, 727-40.

Levine M. (2004) Exploring the criteria for early horse domestication. In: *Traces of ancestry: studies in honour of Colin Renfrew* (ed. by Jones M.), pp. 115-26. McDonald Institute Monographs, Cambridge.

Levine M.A. (1999a) Botai and the origins of horse domestication. *Journal of Anthropological Archaeology* **18**, 29-78.

Levine M.A. (1999b) The origins of horse husbandry on the Eurasian steppe. In: *Late prehistoric exploitation of the Eurasian steppe* (ed. by Levine M.A., Rassamakin Y., Kislenko A. & Tatarintseva N.), pp. 5-58. McDonald Institute Monographs, Cambridge.

Levine M.A., Bailey G., Whitwell K.E. & Jeffcott L.B. (2000) Palaeopathology and horse domestication: the case of some Iron Age horses from the Altai Mountains, Siberia. In: *Human Ecodynamics. Symposia of the Association for Environmental Archaeology* (ed. by Bailey G., Charles R. & Winder N.), pp. 123-33. Oxbow Books.

Levinson G. & Gutman G.A. (1987) Slipped-strand mispairing - a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution* **4**, 203-21.

Lewis G.E. (1937) A new Siwalik correlation. *American Journal of Science* **s5-33**, 191-204.

Lewis P.O. & Zaykin D. (2001) gda (genetic data analysis): computer program for the analysis of allelic data. University of Connecticut, Storrs, CT.

Lindgren G., Backstrom N., Swinburne J., Hellborg L., Einarsson A., Sandberg K., Cothran G., Vilà C., Binns M. & Ellegren H. (2004) Limited number of patrilines in horse domestication. *Nature Genetics* **36**, 335-6.

Lindsay E.H., Opdyke N.D. & Johnson N.M. (1980) Pliocene dispersal of the horse *Equus* and late Cenozoic mammalian dispersal events. *Nature* **287**, 135-8.

Ling Y., Yuehui M., Weijun G., Yuejiao C., Yanping W., Jianlin H., Dapeng J., Lai M. & Mahmut H. (2010a) Identification of Y chromosome genetic variation in Chinese indigenous horse breeds. *Journal of Heredity* **101**, 639-43.

Ling Y.H., Ma Y.H., Guan W.J., Cheng Y.J., Wang Y.P., Han J.L., Mang L., Zhao Q.J., He X.H., Pu Y.B. & Fu B.L. (2010b) Evaluation of the genetic diversity and population structure of Chinese indigenous horse breeds using 27 microsatellite markers. *Animal Genetics* **42**, 56-65.

Linklater W.L., Cameron E.Z., Stafford K.J. & Veltman C.J. (2000) Social and spatial structure and range use by Kaimanawa wild horses (*Equus caballus*: Equidae). *New Zealand Journal of Ecology* **24**, 139-52.

Linz B., Balloux F., Moodley Y., Manica A., Liu H., Roumagnac P., Falush D., Stamer C., Prugnolle F., van der Merwe S.W., Yamaoka Y., Graham D.Y., Perez-Trallero E., Wadstrom T., Suerbaum S. & Achtman M. (2007) An African origin for the intimate association between humans and *Helicobacter pylori*. *Nature* **445**, 915-8.

Lira J., Linderholm A., Olaria C., Brandström M.D., Gilbert M.T.P., Ellegren H., Willerslev E., Lidén K., Arsuaga J.L. & Götherström A. (2010) Ancient DNA reveals traces of Iberian Neolithic and Bronze Age lineages in modern Iberian horses. *Molecular Ecology* **19**, 64-78.

Lister A.M. (2001) Tales from the DNA of domestic horses. *Science* **292**, 218-.

Lister A.M., Kadwell M., Kaagan L.M., Jordan W.C., Richards M.B. & Stanley H.F. (1998) Ancient and modern DNA in a study of horse domestication. *Ancient Biomolecules* **2**, 267-80.

Liu H., Prugnolle F., Manica A. & Balloux F. (2006) A geographically explicit genetic model of worldwide human-settlement history. *American Journal of Human Genetics* **79**, 230-7.

