
Coat colours and mitochondrial lineages of ancient horses to document domestication

Dissertation zur Erlangung des akademischen Grades
„doctor rerum naturalium“
(Dr. rer. nat.)

eingereicht am
Fachbereich für Biologie, Chemie und Pharmazie
der Freien Universität Berlin

vorgelegt von

Dipl.-Biologe
Michael Cieslak

Berlin, Juni 2011

Die experimentellen Arbeiten zur vorliegenden Dissertation wurden in der Zeit von Juni 2007 bis April 2011 im Institut für Zoo- und Wildtierforschung (Evolutionsgenetik), Department für Nutzpflanzen- und Tierwissenschaften der Humboldt-Universität zu Berlin und im Deutschen Archäologischen Institut durchgeführt. Die Dissertation wurde am Fachbereich Biologie, Chemie, Pharmazie der Freien Universität Berlin eingereicht.

1. Gutachter: *Prof. Dr. Heribert Hofer (Leibniz-Institut für Zoo- und Wildtierforschung, Berlin)*

2. Gutachter: *Prof. Dr. Stephan Sigrist (Institut für Biologie, Freie Universität Berlin)*

Tag der Disputation: 06.02.2012

ACKNOWLEDGEMENTS

I. ACKNOWLEDGEMENTS

First of all, I would like to acknowledge and express my deep gratitude to Prof. Dr. Heribert Hofer for supervising me at the end of my doctoral studies and for serving as the chair of my doctoral examination and dissertation committee. Moreover, I would like to thank the further member of the dissertation committee, Prof. Dr. Stephan Sigrist.

In particular, I am truly indebted and thankful to PD Dr. Arne Ludwig who supervised me throughout the years of my doctoral studies. This dissertation would not have been possible without his advice, feedback, input, perspectives, and critical insights that helped focus my efforts.

Moreover, I am sincerely and heartily grateful to Dr. Melanie Pruvost. She introduced me to the work of ancient DNA research and her support, kindness, and helpful suggestions have been invaluable.

I would also like to thank the past and present members of the Institute for Zoo- and Wildlife Research (IZW) for their help and support and for the profitable time I could spend there. This applies particularly to the Research Group of Evolutionary Genetics under the supervision of Prof. Dr. Simone Sommer.

I would also like to take this opportunity to thank Dr. Monika Reissmann and the members of her group at the Department of Breeding Biology and Molecular Genetics (Humboldt University Berlin) and Prof. Dr. Norbert Benecke at the German Archaeological Institute (Berlin) in whose labs I could work.

I would also like to thank my parents, Gisela and Richard Cieslak. Their faith and confidence in me has shaped me to be the person I am today.

And finally, I would like to thank Dung Tieu Thi Phuong, Siti Mariam, Manuel Arnegger, Benjamin Ott and Christoph Rothmeier for the support and motivation that I received from them.

The research conducted in this study was supported by the DFG (Deutsche Forschungsgemeinschaft) requested by Arne Ludwig (LU 852/6-2; LU 852/7-3).

CONTENTS

II. CONTENTS

I. ACKNOWLEDGEMENTS.....	i
II. CONTENTS.....	ii
III. ABBREVIATIONS.....	iii
IV. SUMMARY.....	iv
V. ZUSAMMENFASSUNG.....	v
1. INTRODUCTION.....	1
1.1 Domestication of animals.....	1
1.1.1 <i>Origin of species domestication</i>	1
1.1.2 <i>Purpose of studying domesticates</i>	2
1.2 Genetic documenting of domestication.....	4
1.2.1 <i>Documenting of domestication and lineages</i>	5
1.2.1.1 <i>Question of molecular markers for optimal lineage classification</i>	5
1.2.1.2 <i>Signature of domestication on lineage diversity and geographic lineage distribution</i>	6
1.2.1.3 <i>Documenting place and time of domestication using lineages</i>	6
1.2.2 <i>Documenting of domestication and selected phenotypic character</i>	9
1.2.2.1 <i>Signature of domestication on variation of a selected phenotypic character</i>	9
1.2.2.2 <i>Documenting place and time of domestication using a selected phenotypic character</i>	10
1.3 Identification of a genetic background of a phenotypic trait.....	11
1.4 Coat colouration as phenotypic character under breeding.....	11
1.4.1 <i>Coat colour phenotypes</i>	11
1.4.2 <i>Coat-colour-associated genes</i>	12
1.5 Background and techniques for ancient DNA.....	15
2. AIMS OF THE STUDY.....	18
3. TAXONOMIC CLASSIFICATION.....	21
4. SUMMARY OF PAPERS.....	24
4.1 Summary of paper I.....	24
4.2 Summary of paper II.....	26
4.3 Summary of paper III.....	29
4.4 Summary of paper IV.....	31
5. DISCUSSION.....	33
5.1 Benefits of ancient DNA approaches to document horse domestication.....	33
5.2 Mitochondrial HVR1 part for the equine lineage classification.....	36
5.3 Lineage diversity and geographic lineage distribution of ancient wild horses.....	37
5.4 Lineage diversity and geographic lineage distribution under domestication.....	39
5.5 Coat colour variation of ancient wild horses.....	42
5.6 Coat colour variation under domestication and early breeding.....	43
5.7 Chronological appearance of investigated coat colour phenotypes.....	45
5.8 Conclusion about origin of horse domestication.....	46
5.9 Outlook.....	47
6. REFERENCES.....	49
7. APPENDIX: <i>included papers</i>	53

ABBREVIATIONS

III. ABBREVIATIONS

aDNA	Ancient DNA
AMOVA	Analyses of Molecular Variance
bp	Base pair(s)
BP/BC	Before present/Christ
D-loop	Displacement-Loop
DNA	Deoxyribonucleic acid
HVR1/2	Hyper variable region 1 or 2
MRCA	Most recent common ancestor
mtDNA	Mitochondrial DNA
PCR	Polymerase chain reaction
SNP	Single nucleotide polymorphism

SUMMARY

IV. SUMMARY

The domestication of the horse undoubtedly marks a major turning point in human history. In particular, the horse's role in pack, draft, and riding led to profound changes in human societies. The increased mobility provided by the horse enabled people to discover and migrate faster into previously unknown regions, and consequently increased the range of people's trade contacts. In fact, there is evidence to suggest that the domesticated horse was even partially responsible for the spread of Indo-European languages. And one should not forget the military advantage of the horse-technology, which was revolutionary. But nevertheless, very little is known about when, where, or how the earliest horse domestication happened. Until now, archaeologists have a lack of meaningful diagnostic anatomical and biometrical criteria to distinguish clearly between a domestic and a wild horse on the basis of ancient remains (e.g. bones, teeth). This is mainly due to the fact that early domestic horses most likely were not used for riding, but for milk and meat supply. However, only riding causes clear detectable traces on bones and teeth. Predominantly, it has been assumed that horse domestication began in the Eurasian steppe. And in fact, an undoubtedly important step was recently taken towards decoding the history of the domestic horse. A team of archaeologists has pinpointed horse domestication in Kazakhstan to around 3500 BC. Obviously, however, the real history of horse domestication was more complex. In contrast to a sole domestication event in Kazakhstan, mitochondrial studies of modern domestic horses indicate multiple origins. However, until now, any molecular research on the basis of modern DNA from domestic horses has failed to define even one single geographic place, let alone the date of horse domestication. Nevertheless there is the possibility of multiple domestication events at different places within the vast Eurasian steppe on the one hand, and on the other hand, even the Iberian Peninsula and China come into question as possible locations. On account of the fact that mitochondrial analyses on the basis of modern DNA had failed so far in shedding more light on the origin of horse domestication, in this dissertation the decision was made to use DNA from ancient horse remains. To determine the date and place of one or more domestication origins, two ancient DNA approaches were used. However, each of these two approaches has its own kind of molecular marker. The first approach was based on a neutral molecular marker (part of mitochondrial D-loop) and was therefore usable for a phylogenetic analysis comparing the geographic lineage distribution of the pre-domestic wild horses and the early domestic horses (total is 199 Eurasian and 8 Alaskan samples from the Late Pleistocene to the Middle Ages). In addition, this ancient DNA approach should be helpful in addressing the origin and history of maternal lineages in modern domestic horses and in order to find the cause of the huge mitochondrial DNA diversity within the modern domestic horse population. The second ancient DNA approach was an initial attempt to infer the date and site of domestication on the basis of a detected significant increase of coat colour variation. This could be possible, because of the assumption that wild horses had most likely a uniform coat colour, and that the coat colouration was in all probability an early goal of selective breeding. According to this, there was most likely a rapid increase of coat colour variation soon after domestication. Eight SNPs in six genes (*ASIP*, *MC1R*, *Kit*, *MATP*, *EDNRB*, *SILV*) were investigated to detect basic coat colours (bay, black and chestnut) on the one hand, and two kinds of dilution (silver and cream) and three spotted or painted colourations (overo, tobiano and sabino) on the other (total is 89 Eurasian samples from the Late Pleistocene to the Middle Ages). To improve the distinction between a wild and

SUMMARY

domestic horse, this dissertation includes the additional task of establishing a genetic test for dun coat colouration. This reasoning behind this is the evidence that dun was the original wild phenotype of the ancestor of the domestic horses. However, the molecular analysis revealed no association between the candidate gene, the *melanophilin*, and the dun phenotype. Consequently, it was not possible to distinguish between dun and non-dun horses. This was not as daunting as it first sounds, because the colour-gene-markers already mentioned were sufficient to display a shift in coat colour variation. In fact, given the detected rapid change of coat colouration, it could be concluded that horse domestication started in the Eurasian steppe region by at least ca. 3000 BC. In contrast to this clear outcome, mitochondrial lineages seem to be inappropriate for discovering a precise place of horse domestication within the Eurasian steppe. A phylogeographic pattern was constructible neither for the pre-domestic horses nor for the early domestic horse population within this area. This was most likely due to the fact that the wild horse population living there obviously had a huge number of maternal lineages and a panmictic structure through a high level of genetic exchange across this vast area. Nevertheless, mitochondrial data indicate that the Eurasian pre-domestic horse population can be subdivided into at least two genetically isolated populations. One population inhabited the Eurasian grassland steppe and the other one populated the Iberian Peninsula, which was a former glacial refuge. Furthermore, analyses of mitochondrial and coat colour data revealed that the Iberian Peninsula was not an important and independent horse domestication centre, at least not on a significant scale until the Iron Age, and provided instead strong evidence for introgression of local wild mares. Several previous studies claimed that the large number of maternal lineages of the modern domestic horse population is the consequence of a multiple origin scenario of domestication (each in different regions). However, the present study concludes that this claim is not necessarily true if one assumes that the majority of domesticated lineages originated from the panmictic wild horse population at the Eurasian steppe, which was characterized by a high rate of mitochondrial diversity. Nevertheless, the analysis of the mitochondrial data of early domestic horses may indicate a second wave of domestication during the Iron Age. In fact, the results of this study show that a huge number of maternal lineages were already present in the early domestic horse population, and 30% of these lineages were lost during 5500 years of horse breeding.

V. ZUSAMMENFASSUNG

Die Domestikation des Pferdes markiert ohne Zweifel einen bedeutenden Wendepunkt in der Geschichte der Menschheit. Insbesondere die Rolle des Pferdes im Transport, als Zug- und Reittier, führte zu tiefgreifenden Veränderungen in der Gesellschaft. Die gesteigerte Mobilität durch das Pferd ermöglichte dem Menschen schneller als je zuvor noch unbekannte Regionen zu entdecken und in diese vorzudringen, was dazu führte, dass sich die Reichweite seiner Handelskontakte ausdehnte. Es gibt tatsächlich Hinweise zu der Annahme, dass das domestizierte Pferd für die Verbreitung der Indo-Europäischen Sprachen mitverantwortlich war. Ebenso sollte erwähnt werden, dass die militärische Nutzung des Pferdes auch kriegerische Auseinandersetzungen revolutionierte. Nichtsdestotrotz weiß man sehr wenig darüber, wann, wie und wo die erstmalige Domestikation des Pferdes stattgefunden hat. Bisher mangelt es den Archäologen an aussagekräftigen anatomischen und biometrischen Kriterien um eindeutig zwischen einem domestizierten und einem wilden Pferd anhand von Knochenfunden unterscheiden zu können. Höchstwahrscheinlich sind frühdomestizierte Pferde nicht zum Reiten genutzt worden, sondern dienten der Milch- und Fleischversorgung. Jedoch hinterlässt ausschließlich das Reiten nachzuweisende Spuren an Knochen und Zähnen. Es wird überwiegend angenommen, dass die Domestikation des Pferdes in der eurasischen Steppe stattgefunden hat. Erst kürzlich konnte zweifelsfrei ein wichtiger Schritt in Richtung Entschlüsselung der Geschichte des Hauspferdes gemacht werden. Ein Team von Archäologen lokalisierte die Domestikation des Pferdes um 3500 BC in Kasachstan. Allerdings gibt es Anhaltspunkte, die auf einen komplexeren Ablauf der Pferdedomestikation hinweisen. Im Gegensatz zu einem einmaligen Domestikationsereignis in Kasachstan deuten Studien mit mitochondrialer DNA von modernen Hauspferden auf ein multiples Domestikationsereignis hin. Diese Studien konnten dennoch keine entscheidenden Kenntnisse über den Zeitpunkt und den oder die Orte der Domestikation liefern. Trotzdem werden unterschiedliche Domestikationsstätten innerhalb der eurasischen Steppe diskutiert. Desweiteren werden die iberische Halbinsel und China als unabhängige Domestikationsorte in Betracht gezogen. Dadurch, dass bisherige Analysen mit moderner mitochondrialer DNA keine konkrete Dokumentation der Pferdedomestikation zugelassen haben, basieren die Analysen dieser Dissertation insbesondere auf „ancient DNA“ aus Knochen und Zähnen fossiler bis frühmittelalterlicher Pferde (von Nordostsibirien bis einschließlich der iberischen Halbinsel aus der Zeit von Spätpleistozän bis frühes Mittelalter). Um den Ort und den Zeitpunkt der Domestikation auf molekularer Ebene nachzuweisen, wurden in dieser Arbeit zwei unterschiedliche „ancient DNA“ Lösungsansätze verfolgt. Der erste Lösungsansatz basiert auf einem neutralen Marker (Abschnitt des mitochondrialen D-Loop's) und dient folglich einer phylogenetischen Untersuchung, um die geographische Stammlinienverteilung von prä-domestizierten und früh-domestizierten Pferden zu vergleichen (insgesamt waren es 199 eurasische Pferdeproben und 8 Proben aus Alaska von Spätpleistozän bis Mittelalter). Außerdem sollten diese Analysen dazu dienen, den Ursprungsort von domestizierten mütterlichen Linien zu lokalisieren, als auch die enorme mitochondriale DNA Variabilität innerhalb der heutigen Hauspferdpopulation zu klären. Der zweite „ancient DNA“ Lösungsansatz war der erstmaliger Versuch den Zeitpunkt und den Ort der Domestikation durch einen rapiden Anstieg der Anzahl an unterschiedlichen Fellfarbvarianten festzustellen. Grundlage dieses Lösungsansatzes waren folgende Annahmen: Zum einen wird vermutet, dass die Fellfärbung der ancestralen Wildpferde einer einheitlichen Wildtypfellfärbung