Loftus R.T., Ertugrul O., Harba A.H., El-Barody M.A.A., Machugh D.E., Park S.D.E. & Bradley D.G. (1999) A microsatellite survey of cattle from a centre of origin: the Near East. *Molecular Ecology* **8**, 2015-22.

Loftus R.T., Machugh D.E., Bradley D.G., Sharp P.M. & Cunningham P. (1994) Evidence for 2 independent domestications of cattle. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 2757-61.

Ludwig A., Pruvost M., Reissmann M., Benecke N., Brockmann G.A., Castanos P., Cieslak M., Lippold S., Llorente L., Malaspinas A.S., Slatkin M. & Hofreiter M. (2009) Coat color variation at the beginning of horse domestication. *Science* **324**, 485-.

Luikart G., Fernandez H., Mashkour M., England P.R. & Taberlet P. (2006) Origins and disffusion of domestic goats inferred from DNA markers. In: *Documenting Domestication: New Genetic and Archaeological Paradigms* (ed. by Zeder M.A., Bradley D.G., Emshwiller E. & Smith B.D.), pp. 294-305. University of California Press.

Luikart G., Gielly L., Excoffier L., Vigne J.D., Bouvet J. & Taberlet P. (2001) Multiple maternal origins and weak phylogeographic structure in domestic goats. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 5927-32.

Luís C., Juras R., Oom M.M. & Cothran E.G. (2007) Genetic diversity and relationships of Portuguese and other horse breeds based on protein and microsatellite loci variation. *Animal Genetics* **38**, 20-7.

MacFadden B.J. (2005) Fossil horses: evidence for evolution. *Science* **307**, 1728-30.

Manceau V., Després L., Bouvet J. & Taberlet P. (1999) Systematics of the genus *Capra* inferred from mitochondrial DNA sequence data. *Molecular Phylogenetics and Evolution* **13**, 504-10.

Manel S., Schwartz M.K., Luikart G. & Taberlet P. (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution* **18**, 189-97.

Manica A., Amos W., Balloux F. & Hanihara T. (2007) The effect of ancient population bottlenecks on human phenotypic variation. *Nature* **448**, 346-U6.

Manica A., Prugnolle F. & Balloux F. (2005) Geography is a better determinant of human genetic differentiation than ethnicity. *Human Genetics* **118**, 366-71.

Mantel N. (1967) Detection of disease clustering and a generalized regression approach. *Cancer Research* **27**, 209-&.

Marklund S., Ellegren H., Eriksson S., Sandberg K. & Andersson L. (1994) Parentage testing and linkage analysis in the horse using a set of highly polymorphic microsatellites. *Animal Genetics* **25**, 19-23.

McGahern A., Bower M.A.M., Edwards C.J., Brophy P.O., Sulimova G., Zakharov I., Vizuete-Forster M., Levine M., Li S., MacHugh D.E. & Hill E.W. (2006) Evidence for biogeographic patterning of mitochondrial DNA sequences in Eastern horse populations. *Animal Genetics* **37**, 494-7.

McRae B.H. (2006) Isolation by resistance. *Evolution* **60**, 1551-61.

Meadows J.R.S., Li K., Kantanen J., Tapio M., Sipos W., Pardeshi V., Gupta V., Calvo J.H., Whan V., Norris B. & Kijas J.W. (2005) Mitochondrial sequence reveals high levels of gene flow between breeds of domestic sheep from Asia and Europe. *Journal of Heredity* **96**, 494-501.

Medugorac I., Medugorac A., Russ I., Veit-Kensch C.E., Taberlet P., Luntz B., Mix H.M. & Forster M. (2009) Genetic diversity of European cattle breeds highlights the conservation value of traditional unselected breeds with high effective population size. *Molecular Ecology* **18**, 3394-410.

Megens H.J., Crooijmans R., Cristobal M.S., Hui X., Li N. & Groenen M.A.M. (2008) Biodiversity of pig breeds from China and Europe estimated from pooled DNA samples: differences in microsatellite variation between two areas of domestication. *Genetics Selection Evolution* **40**, 103-28.

Meyer A.H., Valberg S.J., Hillers K.R., Schweitzer J.K. & Mickelson J.R. (1997) Sixteen new polymorphic equine microsatellites. *Animal Genetics* **28**, 69-70.

Mitchell F.J.G. (2005) How open were European primeval forests? Hypothesis testing using palaeoecological data. *Journal of Ecology* **93**, 168-77.

Moragues M., Comadran J., Waugh R., Milne I., Flavell A. & Russell J. (2010) Effects of ascertainment bias and marker number on estimations of barley diversity from high-throughput SNP genotype data. *TAG Theoretical and Applied Genetics* **120**, 1525-34.

Moran P. (1948) The interpretation of statistical maps. *Journal of the Royal Statistical Society Series B* **10**, 243–51.

Morey D.F. (1992) Size, shape and development in the evolution of the domestic dog. *Journal of Archaeological Science* **19**, 181-204.

Naderi S., Rezaei H.R., Pompanon F., Blum M.G.B., Negrini R., Naghash H.R., Balkiz O., Mashkour M., Gaggiotti O.E., Ajmone-Marsan P., Kence A., Vigne J.D. & Taberlet P. (2008) The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 17659-64.

Naderi S., Rezaei H.R., Taberlet P., Zundel S., Rafat S.A., Naghash H.R., El-Barody M.A.A., Ertugrul O. & Pompanon F. (2007) Large-scale mitochondrial DNA analysis of the domestic goat reveals six haplogroups with high diversity. *PLoS ONE* **2**.

Nei M. (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Nobis G. (1971) Vom Wildpferd zum Hauspferd: Studien zur Phylogenie pleistozäner Equiden Eurasiens und das Domestikationsproblem unserer Hauspferde. In: *Fundamenta, Monographien zur Urgeschichte*, Series B, Volume 6. Böhlau, Köln.

Oakenfull E.A., Lim H. & Ryder O. (2000) A survey of equid mitochondrial DNA: Implications for the evolution, genetic diversity and conservation of *Equus*. *Conservation Genetics* **1**, 341-55.

Oksanen J., Guillaume Blanchet F., Kindt R., Legendre P., O'Hara R.B., Simpson G.L., Solymos P., Henry M., Stevens H. & Wagner H. (2011) vegan: Community Ecology Package. R package version 1.17-10.

Olsen S.J. (1988) The horse in ancient China and its cultural influence in some other areas. *Proceedings of the Academy of Natural Sciences of Philadelphia* **140**, 151-89.

Olsen S.L. (2006a) Early Horse Domestication on the Eurasian Steppe. In: *Documenting Domestication: New Genetic and Archaeological Paradigms* (eds. by Zeder M.A., Bradley D.G., Emshwiller E. & Smith B.D.), pp. 245-69. University of California Press.

Olsen S.L. (2006b) Early horse domestication: weighing the evidence. In: *Horses and Humans: the evolution of human-equine relationships* (ed. by Olsen SL, Grant S, Choyke AM & Bartosiewicz L), pp. 81-115. Archaeopress, Oxford.

Outram A.K., Stear N.A., Bendrey R., Olsen S., Kasparov A., Zaibert V., Thorpe N. & Evershed R.P. (2009) The earliest horse harnessing and milking. *Science* **323**, 1332-5.

Pantaléon-Cano J., Yll E.-I., Perez-Obiol R. & Roure J.M. (2003) Palynological evidence for vegetational history in semi-arid areas of the western Mediterranean (Almeria, Spain). *The Holocene* **13**, 109-19.

Perez-Pardal L., Royo L.J., Beja-Pereira A., Curik I., Traore A., Fernandez I., Sölkner J., Alonso J., Álvarez I., Bozzi R., Chen S., de León F.A.P. & Goyache F. (2010) Y-specific microsatellites reveal an African subfamily in taurine (*Bos taurus*) cattle. *Animal Genetics* **41**, 232-41.

Peter C., Bruford M., Perez T., Dalamitra S., Hewitt G. & Erhardt G. (2007) Genetic diversity and subdivision of 57 European and Middle-Eastern sheep breeds. *Animal Genetics* **38**, 37-44.

Pilot M., Jedrzejewski W., Branicki W., Sidorovich V.E., Jedrzejewska B., Stachura K. & Funk S.M. (2006) Ecological factors influence population genetic structure of European grey wolves. *Molecular Ecology* **15**, 4533-53.

Preece R.C. (1991) Radiocarbon-dated molluscan successions from the Holocene of central Spain. *Journal of Biogeography* **18**, 409-26.