ZUSAMMENFASSUNG

entsprach, und zum anderen, dass unterschiedliche Fellfärbungen höchstwahrscheinlich frühe Ziele der selektiven Züchtung unmittelbar nach dem Domestikationsereignis waren. Demnach ist ein rapider Anstieg der Fellfarbvariation in früh-domestizierten Pferden zu erwarten. Es wurden acht SNPs in sechs Genen (*ASIP*, *MC1R*, *Kit*, *MATP*, *EDNRB*, *SILV*) untersucht, um die Grundfarben (Brauner, Rappe, Fuchs), zwei Arten von Farbverdünnungseffekten (Silver und Cream) und drei gefleckt oder gemusterte Fellfärbungen (Overo, Tobiano und Sabino) zu analysieren (insgesamt waren es 89 eurasische Pferdeproben von Spätpleistozän bis Mittelalter). Um noch eindeutiger zwischen einem wilden und einem domestizierten Pferd unterscheiden zu können, beinhaltet diese Dissertation eine weitere Aufgabenstellung: Die Entwicklung eines genetischen Tests für die dun Fellfärbung. Hintergrund dafür ist, dass es sich bei diesem Phänotyp mit großer Wahrscheinlichkeit um die originale Fellfärbung der wilden Vorfahren heutiger domestizierter Pferde handelt. Jedoch konnte durch eine Kandidatengenanalyse keine Assoziation zwischen dem getesteten *melanophilin* Gen und dem dun Phänotyp festgestellt werden. Deshalb bestand keine Möglichkeit zwischen dun und nicht-dun Pferden zu unterscheiden. Dies hatte jedoch keine größeren Auswirkungen auf die Ergebnisse dieser Studie, da die genannten anderen Farbgenmarker ausreichten, um eine signifikante Zunahme der Fellfarbvarianten festzustellen. Tatsächlich konnte ein rapider Anstieg an unterschiedlichen Phänotypen nachgewiesen werden, sodass man annehmen kann, dass die Pferdedomestikation spätestens 3000 BC in der eurasischen Steppe stattgefunden haben muss. In Kontrast zu diesem eindeutigem Ergebnis erscheinen die mitochondrialen Stammlinien für eine präzise Ortsbestimmung der Domestikation innerhalb der weitflächigen eurasischen Steppe als ungeeignet. Die Rekonstruktion eines phylogeographischen Musters war weder für die prä-domestizierte Pferdepopulation noch für die früh-domestizierte Population innerhalb dieser Region möglich. Dies lag anscheinend daran, dass die dort lebende Wildpferdpopulation offensichtlich über eine große Anzahl an mütterlichen Stammlinien verfügte, eine panmiktische Populationsstruktur hatte und ein reger genetischer Austausch innerhalb dieses weitflächigen Gebietes stattfand. Nichtsdestotrotz deuten die mitochondrialen Daten darauf hin, dass die prä-domestizierten Wildpferde in zumindest zwei genetisch isolierte Populationen unterteilt werden können. Eine dieser Populationen bewohnte die eurasische Graslandsteppe, während die andere auf der iberischen Halbinsel vorkam, wobei es sich bei Letzterem um ein ehemaliges eiszeitliches Refugium handelt. Des Weiteren offenbarten die mitochondrialen Daten als auch die Analyse der Fellfarben, dass aller Wahrscheinlichkeit nach die iberische Halbinsel kein bedeutendes und unabhängiges Zentrum der Pferdedomestikation war, zumindest nicht im größeren Maßstab bis hin zur Eisenzeit. Die Ergebnisse deuten stattdessen eher auf vereinzelte Einkreuzungen von iberischen Wildstuten hin. Des Weiteren zeigt diese Studie anschaulich, dass die hohe Anzahl an unterschiedlichen mütterlichen Stammlinien in der Population moderner Hauspferde nicht unbedingt die Folge einer multiplen Domestikation in jeweils unterschiedlichen Regionen sein muss. Alternativ kann angenommen werden, dass ein Großteil der domestizierten Stammlinien von der panmiktischen Wildpferdpopulation der eurasischen Steppe abstammt, für die eine hohe mitochondriale Diversität charakteristisch war. Dennoch deutet die Analyse der mitochondrialen Daten der früh-domestizierten Pferde auf eine zweite Welle der Domestikation während der Eisenzeit hin. Allgemein zeigen die Ergebnisse dieser Studie, dass schon die Population früh-domestizierter Pferde eine hohe Anzahl an mütterlichen Stammlinien aufwies, wovon 30% dieser Linien während den 5500 Jahren Pferdezucht verlorengegangen sind.

1. INTRODUCTION

1.1 *Domestication of animals*

What is domestication? Although, desired traits such as lower aggressiveness and reduced fear of humans can be achieved with animals captured from the wild, many species that are regularly tamed, such as elephants in Asia, zebras in Africa or eagles used for hunting in certain cultures, never become domesticated. Indeed, domestication is a process much more complex and, unquestionably, much more time-consuming than taming because it includes genetic modifications of the species in question by intensive directed selection (artificial selection) for preferred phenotypic traits.

Therefore, domestication can be defined as: a hereditary reorganization of a wild ancestor into a domestic and cultivated form according to the interests of humans. This leads to morphological (e.g. coat colouration) and/or behavioural (e.g. tameness) changes and distinguishes the domestic species from its wild ancestor.

1.1.1 *Origin of species domestication*

More than 30 species of birds and mammals have been domesticated for the use in agriculture and food production (DAD-IS at <http://www.fao.org/dad-is/>). It is commonly assumed that the beginning of domestication is predominantly associated with early sedentary people during the Neolithic period. Such communities had abandoned their hunting-and-gathering culture and established an agricultural settlement. This transition to sedentary communities required technologies like irrigation, food storage, and therefore in particular domestication and breeding to cultivate species for extensive surplus production of food. These profound changes of the social cultural structure in Neolithic societies are thus called the Neolithic revolution (Childe, 1950).

However, the dog as the first domesticated animal is an obvious exception. There is evidence that the domestication of wolves was initiated late in the Mesolithic period when humans were nomadic hunter-gatherers (Savolainen et al., 2002). This was often explained by the fact that the dog was a useful guard and hunter for a hunter-gatherer culture, and therefore was not used preliminarily as a source of food. Up to the present, there are still controversial discussions about the place and time of its domestication, and additionally, about whether the dog was really a product of a primitive hunter-gatherer past. Recently, Pang et al., (2009) suggested that the dog may have originated among sedentary hunter-gatherers or early farmers in southern China less than 16.300 years ago. However, the scientific evidence of such an early East Asian primitive farming culture is unconvincing. The earliest proof of well established agricultural activity in East Asia (Yellow River and Inner Mongolia) dates from around 6000 BC (Barton et al., 2009; Fuller et al., 2008; Liu et al., 2009; Bettinger et al., 2010), but an East Asian domestication may have already begun 2000 years earlier (Lu et al., 2009).

INTRODUCTION

However, it is commonly assumed that the earliest cultural transition from hunting to farming occurred in Neolithic human societies of the “Fertile Crescent”, which includes parts of present-day Iran, Iraq, Syria and Turkey (9500–8500 BC; Harris, 1996; Cauvin, 2000; Diamond and Bellwood, 2003; Figure 1). These are known for cultivating crops and domesticating animals. Prominent examples of successfully domesticated animals from this region are shown in Figure 1 in APPENDIX, PAPER III; sheep (9000 BC), goat (9000–8000 BC), pig (8000 BC), cattle (8000 BC) and cat (7700 BC). Therefore, it is assumed that this region was the origin of the rapidly expanding Neolithic revolution. From the “Fertile Crescent”, domestic animals (Larson et al., 2007) and the knowledge about species domestication first spread to North Africa in the south, to Europe in the west, and also to Asia in the east (Diamond and Bellwood, 2003). As a result, many domestication events gradually began to occur outside of the “Fertile Crescent”, and domestication became a widespread and important custom of different human cultures. For example, it is assumed that the horse was not domesticated until ca. 3500 BC in the Eurasian steppe (Outram et al., 2009), which is remote from the locus of ruminant domestication in the “Fertile Crescent” (Figure 1), and that the domestic donkey originated in northeast Africa about 5000 years ago (Kimura et al., 2010). Furthermore, it should also be mentioned that there are still many controversial hypotheses about single and multiple domestication events of these animals (Jansen et al., 2002; Liu et al., 2006; Tapio et al., 2006; McGahern et al., 2006a; Naderi et al., 2007; Achilli et al., 2009; Driscoll et al., 2009; Lei et al., 2009; Outram et al., 2009; Bonfiglio et al., 2010; Kimura et al., 2010; Larson et al., 2010; Yindee et al., 2010). However, there is strong evidence that pigs and cattle were independently domesticated a second time in Asia (Kantanen et al., 2009; Larson et al., 2010).

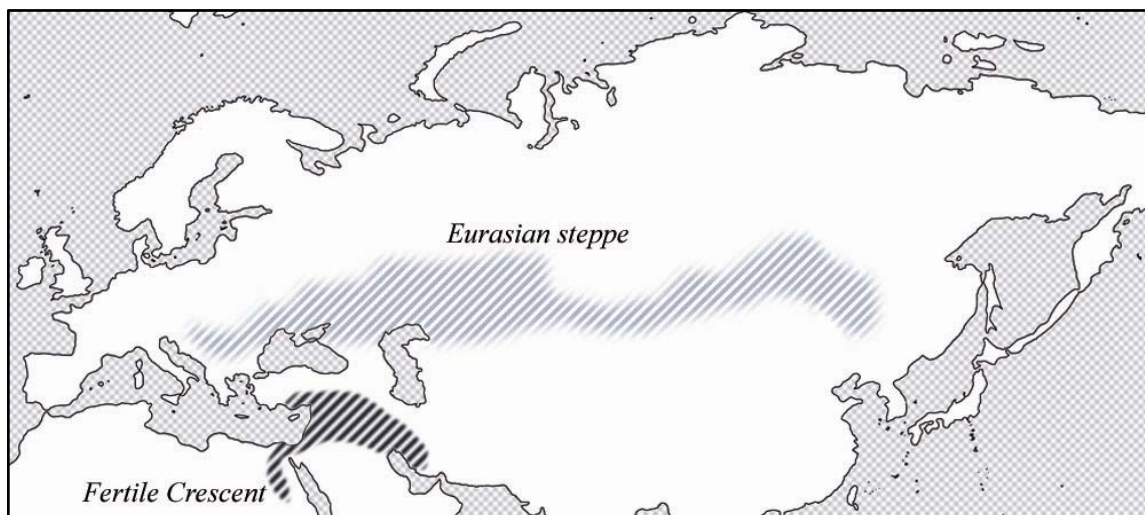


Figure 1. Map of Eurasia and the Middle East. The Fertile Crescent includes parts of present-day Iran, Iraq, Syria and Turkey. The Eurasian grassland steppe stretched from the western borders of the steppes of Hungary to the eastern border of the steppes of Mongolia.

1.1.2 Purpose of studying domesticates

Since Charles Darwin's epic *The Variation of Animals and Plants under Domestication*, first published in 1868, there has been strong interest in the study of domestication as a means of understanding the nature of selection. He described artificial selection as a further

INTRODUCTION

evolutionary mechanism and documented domesticated animals and plants as models for a result of a rapid and drastic evolution. For this reason, the domestic species is a valuable model for investigating the mode of inheritance in the context of evolutionary progress. Specialized areas of genetics, in particular population genetics, genetic diseases, and evolutionary genetics, profit from the domestic animal as a research model. The investigation of domestic species has some major advantages over studies of wild species: pedigrees are known (a major prerequisite for addressing the mode of inheritance), samples are normally not limited and many breeds are inbred, resulting in a fixation of gene variants responsible for breed specific phenotypic traits (e.g. height, meat quality, aggressiveness). This enormous phenotypic diversity of domesticates makes them ideal for studies of the genetic basis for phenotypic variation (Andersson, 2001). In particular, breed-specific disease susceptibilities make the domestic animal a promising model to identify causal gene/s of a disease trait.

Furthermore, the determination of the intraspecific genetic relationship among either a wild species (e.g. subdivision into diverse subspecies) or a domestic species (e.g. subdivision into genetic isolated breeds) can be important for establishing future breeding plans and conservation efforts. However, the reasons for species conservation are not only of cultural and scientific interest. In particular, there is an important economic aspect. For instance, wild species and livestock diversity of a domesticated animal are valuable genetic resources and can avoid an inbreeding depression of a commercial breed by introgression. However, livestock diversity is being increasingly lost, due partly to industrial livestock production. Precisely for this reason, it is important to be aware of the fact that breeds, which today are rare, may carry valuable traits such as tolerance of parasitic and infectious diseases, good maternal qualities, high fertility, longevity, and adaptability to harsh conditions and poor-quality feeds. Traits that may currently be of no commercial interest could likely become a significant asset in the future if environmental and economic conditions change. This would allow farmers to select another stock with the desired characteristic in order to introduce this new trait in his breed by crossbreeding. Therefore, whereas breed improvement programmes during the past century were mainly addressed to commercial breeds of cattle, sheep, goats, pigs and chickens, there is a recent increase of attention to locally adapted breeds (e.g. indigenous and primitive breeds) of these species because of the diversity of their special characteristics, leading to a mini-renaissance (LPP et al., 2010; Mathias et al., 2010). Given that domesticates played and will play a decisive role in the food but also fibres production for the increasing human world population, the importance of well thought out breed improvement programmes are indispensable.

Finally, the genetic diversity of ancient and modern livestock is also of interest for archaeological studies, because it provides representative information about ancient human cultures and movements through the old symbiotic companionship of domesticates and humans (e.g. Lui's et al., 2006; Priskin et al., 2010). Domestic animals were already transferred by humans from their point of origin to other regions even shortly after their domestication. This in turn enabled crossbreeding between different domestic stocks and introgression events of domestic animals with local wild species far from their point of origin. As a result, human migration events, trade contacts, customs, religious practices and military conflicts can leave their footprints in the genetic variation among domestic species. Consequently, both the ancient and the modern geographic distribution of domestic species are valuable information for archaeologists in reconstructing human history.

INTRODUCTION

1.2 *Genetic documenting of domestication*

Molecular signature of domestication. Documenting the temporal and geographic context of domestication is of interest to both biologists and archaeologists. If there is a lack of diagnostic anatomical and biometrical criteria to distinguish between a domestic and a wild species, molecular analyses are valued alternatives. This analysis is possible, because domestication and breeding leave an individual signature in the gene pool of an agricultural animal population. Following facts could be a reason for this signature. On the one hand the living conditions had extensively changed for domestic animals in captivity in comparison to their wild ancestors (e.g. genetic bottlenecks through the domestication or breed formation; inbreeding (increase of homozygosity within breeds); genetic drift (random loss or fixation of specific alleles); human-mediated choice of a mating, because of intentional selection of certain traits). On the other hand individual mutations were fixed within the domestic and the wild population after founder animals were genetic isolated from their ancestral wild population. As a result, one must assume that both populations were distinguishable from each other on the level of the molecular genetics. This in turn also means that it is possible to distinguish between a domestic and an ancestral wild animal. However, it is not necessary to analyze the complete genome of an animal for such an inspection; instead, it can be sufficient to investigate only a single or several specific molecular markers (Randi, 2008; Walsh, 2008). Such markers are selected nucleotide sequence regions of the DNA from the nuclear and mitochondrial genome, either from the coding or non-coding sequence regions. Although these sequence regions mainly demonstrate similarities among individual animals from the same species, each animal differs from others in some locations. Reasons for this variation are mutations, such as single nucleotide polymorphisms (SNPs), short tandem repeats (STRs), and deletion/insertion, as well as recombination, to name some examples. Usually, the highest level of genetic diversity is found in DNA sequence parts with little functional significance. These include, for example, non-coding regions, but also sites within coding regions; however, this applies only when the substitution at this position does not change the function of a molecule. In comparison, the lowest genetic diversity can be typically found for functional regions of molecules (Hartl and Clark, 2007). When considering the human genome, it is conspicuous here that there are fewer SNPs in coding (4%) than in non-coding regions (96%; Crawford et al., 2005).

In fact, molecular markers can be subdivided into neutral and selected markers. Neutral markers of a non-coding sequence region, which shows no effects on the fitness, and also has no link to a selected locus, are especially adapted for the establishment of a lineage (haplotype) classification (Luikart et al., 2003; Calonje et al., 2008). Consequently, such sequence regions are suitable for phylogenetic studies due to their relatively high mutation rate, but also for the establishment of a molecular clock (Zuckerandl and Pauling, 1962). However, selected markers are usually conserved regions of DNA (e.g. exons or sequence regions, which shows effects on the fitness) and therefore especially adapted to detect traits under selection, because molecular variations in these sequence regions are linked with changes of the trait variant from a phenotypic character (e.g. coat colour traits). Such a marker is either a coding sequence region whose allelic variations lead to corresponding phenotypic traits, or a non-coding sequence region, in which case a specific allele of this region has a genetic linkage to a specific allele of a coding region (and therefore also to a specific phenotypic trait).

Undoubtedly, the selection of optimal molecular markers is absolutely crucial in order to achieve the best possible result with regard to the research issue. In summary it can be said that well-selected molecular markers can serve not only to distinguish between a domestic and

INTRODUCTION

an ancestral wild animal, but also to calculate the relationship between animals, to speculate whether the population size has recently changed and to detect traits under selection. These properties of molecular markers can provide valuable information for the study and documentation of a domestication and breeding process.

1.2.1 *Documenting of domestication and lineages*

1.2.1.1 *Question of molecular markers for optimal lineage classification*

Considering the research topic of this dissertation, the selection of molecular markers for the establishment of a lineage classification should be based on their possible validity in connection with documenting a domestication event. The lineage classification is the basic prerequisite of phylogeographical-, mismatch- and molecular clock analyses (Avis, 2000), which can play a supportive role in discovering the place and date of a domestication event.

In fact, the molecular marker for the establishment of a lineage classification must be chosen carefully. Due to the fact that domestication took place only few thousand years ago, when determining an optimal nucleotide sequence, one should take into account that the average mutation rate of this molecular marker has to be relatively high. As noticed in section 1.2, the mutation rate of the mammalian DNA is not uniformly distributed along a DNA strand (Lynch et al., 2006; Yang, 2006). As a result, the phylogenetic resolution varies with analyzing different parts of the DNA (Araki and Tachida, 1997). For example, usually, exons show too little variability for the phylogenies of closely related species. However, intron regions have a large nucleotide variability and can therefore be used as a candidate sequence for phylogenies of closely related species (Igea et al., 2010). In particular neutral regions of DNA are especially adapted for the establishment of a lineage classification, because random has fixed their mutations (undirected mutation). However, whereas some neutral markers are usually better suited for the phylogenetic study of individual animals from the same species, other neutral markers are more adapted to identifying the relationship between individual animals from different species (Bannikova, 2004; Brito and Edwards, 2008). Usually, sequence regions of the mtDNA and the Y-chromosome have been preferred markers to establish clearly distinguishable haplotypic profiles for intra-specific phylogenetic analyses as shown in previous evolutionary studies (e.g. Edwards et al., 2000; Jansen et al. 2000; McGahern et al., 2006a; Ballard and Rand, 2005; Jobling and Tyler-Smith, 2003; Li et al., 2007; Pérez-Pardal et al., 2011). Due to the high copy number of mtDNA in cells, their comparatively fast molecular evolution and their strictly non-recombining maternal inheritance, mtDNA has become a valuable tool in phylogenetic studies (Brito and Edwards 2008; White et al., 2008). In particular, the mitochondrial control region, also known as Displacement-Loop (D-loop), is of interest for phylogenetic studies of closely related lineages, because of their high degree of variation among individual animals from one species. This part has a high number of rapidly evolving sites (Stoneking, 2000). It is usually organized into three different domains: a highly conserved central domain and two flanking parts with high intra-specific genetic variability, the so-called hypervariable regions 1 and 2 (HVR1 and HVR2).

The Y-chromosomal DNA is also a sex-linked marker and therefore a paternal counterpart to mtDNA. In particular the characteristic that the mitochondrial genome and portions of the Y chromosome do not recombine (this implies that their DNA sequence is inherited unchanged through the generations), making the interpretation of the data straightforward. Notably, the Y-chromosomal SNPs occur less frequently in comparison to

INTRODUCTION

the faster-mutating Y-chromosomal microsatellites. Therefore, they may have the potential to identify deeper lineages of Y-chromosomes of more recent evolutionary events (Bosch et al., 1999).

If the optimal haplotype is inferred, haplotype sequences can be ordered into the final genealogical tree or network on the basis of their evolutionary relationship. Several phylogenetic reconstruction methods have been developed, for example, the maximum parsimony (MP), the evolutionary distances (ED), the maximum likelihood (ML) and the bayesian inference (BI) method (Yang, 2006).

1.2.1.2 Signature of domestication on lineage diversity and geographic lineage distribution

Usually, it is assumed that the lineage diversity of an ancestral species population is larger than that of its domestic descendent population. This difference exists, because it is likely that all domestic animals were affected by a human-mediated artificial genetic bottleneck (Zeder et al., 2006). That is to say, a small number of reproductive animals were isolated from the ancestral wild population as founder animals for domestication. For this reason the lineage diversity is especially low in a newly established livestock. Furthermore, the genetic isolation of the domestic and the ancestral wild species population leads to individual lineages (spontaneous mutations over time) within both populations. Moreover, the number of lineages within a livestock can either be decreased or increased by breeding. Inbreeding (Van Dyke, 2003; e.g. through breed selection, genetic bottleneck by the formation of a new breeding stock) or the extinction of local livestock breeds can reduce the number of lineages of a domestic species population, whereas multiple domestication origins, introgressions by wild animals and an extensive breeding can increase the number of lineages.

Furthermore, especially the history of breeding and trade of a domestic species led to the present geographic distribution of its lineages. However, the geographic distribution of the lineages of a wild species metapopulation can be seen as the signature of ecological aspects. For example, the growth of a population usually leads to the migration of a part from this population. Probably, this geographic separation can cause their genetic isolation. Little to no gene-flow between these subpopulations led to population-specific lineage portfolios, which can even be characterized by private lineages.

1.2.1.3 Documenting place and time of domestication using lineages

The lineage diversity and the geographic lineage distribution are crucial information for population genetic analyses in order to reconstruct the status of the connectivity and the adaptive potential of an animal population. In fact, this information permits an insight into past events, which can help to document domestication (Calonje et al., 2008). However, firstly, it must be said that there is a recent progress in ancient DNA (aDNA; cf. section 1.5) extraction and amplification. This enormous progress promises a paradigm shift in common practices to document species domestication on the basis of the lineage diversity and the geographic lineage distribution in the future. Even today, the easier and more cost-effective use of aDNA analysis could allow including dead organism (even when they are older than several thousand years) into population genetic analysis even on a larger scale (Navascués et al., 2010). Especially, the possibility to investigate directly the lineage diversity and geographic lineage distribution of a species during the time of its domestication and of the early domestic species population will certainly lead to a decisive turning point in the genetic documentation of a domestication event regarding its precise temporal and spatial context, but also in describing its breeding history. However, until now, the inclusion of aDNA into

INTRODUCTION

analyses to document species domestication was to a very limited extent. On the one hand, samples from ancient times are very rare and often difficult to organize, and, on the other hand, the difficulty, complexity, heavy workload and the higher expenses of aDNA analysis in comparison to modern DNA analysis consequently limited its application (cf. section 1.5). Therefore, until now, the molecular documentation of animal domestication was mainly based on DNA of modern wild and domestic species as detailed below.

Documenting date of domestication using a molecular clock. It is possible to identify initial ancestral lineages of a domestic animal population on the basis of a coalescent simulation (Hudson, 1990; Schaffner et al., 2005). On the one hand this facilitates the discovery of the wild progenitor and on the other hand, there could, theoretically, be an option to deduce the date of a domestication event on the basis of a molecular clock (Zuckerandl and Pauling, 1962). The molecular clock hypothesis states that lineages (under the neutral theory of molecular evolution; Kimura, 1968, 1983) evolve at a rate that is relatively constant over time. This allows therefore the calculation of the divergence time of lineages on the basis of this “constant” mutation rate. Consequently, the calculated time of the splitting event from initial lineages of a domestic founder population should approximately represent the date of the domestication event.

However, the approach using a molecular clock for the identification of the temporal context of domestication has been increasingly revealed as a less appropriate method. Obviously, it is unsuitable for intraspecific calculations. It is in general very difficult to identify the exact mutation rate of a DNA part, because of the imprecise calibration of a molecular clock. In particular, the very short time span between domestication and the present requires a precise calibration. A molecular clock must first be calibrated against the date of the MRCA (most recent common ancestor) as a benchmark. However, this dating of the MRCA as a fossil record can be very imprecise. Consequently, molecular clocks are more suited to define the time point when two species diverged millions of years ago, but not for a meaningful estimation of such a “short” divergence time (a few thousand years) between a wild progenitor and its domestic descendent. A well-known example of a strongly debated molecular-clock approach to fixing the temporal context of domestication is the calculation of the divergence time between wolf and dog lineages by Vilà et al., (1997). Based upon their molecular clock study, it seemed that dogs diverged from the wolf lineage approximately 135.000 years ago. But this would mean that the dog split off more than 100.000 years earlier than the first archaeological evidence.

Documenting date of domestication using a mismatch analysis. Timing of population expansion can be a value indication for a domestication event. As already described in the previous paragraph, in all likelihood a domestic animal population started with few initial lineages because of a genetic bottleneck. Thereafter, spontaneous mutations followed by extensive breeding can lead to an increase of the lineage diversity. Usually, new mutations have been retained in a domestic animal population while this population is under rapid population growth. The time point of the beginning of the demographic expansion should be equivalent to the start of domestication.

Evidence for population expansion can be inferred, for example, from a star phylogeny (Slatkin and Hudson, 1991), but it can also be obtained from an examination of the distributions of pairwise differences among haplotypes (mismatch distribution; Rodgers and Harpending 1992; Harpending 1994). Several studies used the mismatch distribution to define the beginning of a demographic expansion event. This is possible, because a recent population expansion or a still expanding population shows a unimodal distribution of genetic distances

INTRODUCTION

between pairs of DNA lineages. In contrast, stationary or declining populations are signed by a ragged or multimodal curve, because of a random loss of lineages. The time of the beginning of the demographic expansion can be estimated from the mean of the unimodal curve. However, the expansion time is also estimated by a model, which is based on the molecular clock, and is also subject to the critical aspects as already described in the previous paragraph.

Documenting place of domestication using a phylogeographic structure. One option is to infer the geographic origin of a domestication event through the comparison of a geographic structured lineage distribution from the domestic animal and a phylogeographic structure of its wild ancestor. The discipline of phylogeography combines the evolutionary relationship of lineages with their geographic distribution (Avis, 2000). Presently, there are a number of computational methods to identify a geographic lineage structure of a species. The user can then evaluate, if the entire population can be divided into multiple subpopulations. Three such methods (among others) are the analysis of molecular variance (AMOVA; Excoffier et al., 1992), F-statistics (Weir and Cockerham 1984) and the Mantel-test (Mantel, 1967).