Prentice C., Guiot J., Huntley B., Jolly D. & Cheddadi R. (1996) Reconstructing biomes from palaeoecological data: a general method and its application to European pollen data at 0 and 6 ka. *Climate Dynamics* **12**, 185-94.

Price D.T. (2000) *Europe's first farmers*. Cambridge University Press, Cambridge.

Pritchard J.K., Wen X. & Falush D. (2009) Documentation for structure software: version 2.3.

Pritchard M., Stephens M. & Donnelly P. (2000) Inference of population structure using multilocus genotyping data. *Genetics* **155**, 945-59.

Prothero D.R. & Schoch R.M. (2002) Equus. In: *Horns, Tusks, & Flippers: The Evolution of Hoofed Mammals* (ed. by Prothero D.R. & Schoch R.M.), pp. 213-27. The John Hopkins University Press, Baltimore.

Prugnolle F., Manica A. & Balloux F. (2005) Geography predicts neutral genetic diversity of human populations. *Current Biology* **15**, R159-R60.

R Development Core Team (2010) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ramachandran S., Deshpande O., Roseman C.C., Rosenberg N.A., Feldman M.W. & Cavalli-Sforza L.L. (2005) Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15942-7.

Ramirez-Soriano A., Ramos-Onsins S.E., Rozas J., Calafell F. & Navarro A. (2008) Statistical power analysis of neutrality tests under demographic expansions, contractions and bottlenecks with recombination. *Genetics* **179**, 555-67.

Ramos-Onsins S.E. & Rozas J. (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution* **19**, 2092-100.

Ray N., Currat M. & Excoffier L. (2003) Intra-deme molecular diversity in spatially expanding populations. *Molecular Biology and Evolution* **20**, 76-86.

Riley S.P.D., Pollinger J.P., Sauvajot R.M., York E.C., Bromley C., Fuller T.K. & Wayne R.K. (2006) A southern California freeway is a physical and social barrier to gene flow in carnivores. *Molecular Ecology* **15**, 1733-41.

Rogers A.R. & Harpending H. (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution* **9**, 552-69.

Rossel S., Marshall F., Peters J., Pilgram T., Adams M.D. & O'Connor D. (2008) Domestication of the donkey: timing, processes, and indicators. *Proceedings of the National Academy of Sciences* **105**, 3715-20.

Royo L.J., Alvarez I., Beja-Pereira A., Molina A., Fernandez I., Jordana J., Gomez E., Gutierrez J.P. & Goyache F. (2005) The origins of Iberian horses assessed via mitochondrial DNA. *Journal of Heredity* **96**, 663-9.

Ruth L.S., Hopman T.J., Schug M.D., Aquadro C.F., Bowling A.T., Murray J.D., Caetano A.R. & Antczak D.F. (1999) Equine dinucleotide repeat loci COR041-COR060. *Animal Genetics* **30**, 320-1.

Sacks B.N., Bannasch D.L., Chomel B.B. & Ernest H.B. (2008) Coyotes Demonstrate How Habitat Specialization by Individuals of a Generalist Species Can Diversify Populations in a Heterogeneous Ecoregion. *Molecular Biology and Evolution* **25**, 1384-94.

Savage D.E. & Russel D.E. (1983) *Mammalian palaeofaunas of the World*. Addison-Wesley Publishing Company, Reading, Massachusetts.

Schlötterer C. (2000) Evolutionary dynamics of microsatellite DNA. *Chromosoma* **109**, 365-71.

Sherrat A. (1983) The secondary exploitation of animals in the Old World. *World Archaeology* **15**, 90-104.

Sherratt A.G. (1981) Plough and pastoralism: aspects of the secondary products revolution. In: *Pattern of the Past: Studies in honour of David Clarke* (ed. by Hodder I, Isaac G & Hammond N), pp. 261-305. Cambridge University Press, Cambridge, UK.

Slatkin M. (1993) Isolation by distance in equilibrium and nonequilibrium populations. *Evolution* **47**, 264-79.

Slatkin M. & Hudson R.R. (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129**, 555-62.

Smouse P.E., Long J.C. & Sokal R.R. (1986) Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology* **35**, 627–32.