However, returning to the option of documenting the place of domestication using geographic distributions of lineages, it is generally assumed that the wild population having the most similar or even identical lineage/s compared to the domestic animal population is in all probability the ancestral population from which the domestic one descends. Thus, the geographical region of this ancestral wild population is then most likely the geographical origin of the domestic animal. Nevertheless, it is very difficult or almost impossible to reconstruct ancient original lineage distributions during the time of domestication on the basis of modern DNA alone. There are essentially two main approaches, mostly combined, to achieve this. The more promising approach is to restructure the original phylogeographic pattern on the basis of the extant descendants of the ancestral wild species population (e.g. Larson et al., 2005; Pilot et al., 2010), but there are some highly critical aspects. For instance, there was most likely a change of the original geographic distribution of lineages (of the wild species) in the period of time between domestication and today through particular ecological effects of climate changes. A further consideration is that many natural habitats of wild ancestors are severely endangered by human expansion and hunting, which led to displacements of wild populations, partial extinctions of regional populations or, in extreme cases, the total extinction of the wild ancestor. For example, a full extinction is known for the aurochs, the ancestor of modern domestic cattle, and the ancestral wild horse of the extant domestic horse. Such a lack of naturally distributed wild lineages forces scientists to reconstruct the original phylogeographic pattern by a second approach, which is then based only on the current domestic animal population. There is hope that the geographic lineages distribution of the domestic animal population might indicate this original structure (e.g. Jansen et al., 2002; Lei et al., 2006; Naden et al., 2007). However, there are several reasons arguing against the success of this approach. For example, an extensive human-mediated domestic species exchange probably led to the situation that such patterns became blurred over time. It is known of many domestic animals that they were already transferred by human migration, military conflict, and commercial trade from their point of origin to other regions soon after their domestication. Furthermore, due to the fact that domestic species were dispersed from their centre of origin (probably even from multiple domestication centres) to other regions, crossbreeding between geographically distant domestic animal stocks and introgression events of local wild animals far from the domestication origin were possible. It is also worth pointing out that most of our modern breeds are relatively young, as modern breed formation took place mostly during the past two to three centuries. Consequently,

INTRODUCTION

breeding was often an open process before breed formation took place. However, at least during the Neolithic time, it is likely that most domesticated stocks were small and that gene-flow among them was reduced because of the small population size of humans. As a result, one can assume that during this time there was probably a structure in the geographic lineage distribution. In particular the geographic occurrence of rare introgressed lineages (by local wild animals) within a domestic species population has the potential to indicate their correct original geographic distribution. However, proving such a hypothesis would require aDNA analyses.

As already described in the previous paragraph, the aDNA analysis of species remains (e.g. bones, teeth) have the huge potential to revolutionize the analytical methods listed here for documenting the date and place of a domestication event, but also for analyzing the following breeding history of a species. This potential of aDNA analyses is due to the fact that they offer a direct window to the original geographic lineage distribution of a population from the past. Although, the wild ancestor became almost or completely extinct, the aDNA study allows a reconstruction of an actual phylogeographic pattern during the time of domestication by using ancient samples from this time. Furthermore, aDNA research is a unique possibility to achieve a time depth, which therefore enables the gradual documentation of the history of a domestic lineage on the basis of temporal changes in its geographic distribution and its reproduction. Finally, although the use of the aDNA analysis can certainly contribute to an improved calibration of molecular clocks, it actually renders the rather unsuitable molecular clock method for intraspecific research unnecessary. This is the case when ancient samples are dated either through calibrated carbon dates or from archeological context.

1.2.2 Documenting of domestication and selected phenotypic character

1.2.2.1 Signature of domestication on variation of a selected phenotypic character

Phenotypic characters can either be of morphological (e.g. body size, coat colour) or of behavioural (e.g. tameness, aggression level) nature. They are the product of the expression of an organism's genes, but they also stand under influences of environmental factors. It is possible to permit an evaluation of the relative impact of genes and environment upon the variation of traits within and between populations by the heritability (Visscher et al., 2008).

Wild populations commonly possess a morphological unity and a species-specific phenotype. In contrast, domestic animals are usually characterized by large phenotypic variations not found in any of their wild ancestors (Figure 1 in APPENDIX, PAPER III). This, however, seems surprising, because this phenomenon leads at first glance to the contradiction of two common assumptions. The first assumption is that genetic diversity is manifested by differences in phenotypic characters (Frankham et al., 2007). The second one states that there is usually a lower genetic diversity in a domestic animal population in comparison to its closely related ancestral wild animal population (given, it is a non-endangered species of large population). This in turn is based on the hypothesis that a domestication event began with a small number of wild animals, which were used as founder lineages (historical bottleneck). Most likely, this phenomenon can be explained by following scenario: On the one hand there could be a larger genetic diversity even at conservative loci (e.g. coding regions) of the genome, because of a relaxed selection pressure. Domestic animals live in human keeping and thus have been exposed to reduced natural selection (Price and King, 1968). Therefore, physiological disruptions caused by mutations can be compensated. It is especially worth mentioning that a relaxed selection pressure allows that even deleterious alleles can be selected positively by breeders. In fact, domestic animals are

INTRODUCTION

often characterized by phenotypic traits, which are simultaneously linked to pathogenic effects of various degrees (e.g. leucistic white phenotype, the so-called leopard spotting found in some horse breeds is linked with night blindness; Bellone et al., 2008). Moreover, many domestic species are characterized by a reduction in brain size and also in the size of their sense organs (Kruska, 2005).

On the other hand, probably, there is a larger genetic diversity at certain loci of the genome from a domestic species, when one looks exclusively (including all breeds of a domestic species) at such loci that are responsible for phenotypic characters, which are especially interesting for breeders. This is possible due to the fact that different breeders preferred different traits of a character, which in turn led to the increase of variation of this character. It can be supposed that sooner or later after domestication, domestic species were dispersed from their centre of origins to other regions often with different climatic conditions. These new environments forced people to select for adaptations to the new habitats, eventually resulting in the fixation of favoured allelic variants. During this process, human preferences were not constant over time. For example, during the course of their domestication history, horses, initially domesticated primarily for providing meat and milk (Outram et al., 2009), had become animals for transport, warfare and finally sport riding. These shifts in phenotypic selection caused parallel changes in the genetic targets affected by selection. Consequently, if one now looks at not only one breed but instead at the unit of all breeds of a domestic species as one population, it can then be assumed that the allelic diversity of a common selected loci is higher in the domestic population than in the population of the wild ancestor. In contrast, particularly when considering a single modern pure-bred animal population, one can assume that there is a high homozygosity at their selected charismatic phenotypic traits. Moreover, the last century of industrialized farming has driven many old breeds close to extinction. This means at the same time that the genetic diversity of domestic species is again decreasing.

Finally, it should be mentioned that positively selected traits led to the increase of the rate of non-synonymous substitution in their causative locus. Therefore, it can be speculated on the basis of the ratio of non-synonymous to synonymous mutations, whether a locus was under neutral or positive selection (Tennessen, 2008).

1.2.2.2 Documenting place and time of domestication using a selected phenotypic character

Through improved methods of the aDNA analysis, there is a new potential approach to discovering the date and geographic site of a domestication event on the basis of a detected abrupt phenotypic change of a trait or even an rapid increase in the variation of a phenotypic character (you must remember that the population of a wild species commonly possess a phenotypic unity of a character; cf. 1.2.2.1). If phenotypic traits of a character are easily selectable and were probably an early goal of ancient breeders and concurrent with the domestication event, the date and place of an abrupt increase in the phenotypic variation of a character could simultaneously document the approximately time and site of a domestication event. Furthermore, it should be said that there is also the option of defining (independently of a domestication event) not only the temporal and geographical origin of a selected phenotypic trait, but also of tracking its geographical spread over time. Moreover, here it needs to be added that aDNA usually has a highly fragmented character; therefore, the molecular marker of this trait should preferable be detectable through a short sequence like a SNP.

INTRODUCTION

1.3 Identification of a genetic background of a phenotypic trait

The identification and understanding of the genetic background of a phenotypic trait are important in a number of disciplines, including animal breeding (e.g. meat quality), medicine (e.g. susceptibility to illness) and evolutionary biology (e.g. understanding of the evolutionary development of traits). Currently, different candidate gene approaches to defining the genetic architecture of a phenotypic trait are available. The question of which approach would be suitable depends mainly on the complexity (monogenic or polygenic) of a trait. In general, phenotypic traits are categorized in qualitative (*discrete*, e.g. black versus white) and quantitative (*continuous*, e.g. body size) traits. Whereas qualitative traits are mainly independent of environmental condition, phenotypic expressions of quantitative traits vary continuously in different environmental circumstances (Frankham, 2008). Furthermore, qualitative traits are usually determined simply on the basis of only one or probably a few involved genes. Usually, their inheritance follows Mendel's law. Whereas QTL (quantitative trait loci) regions are generally large and contain several putative causal genes/loci often found on different chromosomes. Therefore, different candidate gene approaches are essential for either qualitative or quantitative traits. Commonly, whole genome scan projects of diverse domestic animals constitute the basis of candidate gene approaches for qualitative as well as quantitative traits. The first step of a traditional approach for a qualitative trait is to examine published studies of similar phenotypes of interest for suggestions about a candidate gene. Then, comparative mapping data, linkage studies or association studies may give information about the genomic region. Whereas linkage analysis helps to identify variants in or near such genes that might be in linkage disequilibrium, association studies look for a statistical correlation between specific genetic variants (alleles) and corresponding phenotypes. Usually, however, the traditional candidate gene approach results in an information bottleneck in the case of the investigation of genetic complex quantitative traits. To examine the complex genetic architecture of such a trait, a variety of different methods are essential. For instance, the combination of QTL analyses (Gelderman, 1975) with gene expression profiling (DNA microarrays; Schena, 1999), so-called expression QTLs (eQTLs), can even describe epistatic effects of genes. The most remarkable progress in this field is the development of a novel web-resource-based candidate gene identification approach, the so-called digital candidate gene approach (DigiCGA; Zhu et al., 2010).

1.4 Coat colouration as phenotypic character under breeding

1.4.1 Coat colour phenotypes

In contrast to their wild ancestors, modern domesticated animals are characterized by a huge variability of coat colourations (Figure 1 in APPENDIX, PAPER III). As coat colour is an easily selectable phenotypic trait, colour traits, therefore, were most likely an important goal of selective breeding since ancient times. Generally, coat colour phenotypes can be categorized into two main subdivisions: (A) patterned phenotypes (e.g. dorsal-ventral pigmentation, white spotted, spotted, striped, leopard) or (B) non-patterned phenotypes (solid coloured ranging from heavily pigmented through brightened to white). Every patterned or non-patterned coat colour phenotype has its basic coat colour. This basic coat colouration is defined by the ratio of the two pigments eumelanin and pheomelanin, and range from black (dominance of eumelanin) through brown (mix of eumelanin and pheomelanin) to

INTRODUCTION

red/yellowish (dominance of pheomelanin) in mammals but also in many birds. The pigments are synthesized within pigment cells (melanocytes) and incorporated in skin (keratinocytes), hair and feathers. Furthermore, there are three different major modes for “white colouration”. Leucistic white results from a lack of pigment cells in skin and hair follicle, whereas albinistic white coat colour is an inability or impairment of melanogenesis (pigment synthesis) within pigment cells. A probable further mode of white colour was most likely based on a controlled pigment synthesis-stop of a functional pigment cell (so far, unproven on the molecular level). An example for such a mode could be the long-term coat colour adaptations and the colour change of winter- and summer-coat in mammals and also in birds. This requires a strong synchronization of the hair- or feather-follicle cycling and the pigment synthesis with respect to a seasonal coat colour change.

1.4.2 *Coat-colour-associated genes*

To date, more than 150 genes have been identified that influence pigmentation in various ways, and another 150 coat-colour-associated genetic loci have also been detected (Montoliu et al., July 1, 2010, <http://www.espcr.org/micemut>). Each colour associated locus is responsible for one or more pigmentation traits either alone or in combination with additional loci.

Despite the high number of discovered coat-colour-associated loci, their involvement within genetic pathways and also the interaction of these pathways in hierarchical successively and parallel operating systems are still only vaguely described. The molecular situation is further exacerbated on the phenotypic level. Frequently, coat-colour-associated genes were only categorized on the basis of their linked phenotypes like white-, white spotted-, and dilution-genes etc. However, different genes can be responsible for highly similar coat colour phenotypes in different individuals of a species or in different species. Therefore, any phenotypic classification of coat colouration blurs the underlying differences in the genetic basis of colour variants. And in addition, it obscures pleiotropic effects (they can affect performance-related traits or they even have pathogenic effects) which are frequently linked to coat-colour associated mutations.

To understand coat colour pigmentation and their pleiotropic effects at the level of complex interplays between many coat colour-associated genes, it is unavoidable to consider colouration as a complex process of many acting genes along a time axis starting from the time of the embryonic stage up to the stage of a pigmented coat. Therefore, it is helpful to subdivide the coat colouration into the following categories: i) pigment cell development; ii) pigment synthesis (includes melanosome formation); iii) pigment transport and transfer and iv) survival of pigment stem cells (Figure 2; Figure 2,3 in APPENDIX, PAPER III).

i) Pigment cell development: The pigment cell development (Silver et al., 2006) in turn can be subdivided into the following sub-items: a) the derivation and specification of precursor pigment cells (melanoblast) from the neural crest, b) the migration of melanoblast to allocated positions in the skin and c) the homing and differentiation of pigment cells. If they fail to do so, those areas of the skin will lack pigment cells resulting in white coat patches (a phenotype called leucism). All steps of the development depend on a cascade of molecular signal pathways (e.g. endothelin receptor type B – EDNRB/endothelin 3 – EDN3 and v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog – KIT/KIT ligand – KITLG) and transcription factors (e.g. microphthalmia-associated transcription factor – MITF, paired box 3 – PAX3 and SRY-(sex determining region Y)-box 10 – SOX10). Notably, mutations that

INTRODUCTION

disrupt the development of pigment cells can lead to leucistic white colour through their absence in skin and hair follicles (Figure 2).

Due to the fact that pigment cells derive from the embryonic neural crest (as noticed above), which is a multipotent tissue, and can also develop into different cell types including neurons, glia-cells, bone cells, smooth muscle and cartilage cells, mutations of coat-colour associated genes of in particular the subitems a) and b) can, consequently, also be linked (pleiotropic effects) with any physiological disruptions (Besmer et al., 1993; Metallinos et al., 1998; Santchi et al., 1998; Yang et al., 1998; Hallsson et al., 2000; Kawaguchi et al. 2001; McCallion and Chakravarti, 2001; Matsushima et al., 2002; Shibamura et al., 2002; Bellone et al., 2008; Stritzel et al., 2009; Minvielle et al. 2010).

ii) Pigment synthesis (including melanosome formation): The pigment synthesis (eumelanin and pheomelanin) takes place within the melanosomes (specialized organelles within pigment cells). But before the pigment production can start, melanosomes have to be formed (Nordlund et al., 2006). This formation can be displayed by a progress in maturity. Several cellular signals are necessary for the trafficking, targeting, docking and fusion of melanosomes (e. g. lysosomal trafficking regulator – LYST, melan-A – MLANA, adaptor-proteins, biogenesis of lysosomal organelles complex-1 subunit1 – BLOC1S1, RAB proteins, soluble NSF attachment protein receptors – SNAREs and regulators of G protein signalling – RGS proteins). Certain LYST mutations have been shown to result in diluted colour (Ogawa et al., 1997; Westbroek et al., 2007).

After the establishment of the fibrillar internal matrix in melanosomes by structural proteins and the organization of melanosomal enzymes, the pigment production can start (Nordlund et al., 2006). Such melanosome-specific components are the structure proteins (e.g. silver homolog – SILV, also known as PMEL17), membrane transport channels (oculocutaneous albinism II – OCA2, also known as P-protein, solute carrier family 45 member 2 – SLC45A2, also known as MATP, solute carrier family 36 member 1 – SLC36A1, solute carrier family 7 member 11 – SLC7A11 and solute carrier family 24 member 5 – SLC24A5) as well as melanogenesis-related enzymes (TYR, tyrosinase-related protein 1 – TYRP1 and dopachrome tautomerase – DCT). The synthesized melanins are deposited on the intraluminal fibrils within melanosomes, resulting in a progressively pigmented internal matrix.

Usually, the basic coat colouration (black, brown and reddish/yellowish) is determined by the synthesized eumelanin and pheomelanin ratio. This, in turn, is primarily controlled by the agouti signalling protein (ASIP) and melanocortin 1 receptor (MC1R) ligand-receptor system (Nordlund et al., 2006). Via TYR activation, MC1R signalling induces melanocytes to synthesize mainly eumelanin. MC1R's melanogenic activity is itself activated by post-translationally derived peptides (for instance, alpha melanocyte stimulating hormone - α MSH, beta melanocyte stimulating hormone - β MSH or adrenocorticotrophic hormone - ACTH) derived from proopiomelanocortin (POMC). However, such a MC1R's melanogenic activation can be reduced by the ASIP protein. Then, ASIP is binding to MC1R and leads in general to a brown coat colour, because both eumelanin and pheomelanin are produced. If the MC1R is inoperative through a loss-of-function mutation, this causes a switch to the synthesis of pheomelanin (reddish/yellowish coat colouration). In contrast, missense mutations in the code for the *ASIP* gene result in a protein that cannot bind to MC1R. Then, so long as one functional *MC1R* allele is present, MC1R signalling leads to the synthesis primarily of eumelanin (dark/black coat colour).

Normally, eumelanin is generated from L-tyrosine in a series of reactions in the presence of the key enzyme tyrosinase (TYR), whereas pheomelanin formation also starts

INTRODUCTION

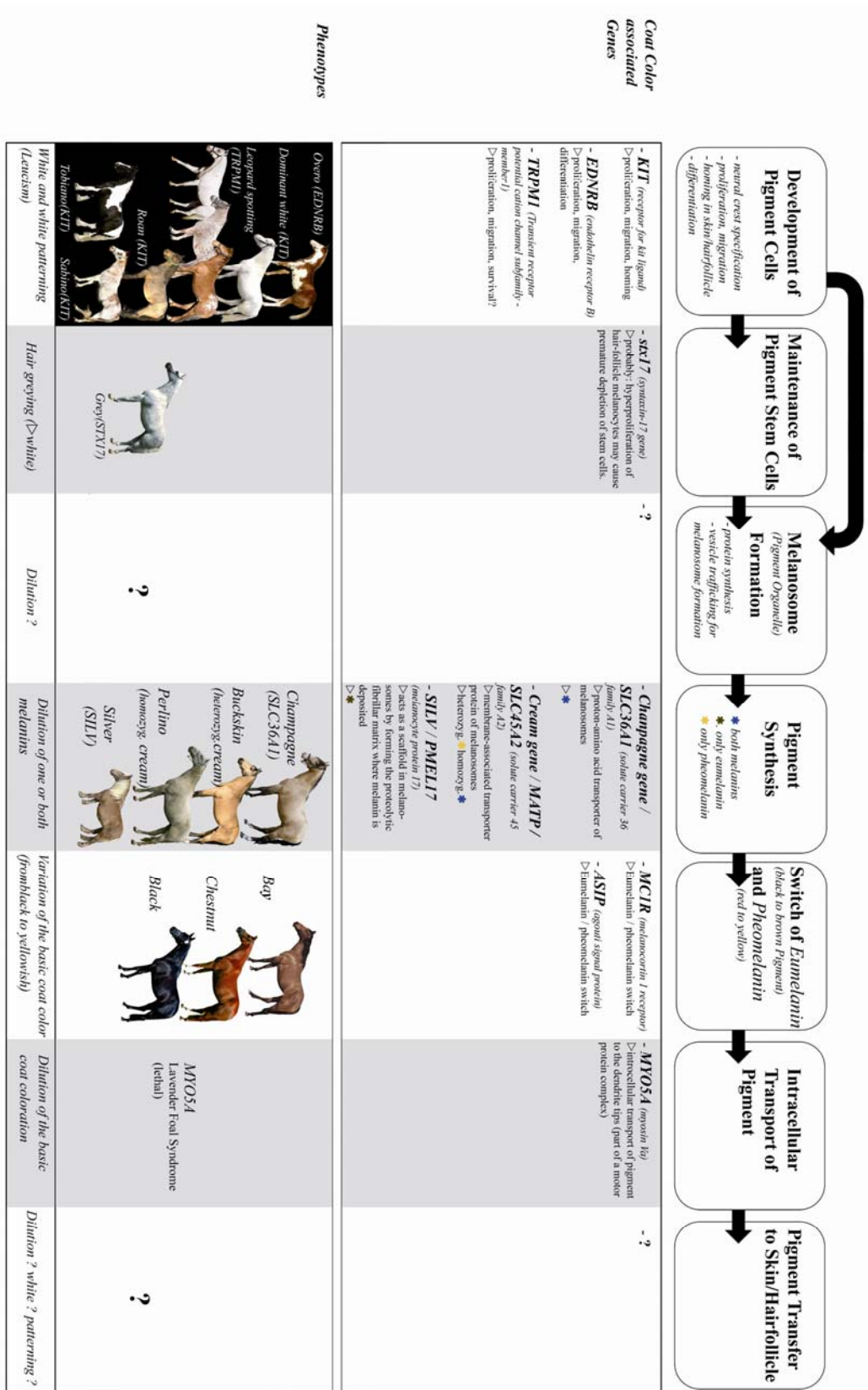


Figure 2. The upper boxes show stages of the pigmentation process, the lower boxes show the corresponding active coat-colour-associated genes and corresponding phenotypes. Gene names as follows *ASIP*, *agouti signalling protein*; *EDNRB*, *endothelin receptor type B*; *KIT*, *v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog*; *MATP* (*SLC45A2*), *solute carrier family 45 member 2*; *MC1R*, *melanocortin 1 receptor*; *MYO5A*, *myosin VA*; *PMEL17* (*SILV*), *silver homolog*; *SLC36A1*, *solute carrier family 36 member 1*; *STX17*, *syntaxin 17*; *TRPM1*, *transient receptor potential cation channel subfamily M member 1*.

INTRODUCTION

from L-tyrosine but proceeds only in the presence of cysteine and simultaneously in the absence of activated TYR (Sharma et al., 2002). Mutations in *TYR* result in unpigmented skin, fur and feathers (albinism), which are caused by a disruption of pigment synthesis. Dissimilar mutations within the mentioned genes of melanosome-specific components can affect the basic coat colouration in its intensity through hypo- (lightening) and hyper-pigmentation (darkening), or, if mutations have an effect on only one of both melanins, it can even modify the colour of the basic coat colouration. Commonly, however, MC1R is mainly responsible for the type of melanin produced. Also, both MC1R and ASIP can act locally as an extracellular modifier of colouration by influencing the distribution of eumelanin and pheomelanin in different body parts.

Pleiotropic effects are known in particular for mutations in genes, which are involved in trafficking, targeting, docking and fusion of melanosomes (e.g. Westbroek et al., 2007), but also in structure proteins (e.g. Clark et al., 2006). However, MC1R and ASIP have, more or less, limited pleiotropic effects.

iii) Pigment transport and transfer: For the pigmentation process to be completed, maturing melanosomes have to be transported from the perinuclear region to the peripheral actin-rich region (Nordlund et al., 2006). This transport takes place via kinesin along the microtubular network. The peripheral transport of melanosomes results in a final “peripheral melanosome positioning” and is achieved by a motor protein complex (melanophilin - MLPH, RAB27A and myosin VA - MYO5A). However, mutations in these genes can lead to diluted coat colouration (Philipp et al. 2005; Ishida et al. 2006; Vaez et al. 2008; Welle et al. 2009; Brooks et al. 2010; Bellone 2010).

Once melanosomes are positioned at the dendrite tip, they can be transferred to the surrounding tissues (skin, hair and feathers), resulting in tissue colouration. However, the genetic mechanisms of this melanosome-to-tissue transfer are not yet known.

iv) Survival of pigment stem cells: Faulty self-maintenance and proliferation mechanisms of follicular stem cells can be responsible for age-dependent greying (Nordlund et al., 2006). For example, when alleles reduce melanocyte stem cell survival, a growing number of hair follicle melanoblasts do not complete their development, resulting in a lack of pigment production and progressive graying. Usually, grey animals are born coloured. They gradually turn into a white phenotype only later during ageing, because pigment cells responsible for hair colouration will be recruited anew from the stem cell pool for each hair life cycle. Mutations in several genes like *MITF*, *KIT*, *NOTCH* (*notch 1*), *BCL2* (*B-cell CLL/lymphoma 2*) and *STX17* (*syntaxin 17*) are strongly associated with greying hair (Pielberg et al., 2008; Schouwey and Beermann, 2008; Blanpain and Sotiropoulou, 2010).

1.5 ***Background and techniques for ancient DNA***

Currently, molecular genetics allow DNA extraction, amplification, and sequencing from ancient tissues of a dead organism. However, unlike modern DNA analyses, aDNA studies have to deal with serious difficulties, because aDNA is highly degraded. Therefore, it was the advent of the Polymerase chain reaction (PCR) technique (Mullis and Faloona, 1987), which significantly contributed to the possibility of analyzing DNA from ancient remains (e.g. bones, teeth; Pääbo et al., 1989; Hofreiter et al., 2001). After the death of an organism, the DNA is rapidly degraded by enzymes. However, under rare specific circumstances, the DNA

INTRODUCTION

can become adsorbed by a mineral matrix that may escape for enzymatic and microbial degradation (Hofreiter, 2008). Generally, due to the low amount of aDNA in fossil remains, the aDNA is extremely difficult to extract. As a result, DNA-frugal methods are necessary to analyze aDNA. A further technical challenge is that fossil DNA is fragmented to a small average size, normally between 100 and 500 bp (base pairs).

The first aDNA study was done with mtDNA from tissue of the extinct zebra-like species, the *Equus quagga* (Higuchi et al., 1984). As was recently shown, it is even possible to sequence complete nuclear genes from Pleistocene specimens (several tens of thousands of years old; Römpler et al., 2006a). However, due to the much higher copy number of mtDNA contained in cells, as compared to nuclear DNA's two copies, mtDNA sequences are much easier to analyze from ancient specimens. Consequently, most aDNA studies to date have concentrated on mtDNA (Hofreiter, 2008), whereas molecular markers of nuclear DNA were only occasionally used in previous aDNA studies (Römpler et al., 2006a; Lalueza-Fox et al., 2007). But, in recent years, the combination of aDNA study and the developed technologies in high-throughput DNA sequencing gave rise to new scientific fields over time, such as paleogenomics and population genetics (Hofreiter, 2008).