Sokolov V.E. & Orlov V.N. (1986) Introduction of Przewalski's horses into the wild. In: *The Przewalski horse and restoration to its natural habitat in Mongolia* (Food and agriculture organization of the United Nations (FAO), Agriculture and Consumer Protection Department, Rome.

Sommer R.S., Benecke N., Lõugas L., Nelle O. & Schmölcke U. (2011) Holocene survival of the wild horse (*Equus ferus*) - a matter of open landscape? *Journal of Quaternary Science* **26,** 805-12.

Städler T., Haubold B., Merino C., Stephan W. & Pfaffelhuber P. (2009) The impact of sampling schemes on the site frequency spectrum in nonequilibrium subdivided populations. *Genetics* **182**, 205-16.

Stenseth N.C., Shabbar A., Chan K.-S., Boutin S., Rueness E.K., Ehrich D., Hurrell J.W., Lingjærde O.C. & Jakobsen K.S. (2004) Snow conditions may create an invisible barrier for lynx. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 10632-4.

Steppan K. (2006) The Neolithic human impact on wild horses in Germany and Switzerland. In: *Horses and Humans: the evolution of human-equine relationships* (ed. by Olsen S.L., Grant S., Choyke A.M. & Bartosiewicz L.), pp. 209-20. Archaeopress, Oxford.

Stewart J.R. (2007) Neanderthal extinction as part of the faunal change in Europe during Oxygen Isotope Stage 3. *Acta Zoologica Cracoviensia* **50A**, 93-124.

Sunnucks P. (2000) Efficient genetic markers for population biology. *Trends in Ecology & Evolution* **15**, 199-203.

Swinburne J.E., Lockhart L., Aldridge V., Marti E., Breen M. & Binns M.M. (2000) Characterisation of 25 new physically mapped horse microsatellite loci: AHT24-48. *Animal Genetics* **31**, 237-8.

Szpiech Z.A., Jakobsson M. & Rosenberg N.A. (2008) ADZE: a rarefaction approach for counting alleles private to combinations of populations. *Bioinformatics* **24**, 2498-504.

Taberlet P., Griffin S., Goossens B., Questiau S., Manceau V., Escaravage N., Waits L.P. & Bouvet J. (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research* **24**, 3189-94.

Tanabe K., Mita T., Jombart T., Eriksson A., Horibe S., Palacpac N., Ranford-Cartwright L., Sawai H., Sakihama N., Ohmae H., Nakamura M., Ferreira M.U., Escalante A.A., Prugnolle F., Björkman A., Färnert A., Kaneko A., Horii T., Manica A., Kishino H. & Balloux F. (2010) *Plasmodium falciparum* accompanied the human expansion out of Africa. *Current Biology* **20**, 1283-9.

Tanaka K., Iwaki Y., Takizawa T., Dorji T., Tshering G., Kurosawa Y., Maeda Y., Mannen H., Nomura K., Dang V.-B., Chhum-Phith L., Bouahom B., Yamamoto Y., Daing T. & Namikawa T. (2008) Mitochondrial diversity of native pigs in the mainland South and South-East Asian countries and its relationships between local wild boars. *Animal Science Journal* **79**, 417-34.

Tapio M., Ozerov M., Tapio I., Toro M., Marzanov N., Cinkulov M., Goncharenko G., Kiselyova T., Murawski M. & Kantanen J. (2010) Microsatellite-based genetic diversity and population structure of domestic sheep in northern Eurasia. *BMC Genetics* **11**, 76.

Tarasov P.E., Volkova V.S., Iii T.W., Guiot J., Andreev A.A., Bezusko L.G., Bezusko T.V., Bykova G.V., Dorofeyuk N.I., Kvavadze E.V., Osipova I.M., Panova N.K. & Sevastyanov D.V. (2000) Last glacial maximum biomes reconstructed from pollen and plant macrofossil data from northern Eurasia. *Journal of Biogeography* **27**, 609-20.