Since aDNA studies began, the results have received much attention in the scientific community and have been published in the most prestigious scientific journals (Hofreiter, 2008). However, the analysis of aDNA also came under fire for two reasons especially. On the one hand, criticisms focused on the risk of contamination of the ancient samples with modern DNA (Hoelzel, 2005). There is a high risk of contamination due to the fact that a low amount of aDNA is preserved, while the PCR technique is very sensitive and can also, therefore, start their amplification from tiny amounts of contaminant DNA (O'Rourke et al., 2000). On the other hand, criticisms focused as well on the post-mortem damages (Hoelzel, 2005). These post-mortem aDNA modifications can lead to changed base-pairing through nucleotide substitutions (miscoding lesions). Such exchanges of single nucleotides can be categorized in the frequently occurring transitions (exchange of a purine for a purine, or a pyrimidine for a pyrimidine) and the less common transversions (exchange of a purine for a pyrimidine, or a pyrimidine for a purine). Notably, there are two types of transitions: (A/G)/(T/C) (type I) and (C/T)/(G/A) (type II) (Hansen et al., 2001), whereas the second type is more common (Binladen et al., 2006). Therefore, a dedicated laboratory exclusively devoted to aDNA and special criteria of authentication are obligatory for validating the aDNA results. Such efforts for making aDNA work more authentically must be adhered to during the entire procedure from archaeological field up to lab works. Therefore, it begins with the proper handling of excavated ancient tissues by archaeologists to avoid contaminations. Moreover, as mentioned, in particular an exclusively aDNA laboratory is the basic precondition for the successful analysis of aDNA. For instance, it is necessary to physically separate pre-PCR and post-PCR work (Herrmann and Hummel, 1994); it is advisable to equip the pre-PCR laboratory with a UV-filtered ventilation system and positive pressure airflow; the lab equipment should be properly bleached and exposed to UV light to destroy contaminant DNA; and the wearing of clean gloves, masks, boots and special lab coats should be obligatory. If the contaminated surface (prelaboratory contaminations) of ancient tissues is removed, the optimal DNA extraction and amplification methods should be carried out. Once aDNA is amplified, it can be treated the same as modern DNA samples, which means that no special laboratory or equipment are needed. However, it cannot directly determine whether amplified DNA is authentic, or if it is either contaminated or post-mortem damaged. Therefore, the reproducibility test particularly is a basic part of the aDNA study (Hofreiter et al., 2001a). The identity of an authentic DNA is most likely validated when the same DNA is extracted, amplified and sequenced from different bones of the same individual, ideally in

INTRODUCTION

different laboratories. Ancient DNA authentication also requires contamination controls during the laboratory procedures. For example, blank extraction can monitor possible contaminations, and both positive and negative controls should be standard for the PCR amplification of aDNA samples (Yang et al., 2003).

2. AIMS OF THE STUDY

Cave paintings and ancient bones at archaeological sites from Late Pleistocene leave no doubt that before domestication the horse was a prime food source for Stone-Age hunters. However, after domestication, the horse underwent a transition from a food source to a means of transportation in pack, draft, and horseback riding. In civilizations with horse-technology, the character of the culture changed dramatically (Kelekna, 2009). Horse-technology is often cited as the crucial event for the rapid spread of Indo-European languages (Anthony, 1986), bronze metallurgy and, in particular, specialized forms of warfare. Horses in combination with wheeled-technology (wagons and carts) were especially irreplaceable in agriculture and warfare. As a result, the political, military and economic impact of horse usage was profound and therefore played an indispensable part in empire expansions. For instance, Genghis Khan initiated the invasion of a territory extending from the modern-day regions of Hungary and Russia to Syria, Vietnam and Korea—the largest empire ever on earth—by means of horse-mounted warriors (Weatherford, 2004).

Despite the historical importance of the domestic horse, the date and place of its domestication continue to be hotly debated. The difficulty of establishing diagnostic anatomical and biometrical criteria to distinguish between a wild and a domestic horse has made it complicated for archaeologists to document horse domestication so far. Only recently, a team of archaeologists has found the earliest known traces to date of horse domestication in North Kazakhstan within the Eurasian grassland steppe, dating to the mid-fourth millennium BC (Outram et al., 2009). This point in time indicates that humans already had generations of experience in working with other domestic animals (e.g. dogs, sheep, goats, pigs, cattle and cats; Figure 1 in APPENDIX, PAPER III). However, until now, there are no unambiguous archaeological pictorial depictions of a horse rider or finds of an ancient horse chariot, which can be dated significantly earlier than 2000 BC. The earliest representations of a rider mounted on a horse were documented in Afghanistan 2100–1800 BC and in Ur III 2050–2040 BC (Owen, 1991), and the earliest horse chariot found by archaeologists dates from 2026 BC (Ural; Anthony and Vinogradov, 1995; Figure 1 in APPENDIX, PAPER I). To complicate matters further, especially molecular studies on the basis of mitochondrial data that do not emanate from a single domestication origin imply multiple domestication events (e.g. Vilà et al., 2001; Jansen et al., 2002; McGahern et al., 2006a). There were probably multiple origins within the very large Eurasian steppe, and even the Iberian Peninsula (e.g. Jansen et al., 2002; Royo et al., 2005) and China (e.g. Lei et al., 2009) come into question as possible locations. Until now, there were some factors, which prevented a successful application of genetic analyses to reconstructing the date and place of horse domestication on the basis of DNA lineages. In particular, the absence of an original phylogeographic pattern, which can be important for discovering the place of a domestication origin, led to this failure. This was due, on one side, to the fact that the ancestral wild horse is extinct and on the other, that until now, all attempts to reconstruct such a pattern on the basis of modern domestic lineages were without satisfactory success (e.g. Jansen et al., 2002; McGahern et al., 2006a, Lei et al., 2009). Several previous studies have demonstrated the difficulty of restricting mitochondrial lineages to a particular geographic area or a specific breed. Also, the extremely low nucleotide diversity on the Y-chromosome has made it unusable until recently (Wallner et al., 2004). Only a short time ago, the first description of two different Y-chromosomal haplotypes in Chinese indigenous breeds was given (Ling et al., 2010). However, in contrast to the Y-chromosome, an impressive amount of variability was discovered in modern horses for the

AIMS OF THE STUDY

mitochondrial control region. Consequently, this implies that a sex bias has been assumed for the founders of domestic horses. Probably, many females but only a few stallions were domesticated (Vilà et al., 2001; Jansen et al., 2002; Wallner et al., 2004). In particular this large number of maternal lineages of the modern domestic horse population led several previous studies to assume a multiple origin scenario of domestication.

Since, until now, any molecular research on the basis of DNA of modern horses has failed to document domestication. Therefore, in this dissertation the decision was made to use aDNA approaches. This study includes an ancient sample set of wild and early domestic horses covering a geographic range from China to Spain and spanning a time frame from the Late Pleistocene up to the Middle Ages. The first approach is a phylogenetic analysis comparing the geographic lineage distribution of the pre-domestic wild horses and the early domestic horses throughout Eurasia to infer a domestication origin (total is 199 Eurasian and 8 Alaskan horse samples). The lineage classification is established on the basis of the HVR1 sequence within the mitochondrial D-loop sequence, which has proven to be an important tool for understanding intraspecific relatedness (e.g. Guo et al., 2006; Driscoll et al., 2009; Bonfiglio et al., 2010; Kimura et al., 2010; Larson et al., 2010). The potential of aDNA analyses is due to the fact that they offer a direct window to the original geographic lineage distribution during the time of domestication. Furthermore, aDNA research is a unique possibility to achieve a time depth, which therefore enables the gradual documentation of the history of a domestic lineage on the basis of temporal changes in its geographic distribution.

However, the second aDNA approach should be an initial attempt to infer the date and site of domestication on the basis of a detected significant increase of coat colour variation in ancient horses (total is 89 Eurasian horse samples from the Late Pleistocene to the Middle Ages). This could be possible because of the following assumptions: (1) The coat colour of wild horses was most likely uniform; (2) the coat colour is an easily selectable phenotypic trait for breeders; and (3) the coat colour was in all probability an early goal of selective breeding. According to this, there was most likely a rapid increase of coat colour variation soon after domestication. Eight SNPs in six genes (*ASIP*, *MC1R*, *Kit*, *MATP*, *EDNRB*, *SILV*) should be investigated to detect basic coat colours (bay, black and chestnut) on the one hand, and two kinds of dilution (silver and cream) and three spotted or painted colourations (overo, tobiano and sabino) on the other. These molecular markers should not only be used to serve directly to distinguish between a wild and a domestic horse. Instead, in particular the increase of colour variation argues in favour of human-mediated selection.

Primary aim of this dissertation is:

Documenting the date and geographic location of horse domestication, and consequently, examination of the hypothesis of multiple origins of horse domestication.

In order to achieve this primary aim, the following secondary aims or objectives can be identified:

- *Reconstruction of a phylogeographic pattern of the pre-domestic and the early-domestic horse population.*

AIMS OF THE STUDY

- *Comparison of the maternal lineage diversity within the pre-domestic and early-domestic horse population, and the relevance of this diversity for the approximate number of actually domesticated maternal lineages.*
- *Documenting introgressions by wild mares into the domestic horse population.*
- *Reconstruction of the history of a single maternal lineage (e.g. geographic origin, domestication or introgression of this lineage, intensity of breeding of this lineage, human mediated geographic distribution of this lineage, etc.).*
- *Hypothesizing ancient human migration events (e.g. trade, warfare etc.) on the basis of geographic lineages' distribution.*
- *Documenting of coat colour variation in wild horses.*
- *Documenting of the chronological appearance of investigated coat colour phenotypes in domestic horses.*
- *Reconstruction of a coat colour trait geographic pattern within the early-domestic horse population.*

This dissertation also includes a candidate gene approach of *MLPH* (*melanophilin*) to establish a genetic test for dun colouration, because there is evidence that dun (Figure 1 in APPENDIX, PAPER IV) was the original wild-type colouration of the ancestral wild horse. Such a diagnostic test could improve the distinction between a wild and domestic horse. Therefore, this study possesses a further aim:

- *Examination of a candidate gene, the melanophilin, to establish a genetic diagnostic test for the equine dun phenotype.*

Finally, this dissertation includes extensive review about the coat colour in domestic animals with regard to their wild ancestors. The two-fold goal of this review is to deepen the understanding of genetic backgrounds of coat colour traits on the basis of the current state of knowledge, as well as to contribute to a better comprehension of why it is more crucial to categorize coat colour traits on their molecular mechanisms than on any phenotypic classification. In particular, the chronologically configured activities during the colouration process provide a better understanding of epistatic and pleiotropic effects of coat-colour-associated loci. In addition, this view of molecular mechanisms of colouration should also facilitate a deeper understanding of the underlying causes that have resulted in the observed increase of colour variation in domesticated animals compared to their wild ancestors. Such a view could be also helpful in finding more precise answers to the following questions: From when on did the phenotype variation in domestic animals increase? Was genetic drift or rather deliberate breeding the main factor driving this process? Are similar coat colour phenotypes of wild and domestic animals akin to molecularly controlled? Could there be a chronology order of when which coat colour trait appeared for the first time in the history of a domestic species?

3. TAXONOMIC CLASSIFICATION

Taxonomic classification of the extant members of the genus *Equus*

Class: *Mammalia*
Order: *Perissodactyla*
Family: *Equidae*
Genus: *Equus*

The term *Equus* is derived from the Latin word for “horse”. This genus includes many extinct equine forms and the extant equine species.

The extant stenoid horses:

- The African zebras: the plains zebra (*Equus quagga*), the mountain zebra (*Equus zebra*) and the grévy's zebra (*Equus grevyi*).
- The African wild ass (*Equus asinus*), which is the ancestor of the domestic donkey (*Equus asinus asinus*).
- The Asian asses: the Asian wild ass (*Equus hemionus*) and the Tibetan kiang (*Equus kiang*).

The extant caballoid horses:

- Subspecies: the modern domestic horse (*Equus ferus caballus*) is the descendant of an extinct species, the wild horse (*Equus ferus*).
- “Subspecies”: the Przewalski’s horse (mainly referred to as *Equus ferus przewalski*).

The dichotomous grouping of the genus *Equus* into two main groups: the stenoid (zebroid; zebras and asses) and the caballoid (“true”) horses is a historical classification (Boule 1899, Hopwood 1936, McGrew 1944; Forstén, 1992). Molecular studies however showed that zebras and asses form separate clades to varying degrees (Oakenfull and Clegg, 1998). This is also shown in Figure 3).

All extant equine members represent a group of closely related species, and molecular studies indicate that the MRCA of all the extant equines lived around 5.6 (3.9–7.8) million years ago. Later, around 4.8 (3.2–6.5) million years ago the lineages of the caballine horse (*Equus ferus*, *Equus ferus caballus* and *Equus ferus przewalski*), the donkey and the zebra diverged from one another (Weinstock et al., 2005).

TAXONOMIC CLASSIFICATION

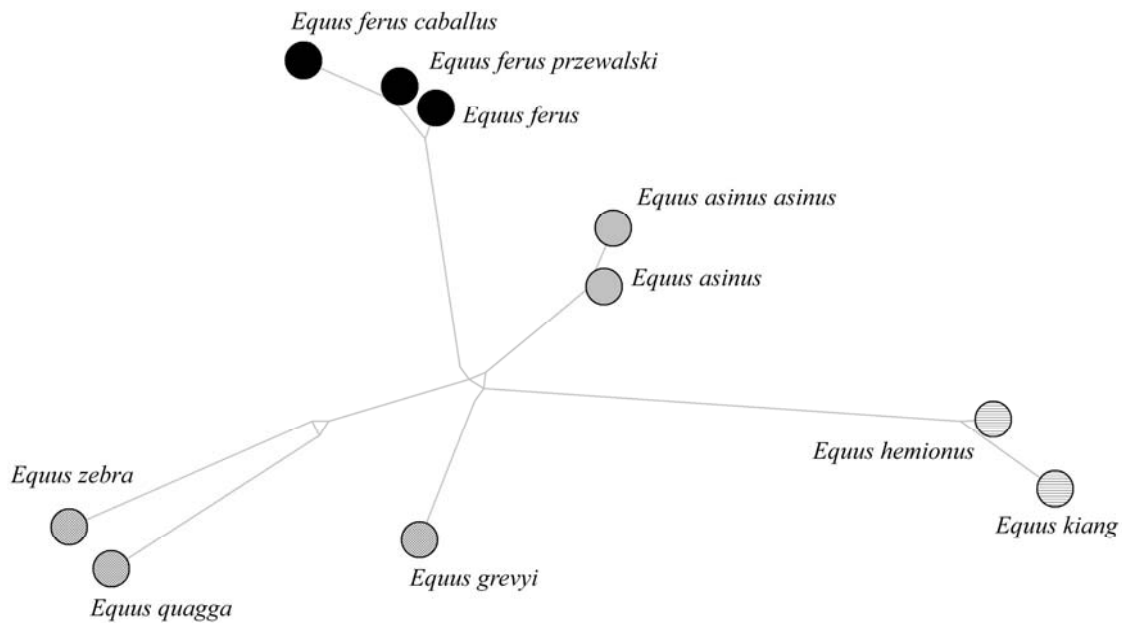


Figure 3. A median joining phylogenetic network constructed for *Equus ferus* (FJ204360), *Equus ferus przewalski* (AF014409), *Equus ferus caballus* (AY246220), *Equus asinus* (DQ368570), *Equus asinus asinus* (AY569544), *Equus hemionus* (AF220937), *Equus kiang* (AY569542), *Equus grevyi* (AF220929), *Equus quagga* (EU650599), *Equus zebra* (AY651947) mtDNA sequences using 242 bp (15,498-15,740; reference sequence: X79547) of the control region to show their relationships.

History of the extant caballoid horses. The modern domestic horse and the Przewalski's horse are the remaining extant caballoid horses. Whereas the domestic horse is definitely a descendant of the extinct wild horse, there is still some question as to what extent the Przewalski's horse (mainly referred to as *Equus ferus przewalski*) is related to the wild horse and to the domestic horse, and is currently not possible to classify. In any case, it is certain that they possess a common caballoid horse as ancestor.

Fossils indicate the first occurrence of caballoid horses in Eurasia around the early Middle Pleistocene (Uerpmann, 1987) and in West Europe in the middle Pleistocene (Uerpmann, 2005). Obviously, equines spread in diverse waves of migration from the North America over the Bering Strait to the Eurasia (during glacial periods when the sea level was low). However this assumption is based only on speculation, but in fact it is still unclear, whether the evolutionary origin of caballine horses was in America or in Eurasia. This is also due to the fact that the taxonomical situation for the Pleistocene caballoid horses seems to be very difficult, because there was a huge regional and temporal variety of the morphology (e.g. in body size and cranial morphology). This probably indicates a remarkably intraspecific plasticity during Pleistocene period. In all probability, such an extreme environment-driven adaptive morphology (ecomorphotype) would explain the overestimated taxonomic diversity of Pleistocene caballoid horses in the last century by paleozoologists. It is not surprising that nowadays a reclassification is taking place, which is based on the weight of molecular investigation methods. On the basis of mitochondrial data, Weinstock et al., (2005) documented that Late Pleistocene-Holocene caballines (*Equus ferus*) can be grouped into two major clades. One was a Holarctic clade distributed from Europe to North America and was most likely the ancestral stock for horse domestication in Eurasia. The other clade was obviously restricted to North America.

TAXONOMIC CLASSIFICATION

In agreement with many other Late Pleistocene-Holocene mammals, the equines from North America ultimately became extinct around 9000 BC, the reasons for which are still being debated. Most likely, crucial climatic changes during this time period led to steady reassimilations and, in turn, to a strong selection of horses. Similarly, many Eurasian horse populations disappeared or were displaced during the Late Pleistocene-Holocene. For instance, around the sixth millennium BC, wild horses had obviously vanished from the regions west of the Eurasian steppe. In spite of this, the Pontic-Caspian steppe continued to be inhabited or was repopulated by horses. Moreover, some populations even disappeared from Pleistocene refuges (Italian- and Balkan-Peninsula, Levant; Davis, 1980; Bedetti et al., 2001; Forsten and Dimitrijevic, 2004). Only the Iberian population persisted until at least the middle of the Holocene period (Uerpmann, 2005). Furthermore, there is an obvious lack of horse bones in Siberian archaeological records of three millennia until the third millennium BC. During the third millennium BC, a dramatic reappearance of what were most likely domesticated horses in funerary contexts can be observed in many regions of Europe, whereas the horse remains in Chinese human burials first became more common during the Bronze Age (2nd and 1st millennia BC; Flad et al., 2007).

Obviously, only small pockets of wild horse populations remained up to the last centuries. Whereas the European wild Tarpan population became extinct in 1879 in Ukraine, the Przewalski's horses survived in captivity. Meanwhile, captive Przewalski's horses have been reintroduced into the wild in parts of Mongolia and China. Today, the population of Przewalski's horses existing in zoos and in the wild descends from only 13 horses (eight females and five males; Volf et al., 1991). However, this genetic bottleneck most likely led to the limited number of only two mitochondrial lineages and one Y-chromosomal paternal lineage in the current population.

In recent decades, substantial disagreement continues between scientists about the degree of relationship between the modern domestic and the Przewalski's horse. The current state of investigation identifies a different number of chromosomes for the domestic horse ($2n=64$) than for the Przewalski's horse ($2n=66$), and recognizes that the Przewalski's horse and the domestic horse are the only equids that can cross-breed and produce fertile offspring. Furthermore, X-chromosomal, autosomal and mitochondrial data do not place Przewalski's horses in a clade separate from the domestic horse within a phylogenetic tree. This suggests a close relationship between domestic and Przewalski's horses (Lau et al., 2009). However, Y-chromosome data suggest that the split between Przewalski's horse and the domestic horse occurred 120,000–240,000 years ago (Wallner et al., 2003). Probably, this could indicate an interbreeding of only Przewalski's horse males and domestic horse females. Therefore, the true relationship between the two equines still remains open.

4. SUMMARY OF PAPERS

4.1 *Summary of paper I*

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

Background

Despite the high impact of horse-technology on the history of human societies, the date and place of horse domestication have remained cryptic. Genetic analyses on the basis of mtDNA in modern horse breeds and also limited ancient mtDNA analyses failed to answer these questions. Although Outram et al., (2009) recently provided evidence that horse domestication took place in North Kazakhstan about 3500 BC, mitochondrial studies of modern domestic horses suggest multiple origins in contrast to a single domestication event in North Kazakhstan, but without any clear indication of a date or geographic location (e.g. Vilà et al., 2001; Jansen et al., 2002; McGahern et al., 2006a). Therefore, in contrast to previous studies, this research addressed the question of the time and place of horse domestication by analyzing nuclear genes responsible for coat colouration of ancient horse remains. Presently, several genetic equine coat colour tests have been described (Rieder, 2009). In this study, a set of eight SNPs in six genes (*ASIP*, *MC1R*, *Kit*, *MATP*, *EDNRB*, *SILV*) were used for detecting basic coat colours (bay, black and chestnut), two kinds of dilution (silver and cream) and three spotted or painted colourations (overo, tobiano and sabino). Since coat colour is an easily selectable phenotypic trait, which was in all probability a goal of selective breeding already in ancient times, the detection of a significant increase of coat colour variation thus has great capability to elucidate the time and place of horse domestication. The investigated horse remains ranged from the Iberian Peninsula, Middle and Eastern Europe to China and Siberia and stemmed from different time points and cultural horizons (Late Pleistocene up to Medieval).

Aim

The aim of this study was to determine the date and place of the domestication origin of the horse. Notably, this aDNA approach to document domestication on the basis of a rapid increase of coat colour variation has not been utilized so far. Therefore, this study has also the general potential to propose a new approach to documenting domestication. In fact, this is the first analysis, which can shed more light upon coat colour variation in the wild horse population from Late Pleistocene up to the middle Holocene.

Results

Eight mutations in six genes that are responsible for coat colour variation were successfully typed from 89 ancient samples (out of 152 tested). The success rate of aDNA amplification depended on the climate (91% in Siberia, 52% in East Europe and 35% in Spain). Interestingly, there was no colour variation in the Siberian and European Pleistocene wild horses, suggesting that these horses were bay or bay-dun coloured. However, the Iberian Early Holocene wild horses showed both the black and the bay phenotype equally. Likewise,

SUMMARY OF PAPERS

analysis of East European lowland wild horses from the Neolithic and Chalcolithic periods reveals a similar result. It is assumed, that analyses of these Neolithic and Chalcolithic horses represent the millennia before the beginning of domestication. Unfortunately, this period is characterized by a lack of horse bones in Siberian archaeological records from 6000 up to 3000 BC, which prevents the inclusion of the Siberian region into a study for this time span. At the beginning of the 3rd millennium BC, there was a significant increase in the number of different coat colourations in both Siberia and East Europe. While the earliest chestnut allele was detected in Romania from the late 5th millennium BC, the first horse, however, with a chestnut phenotype was identified in Siberia from around the 3rd millennium BC. Later, this phenotype reached a proportion of almost 30% during the Bronze Age. Mutations responsible for coat colour spotting or dilution appeared later. The first sabino spotting phenotype appears during the 3rd millennium BC in Siberia and was later present in Armenia and Moldavia during the Middle Bronze Age. The second spotting phenotype, Tobiano, was only found in a single Eastern European sample during the second half of 2nd millennium BC. However, the cream (buckskin) and (black) silver dilution was first observed in Siberia from about a thousand years ago. Interestingly, there was no colour change in Iberian samples until Middle Ages.

Discussion

The study clearly proves that it is possible to document a transformation from a wild species to a domestic one on the basis of an increase of coat colour variation. Therefore, the analysis of nuclear SNPs responsible for coat colour phenotypes from aDNA is a very promising approach to pinpointing the date and place of a domestication event. Given the rapid increase of the number of different coat colour variants, it can be concluded that horse domestication started in the Eurasian steppe region around at least 3000 BC, because the observed changes in colouration are best explained by selective breeding. This supports the results of Outram et al., (2009), who provided evidence that horse domestication took place in Kazakhstan about 3500 BC. Furthermore, the fact that black wild horses did not inhabit Siberia or East and Central Europe during the Late Pleistocene, but did occur later in these regions during the Early Holocene period probably indicates a postglacial immigration or an adaptation to an increased forest cover as was recently shown for wolves. However, a human-mediated selection is less likely, because there was no coat colour phenotype other than bay and black during this time (also tested by a binomial test). Due to the fact that there was no observed colour change in Spanish samples until the Middle Ages, it can be concluded that there was no important and independent horse domestication centre, or at least not on a significant scale.

SUMMARY OF PAPERS

4.2 *Summary of paper II*

Cieslak, M. Pruvost, M. Benecke, N. Hofreiter, M. Morales, A. Reissmann, M. & Ludwig, A. (2010) Origin and history of mtDNA lineages in domestic horses. *PLoS ONE* 5(12): e15311.

Background

Despite the pivotal role horses have played in the history of human societies, the time and place of horse domestication is still unclear. Modern domestic horses represent a genetic paradox: while they have a very great number of maternal lineages (mtDNA; Vilà et al., 2001; Jansen et al., 2002; McGahern et al., 2006a), their paternal lineages are extremely homogeneous on the Y-chromosome (Wallner et al., 2004; Ling et al., 2010). Consequently, a sex bias has been assumed for the founders of domestic horses. It is likely that only a few stallions but many females were domesticated. Based on this huge number of domestic maternal lineages, several previous studies convincingly argued for multiple domestication origins. However, although over the last decade mtDNA has become a valuable tool to document the origin of a domestic species, until now, previous studies on horses have failed to do just that. This was mainly due to an absence of an original phylogeographic pattern, which is a basic prerequisite for hypothesizing a geographic origin of a domestication event. Due to the fact that the wild ancestor of the domestic horse became extinct, as well as the difficulty in restricting modern mitochondrial lineages to a particular geographic area or specific breed (e.g. Vilà et al., 2001; Jansen et al., 2002; McGahern et al., 2006a), the decision was made to use an aDNA approach to documenting horse domestication. Indeed, the analysis of the maternal lineage distribution of pre-domestic and early-domestic horses is very promising. Therefore, this study is based on analyzed HVR1 sequences (part of the D-loop) from 207 ancient horse remains (122 previously published ancient sequences and 85 samples newly analyzed in this study), which ranged from Alaska/northeast Siberia to the Iberian Peninsula and cover a temporal window from the Late Pleistocene period to the Middle Ages.

Aim

The primary goal of this study is to document the geographic mitochondrial lineage distribution of pre- and early-domestic horses in order to infer the date and the geographic origin(s) of horse domestication. In particular, the Iberian sample set of this study allows to hypothesizing about an Iberian domestication centre. Furthermore, the dataset also promises to shed new light on the question of whether the large mtDNA variability was also present in the pre- and early-domestic horse populations, and to theorize about origins and histories of mtDNA lineages in modern domestic horses. Finally, it is possible to hypothesize about cases of introgression of local wild mares and ancient crossbreeding between geographically isolated domestic livestock.

Results

85 ancient Eurasian samples (out of 157) were successfully analyzed. This ancient dataset was supplemented by 122 aDNA sequences archived in GenBank (in total n=207). Due to the fact that this study revealed a fourth mutational hotspot (three of them were identified by Jansen et al., 2002), which was excluded from phylogenetic analyses, it was forced to define new haplotypes and provide them with a new nomenclature. The huge number of 87 different ancient haplotypes (out of 207) was observed within the sample set from Late Pleistocene up to the Middle Ages. The ancient haplotypes were grouped into 19 haplogroups from A to K and from X1 to X7 based on the structure of the median joining network. However, the

SUMMARY OF PAPERS

modern sample set of 1754 samples (archived in GenBank) was grouped into 274 different haplotypes. 56 of the 87 ancient haplotypes were also observed in current domestic horse population, but only 39 haplotypes were confirmed to have survived in modern breeds. Thus, at least 30% of haplotypes of early domestic horses have become extinct during the last 5500 years. Today, X2 and D3 are the most common ancient haplotypes in modern horses. Moreover, the ancient sample set revealed no clear phylogeographic pattern for pre- or early-domestic horse populations, with the exception of the Iberian wild horse population. The Iberian samples from ca. 5000–1000 BC had the lowest values of both nucleotide and haplotype diversity. Moreover, the average nucleotide diversity of wild horses (Late Pleistocene and pre-domestic samples) was about 25% lower than the average for domestic horses (when the Iberian samples are excluded). Interestingly, the geographic haplotype distribution of Late Pleistocene horses points to a panmictic horse population with a huge haplotype diversity, ranging from northeast Siberia to the Pyrenees. Furthermore, the Early Holocene and the Chalcolithic samples indicate, separate sub-populations of the Eurasian steppe region and the Iberian Peninsula. Furthermore, the mismatch distribution curves of the Iberian pre-domestic and the Chalcolithic/Bronze Age population, as well as the European samples of pre-domestic times were extremely ragged. Meanwhile, the mismatch distributions of the west and south Siberian and European sample sets from early-domestic times show clear unimodal curves. Interestingly, the curves of the sample sets from west and south Siberia were marked by a weak bimodal peak and whereas the East Asian sample set from early domestic times has a ragged, bi- or trimodal character.