Tarasov P.E., Webb III T., Andreev A.A., Afanas'Eva N.B., Berezina N.A., Bezusko L.G., Blyakharchuk T.A., Bolikhovskaya N.S., Cheddadi R., Chernavskaya M.M., Chernova G.M., Dorofeyuk N.I., Dirksen V.G., Elina G.A., Filimonova L.V., Glebov F.Z., Guiot J., Gunova V.S., Harrison S.P., Jolly D., Khomutova V.I., Kvavadze E.V., Osipova I.M., Panova N.K., Prentice I.C., Saarse L., Sevastyanov D.V., Volkova V.S. & Zernitskaya V.P. (1998) Present-day and mid-Holocene biomes reconstructed from pollen and plant macrofossil data from the former Soviet Union and Mongolia. *Journal of Biogeography* **25**, 1029-53.

Tautz D. & Schloetterer C. (1994) Simple sequences. *Current Opinion in Genetics and Development* **4**, 832-7.

Telegin D.Y. (1986) Dereivka: a settlement and cemetery of Copper Age horse keepers on the Middle Dnieper, Oxford.

Tozaki T., Kakoi H., Mashima S., Hirota K., Hasegawa T., Ishida N., Miura N., Choi-Miura N.H. & Tomita M. (2001a) Population study and validation of paternity testing for Thoroughbred horses by 15 microsatellite loci. *Journal of Veterinary Medical Science* **63**, 1191-7.

Tozaki T., Mashima S., Hirota K., Miura N., Choi-Miura N.H. & Tomita M. (2001b) Characterization of equine microsatellites and microsatellite-linked repetitive elements (eMLREs) by efficient cloning and genotyping methods. *DNA Research* **8**, 33-45.

Troy C.S., MacHugh D.E., Bailey J.F., Magee D.A., Loftus R.T., Cunningham P., Chamberlain A.T., Sykes B.C. & Bradley D.G. (2001) Genetic evidence for Near-Eastern origins of European cattle. *Nature* **410**, 1088-91.

Trut L.N. (1999) Early canid domestication: the farm-fox experiment. *American Scientist* **87**, 160–9.

Uerpmann H.P. (1978) Metrical analysis of faunal remains from the Middle East. In: *Approaches to faunal analysis in the Middle East* (eds. by Meadow R.H. & Zeder M.A.), pp. 41-5. Peabody Museum, Cambridge, Massachusets.

Uerpmann H.P. (1990) Die Domestikation des Pferdes im Chalkolithikum West- und Mitteleuropas. *Madrider Mitteilungen* **31**, 110-53.

van Asperen E.N. (2010b) Implications of age variation and sexual dimorphism in modern equids for Middle Pleistocene equid taxonomy. *International Journal of Osteoarchaeology*, early view.

Van Haeringen H., Bowling A.T., Stott M.L., Lenstra J.A. & Zwaagstra K.A. (1994) A highly polymorphic horse microsatellite locus: VHL20. *Animal Genetics* **25**, 207.

Van Oosterhout C., Hutchinson W.F., Wills D.P.M. & Shipley P. (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**, 535-8.

Vilà C., Leonard J.A. & Beja-Pereira A. (2006) Genetic documentation of horse and donkey domestication. In: *Documenting Domestication: New Genetic and Archaeological Paradigms* (eds. by Zeder MA, Bradley DG, Emshwiller E & Smith BD), pp. 342-53. University of California Press, Berkeley, Los Angeles, London.

Vilà C., Leonard J.A., Götherström A., Marklund S., Sandberg K., Liden K., Wayne R.K. & Ellegren H. (2001) Widespread origins of domestic horse lineages. *Science* **291**, 474-7.

Von Koenigswald W. (2002) *Lebendige Eiszeit*. Arbeitsgemeinschaft für Biologisch-Ökologische Landeserforschung, Münster.

VonHoldt B.M., Pollinger J.P., Lohmueller K.E., Han E., Parker H.G., Quignon P., Degenhardt J.D., Boyko A.R., Earl D.A., Auton A., Reynolds A., Bryc K., Brisbin A., Knowles J.C., Mosher D.S., Spady T.C., Elkahloun A., Geffen E., Pilot M., Jedrzejewski W., Greco C., Randi E., Bannasch D., Wilton A., Shearman J., Musiani M., Cargill M., Jones P.G., Qian Z., Huang W., Ding Z.-L., Zhang Y.-p., Bustamante C.D., Ostrander E.A., Novembre J. & Wayne R.K. (2010) Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature* **464**, 898-902.

Wallner B., Piumi F., Brem G., Muller M. & Achmann R. (2004) Isolation of Y chromosome-specific microsatellites in the horse and cross-species amplification in the genus Equus. *Journal of Heredity* **95**, 158-64.

Warmuth V., Eriksson A., Bower M.A., Cañon J., Cothran G., Distl O., Glowatzki-Mullis M.-L., Hunt H., Luís C., do Mar Oom M., Yupanqui I.T., Ząbek T. & Manica A. (2011) European domestic horses originated in two Holocene refugia. *PLoS ONE* **6**, e18194.

Weber J.L. (1990) Informativeness of human (DC-DA)N.(DG-DT)N polymorphisms. *Genomics* **7**, 524-30.

Weinstock J., Willerslev E., Sher A., Tong W.F., Ho S.Y.W., Rubenstein D., Storer J., Burns J., Martin L., Bravi C., Prieto A., Froese D., Scott E., Lai X.L. & Cooper A. (2005) Evolution, systematics, and phylogeography of Pleistocene horses in the New World: A molecular perspective. *PLoS Biology* **3**, 1373-9.

Weir B.S. & Cockerham C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution* **38**, 1358-70.

Wendorf F. & Schild R.N. (1998) Nabta Playa and its role in northeastern African prehistory. *Journal of Anthropological Archaeology* **17**, 97-123

Whitlock M.C. (2011) G_{ST}' and *D* do not replace F_{ST}. *Molecular Ecology* **20**, 1083-91.

Wright S. (1943) Isolation by distance. *Genetics* **28**, 114-38.

Wright S. (1965) The interpretation of population-structure by F-statistics with special regard to systems of mating. *Evolution* **19**, 395-420.

Xuebin Q., Jianlin H., Lkhagva B., Chekarova I., Badamdorj D., Rege J.E.O. & Hanotte O. (2005) Genetic diversity and differentiation of Mongolian and Russian yak populations. *Journal of Animal Breeding and Genetics* **122**, 117-26.

Yamamoto S., Morita K., Koizumi I. & Maekawa K. (2004) Genetic differentiation of white-spotted charr (*Salvelinus leucomaenis*) Populations after habitat fragmentation: spatial–temporal changes in gene frequencies. *Conservation Genetics* **5**, 529-38.

Yang L., Tan S., Yu H., Zheng B., Qiao E., Dong Y., Zan R. & Xiao C. (2008) Gene admixture in ethnic populations in upper part of Silk Road revealed by mtDNA polymorphism. *Science in China Series C: Life Sciences* **51**, 435-44.

Yao Y.-G., Kong Q.-P., Wang C.-Y., Zhu C.-L. & Zhang Y.-P. (2004) Different Matrilineal Contributions to Genetic Structure of Ethnic Groups in the Silk Road Region in China. *Molecular Biology and Evolution* **21**, 2265-80.

Yao Y.G., Lu X.M., Luo H.R., Li W.H. & Zhang Y.P. (2000) Gene admixture in the Silk Road of China: evidence from mtDNA and melanocortin 1 receptor polymorphism. *Genes and Genetic Systems* **75**, 173-8.

Ząbek T., Nogaj A., Radko A., Nogaj J. & Slota E. (2005) Genetic variation of endangered Polish Bilgoraj and two common horse breeds at the microsatellite loci. *Journal of Applied Genetics* **46**, 299 - 305.

Zeder M.A. (2006) Archaeological approaches to documenting animal domestication. In: *Documenting Domestication: New Genetic and Archaeological Paradigms* (ed. by Zeder MA, Bradley DG, Emshwiller E & Smith BD). University of California Press, Berkeley and Los Angeles.

Zeder M.A. & Hesse B. (2000) The initial domestication of goats (*Capra hircus*) in the Zagros mountains 10,000 years ago. *Science* **287**, 2254-7.

Zhang Y., Sun D. & Yu Y. (2007) Genetic diversity and differentiation of Chinese domestic buffalo based on 30 microsatellite markers. *Animal Genetics* **38**, 569-75.

Zhao T.M. & Lee T.D. (1989) Gm and Km allotypes in 74 Chinese populations: a hypothesis of the origin of the Chinese nation. *Human Genetics* **83**, 101-10.