Discussion

It was not possible to discover the date and place of a domestication origin on the basis of the ancient mitochondrial data. This was mainly due to the fact that the ancient wild horse data indicated a panmictic horse population within the Eurasian steppe. A huge mitochondrial diversity and probably a high level of genetic exchange across this vast area led to the loss of a phylogeographic pattern. Although this region in particular is assumed to have played a crucial role in horse domestication (Outram et al., 2009), the absence of such a phylogeographic pattern makes it impossible to speculate about precise places within this immense region on the basis of mitochondrial lineages. Nonetheless, mitochondrial data indicate that the Eurasian pre-domestic horse population can be subdivided into at least two genetically isolated populations: a panmictic population in the Eurasian grassland steppe and a population on the Iberian Peninsula. Probably, there was a third more or less isolated Chinese population, but the dataset of this study had a bias of pre-domestic samples from East Asia and therefore prevents conclusions about the role of China with regard to its contribution of private lineages to the domestic horse population. However, there might be evidence that the two domestic lineages had their roots in East Asia, and, in addition, that both are also common in East Asian primitive breeds today. There was as well, though, a lack of a phylogeographic pattern for the early-domestic horse population, which would not be surprising if the majority of founder lineages from the early domestic horse stock were captured from the panmictic horse population from the Eurasian steppe. Nevertheless, the data of early-domestic horses speak for multiple introgressions or, alternatively, for a second wave of domestication during the Iron Age, because many new haplotypes were discovered in the gene pool of domestic horses from this time period. The only region with private wild lineages was found on the Iberian Peninsula. Obviously, the Pyrenees effectively cut off the Iberian Peninsula from the rest of Eurasia and turned it into a glacial refuge. Pre-domestic Iberian horses up to the Chalcolithic period were shown by our data to be genetically isolated. Currently, these private Iberian lineages are occasionally present for the most part in Iberian,

SUMMARY OF PAPERS

West European and South American breeds, as well as in Arabian horses. The occasional occurrence of these ancient Iberian lineages speaks against an independent domestication origin in Iberia, at least on a significant scale, and provides instead strong evidence for introgression of local wild mares. Finally, considering the data of this study from the Bronze Age horses, it appears that domesticated horses spread throughout Eurasia soon after their initial domestication. Most likely, horse-riding was the catalyst for such a rapid dispersion. The high mobility of horses combined with a huge mitochondrial diversity in early-domestic horses, intensive horse breeding, and hybridizations and introgressions of wild mares have left their footprints on the genetic pedigree of the domestic horses. After 5500 years of horse breeding, this results in a strong mixture of maternal lineages currently throughout the world.

4.3 *Summary of paper III*

Cieslak, M. Reissmann, M. Hofreiter, M. & Ludwig, A. (2011) Review: Colours of domestication. *Biological Reviews* 86: 885-899.

Background

The large coat colour phenotypic variation of domestic species is one of the most obvious differences between domestic animals and their wild ancestors. Whereas the coat colouration of wild animals is usually uniform and species-specific, domestic species come in various colours and heterogeneous patterns (Figure 1 in APPENDIX; PAPER III). It is quite clear that this huge allelic variability of coat-colour-associated genes is a mark of domestication and breeding. However, presently, it is still largely unknown how colour phenotypes were distributed in early domestic animals and when they occurred for the first time.

During the last decades, coat colouration has been investigated in numerous studies, particularly among domestic species, because breeders are highly interested in the development of coat colour genetic tests (e.g. Schmutz and Berryere, 2007; Rieder, 2009). These tests enable breeders and farmers to identify carriers of favoured and unfavoured coat colour alleles, which can be important for the breeding choice in order to preserve offspring with the desired coat colour phenotype.

Although to date more than 300 genetic loci and more than 150 identified coat-colour-associated genes, which influence pigmentation in various ways, have been discovered (Montoliu et al., July 1, 2010 <http://www.espcr.org/micemut>), the genetic pathways influencing coat colouration are still only poorly described. Frequently, coat-colour-associated genes were categorized by any phenotypic classification of coat colouration, but this blurs underlying differences in the genetic basis of colour variants. This is due, on the one hand, to the fact that different genes can be responsible for highly similar coat colour phenotypes in different individuals of a species or in different species. On the other hand, it is also possible that different mutations in a single gene result in greatly varying phenotypes.

Aim

Certainly, one of the objectives of this review is to deepen the understanding of genetic backgrounds of coat colour traits on the basis of the current state of knowledge. However, the primary point of this review is to show that there is the need to abandon the phenotypic classification of coat-colour-associated genes. Instead, this review insists on the categorization of coat-colour-associated genes on the basis of their molecular mechanisms of colouration, with a special emphasis on when and where the different coat-colour-associated genes act. In particular, the chronologically configured activities during the colouration process provide a better understanding of epistatic and pleiotropic effects of coat-colour-associated loci. Moreover, this view of molecular mechanisms of colouration should facilitate a deeper understanding of the underlying causes that have resulted in the observed increase of colour variation in domesticated animals compared to their wild ancestors. Such a view could also be helpful in finding more precise answers to questions such as when did their phenotype variation increase, was genetic drift or rather deliberate breeding the main factor driving this process, and are similar coat colour phenotypes of wild and domestic animals akin to molecularly controlled?

SUMMARY OF PAPERS

Results and discussion

In fact, the categorization of coat-colour-associated genes on the basis of their molecular mechanisms with a special emphasis on when and where they act during the colouration process is vital to hypothesize about the history of coat colour traits in domestic animals. The huge allelic variability of coat-colour-associated genes in domestic animals resulted from relaxed natural selection within human caretaking. Therefore, artificial selection also accepts deleterious coat colour traits, which would, in nature, be under purifying selection either because of their linked pathogenic effects through pleiotropy or through their unfavourable colouration. Consequently, such traits can become fixed in a domestic livestock.

In particular, the fact that many coat colour traits based on a SNP supports the assumption that their genetic fixation in a domestic livestock was possible within only a few generations of intentional breeding. Therefore, one can assume that several coat colour traits were early-goals of selective breeding soon after the domestication event. However, interestingly, it cannot be excluded that the first appearance of some coat colour traits was probably a random by-product of animal selection at the beginning of domestication. This would be possible when a coat colour trait is molecularly linked (through pleiotropy) to traits positively selected by humans. For instance, taming experiments on foxes, minks and rats have shown that selection for the trait of tameness probably led to an increase of colour variety (e.g. Trut, 1999; Trut et al., 2001; Prasolova et al., 2009; Roulin and Ducrest, 2011). However, until now, there is a lack of clear evidence for this molecular link.

Furthermore, it is probable that the less pleiotropic character of *ASIP* and *MC1R* features these two coat-colour-associated genes as predestined to be involved in natural coat colour adaptation processes. This could explain why, until now, an allelic variation was found only at these genes in wild animals. However, in this context it must be said that, in total, only a few investigations were done on wild animals, but then focused exclusively on *ASIP* and *MC1R*.

Moreover, the fact that many coat colour traits in domestic animals are linked to deleterious pleiotropic effects (e.g. deafness of white-patterned dogs and night blindness of homozygous leopard-spotted horses) suggest that several similar coat colour phenotypes of wild and domestic animals are not identical molecular controlled. In particular one can assume that white coat colour phenotypes of wild and domestic animals are not often controlled by the same genes. If one considers, for instance, leucistic or albinistic alleles, it is conspicuous here that they can lead to a lack of pigment in coat and skin. But a lack of the pigment layer in skin is very unlikely for wild species, because it has to shield DNA of surface cells and tissues from harmful UVR.

SUMMARY OF PAPERS

4.4 *Summary of paper IV*

Cieslak, M. Reissmann, K. & Ludwig, A. (Revised: 2010) Examination of *melanophilin* as a candidate gene for the equine dun phenotype. *BMC Research Notes*.

Background

During the last decade, coat colouration in horses has been extensively investigated in numerous studies; therefore, several genes have been identified, which are involved in specific coat colour traits (Rieder, 2009). However, the investigation of the genetic background from the dun phenotype has been unsuccessful so far. The dun trait occurs in certain horse breeds. Breeding has shown that the dun colouration is inherited as a dominant trait, and Dn^+ has been suggested as the dominant allele, and Dn^{nd} has been suggested as the recessive allele (Sponenberg, 2009). This phenotype is characterized by the dilution of the basic coat colour, whereas the extent of dilution at the lower legs and head is less obvious. In addition, dun horses possess a dorsal stripe, while the presence of other patterns, especially darker stripes on the lower legs and the shoulders can vary (Figure 1 in APPENDIX, PAPER IV). The dun coat colour is thought to be a primitive trait in the horse. In particular the fact that the Przewalski's horse, the African ass and the Asian asses are characterized by a dun colour indicates that this phenotype was most likely the original phenotype of the extinct wild ancestor of domestic horses. In our study we wanted to find out whether there can be observed an association between the SNPs within the 16 exons of the candidate gene *melanophilin* (*MLPH*) and the dun phenotype. *MLPH* forms together with *MYO5A* and *RAB27A*, a motor protein complex, which plays a decisive role in the intracellular melanin transport in pigment cells (Van Gele et al., 2009). Today, it is known from mutative disruptions of this protein complex that it can lead to dilution of the basic coat colour (e.g. Philipp et al., 2005; Drögemüller et al., 2007; Vaez et al., 2008; Welle et al., 2009).

Aim

The aim of this study is to establish a genetic test for the dun phenotype. Such diagnostic tests are very important for breeders, because of breeding selection to preserve offspring with the desired coat colour phenotype. This can be vital for a successful breeding program, because certain coat colourations can substantially improve the marketability of an animal. Furthermore, such a diagnostic test would also be an interesting tool to investigate the DNA of an extinct ancestor of the domestic horse, if it is true that dun was their common phenotype.

Results

We found seven substitutions (four of them were synonymous and the other three were non-synonymous) within the 16 exons of seven non-dun horses and nine dun horses including two Przewalski's horses. However, whether alone or in combination, none of these SNPs were ever linked to the dun phenotype. Therefore, the aim to establish a genetic test for the dun phenotype has failed.

Discussion

Although the *melanophilin* gene is known to be involved in several diluted phenotypes in other species (e.g. in humans, mice, dogs, cats and chickens), the detected SNPs of this investigation revealed no facts showing that there is any obvious association with the dun phenotype. Nevertheless, it could not be excluded that it is still involved in the dun

SUMMARY OF PAPERS

phenotype, because this research did not include further important sequence regions of the melanophilin gene like the UTR regions, the promoter region and the introns (potential splice site locations).

5. DISCUSSION

5.1 *Benefits of ancient DNA approaches to document horse domestication*

This dissertation is mainly based on the analysis of DNA from ancient horse remains, which date back from the Late Pleistocene period up to the Middle Ages. Due to following reasons, it was chosen to use aDNA approaches to document horse domestication. Until now, previous studies, which were based on modern DNA from the present domestic horse population, have failed to shed light on the history of horse domestication so far (e.g. Vilà et al., 2001; Jansen et al., 2002; McGahern et al., 2006a). In particular, the lack of an original phylogeographic structure and the enormous large maternal diversity in the modern domestic horse population prevented the acquisition of profound insights into the history of horse domestication. Today, there is difficulty in restricting the high quantity of modern mitochondrial lineages to a particular geographic area or specific breed (e.g. Vilà et al., 2001; Jansen et al., 2002; McGahern et al., 2006a; McGahern et al., 2006b; Cai et al., 2008; Lei et al., 2009; Figure 3 in APPENDIX, PAPER II). To complicate matters further, the wild ancestor of the domestic horse is already extinct.

Several previous studies have concluded that multiple origins of domestication (e.g. Vilà et al., 2001; Jansen et al., 2002; McGahern et al., 2006a) and an extensive Eurasian-wide exchange of horses after their domestication led to the current situation, in which there is no clear phylogeographic structure in the modern horse population (e.g. Jansen et al., 2002; McGahern et al., 2006a). Consequently, some previous studies started to analyze mtDNA of ancient horse remains in order to discover the original geographic lineage distribution (e.g. Vilà et al., McGahern et al., 2006b; Cai et al., 2008; Lei et al., 2009; Lira et al., 2009). Due to the following facts, however, the conclusions about horse domestication on the basis of aDNA were very limited so far. On the one hand, samples from ancient times are very rare and often difficult to organize, and, on the other hand, the difficulty, complexity, heavy workload and the higher expenses of aDNA analysis in comparison to modern DNA analysis consequently limited its application. As a result, all previous aDNA studies were characterized by a small number of ancient samples and were also restricted to a few geographic regions. Moreover, the dataset of ancient samples from previous studies has huge time- and/or spatial-gaps (with the exception of Lira et al., 2009), which are very unfavourable for drawing consistent conclusions. Finally, until now, no samples have been analyzed from the Eurasian steppe from the time period when horse domestication is believed to have occurred (ca. 4th millennium BC). Although samples from this region and time period are especially promising for restructuring the original Eurasian phylogeographic pattern, previous studies, on the contrary, analyzed ancient samples, which date back to the Late Pleistocene period, long before the suspected domestication event, and which also originate mostly far away from the Eurasian steppe (e.g. Vila` et al., 2001; McGahern et al., 2006b; Weinstock et al., 2007; Seco-Morais et al., 2007). The study of Lira et al., (2009) is the only investigation of ancient horse remains from the time period when horse domestication is believed to have occurred. However, their sample set includes only horses from the Iberian Peninsula. In fact, the majority of equine aDNA studies investigated samples from the Iron Age and the Middle Ages (e.g. Di Bernado et al., 2004; McGahern et al., 2006b; Weinstock et al., 2007; Cai et al., 2008; Lei et al., 2009; Priskin et al., 2010) when the original phylogeographic pattern was probably blurred as the result of an extensive human mediated transportation among different Eurasian regions, assuming that around 2000 years have

DISCUSSION

passed since the first horse was domesticated (Outram et al., 2009). Therefore, Chalcolithic and Bronze Age samples of domestic horses could be more meaningful. In fact, the Iberian Bronze Age samples confirm its private Iberian character (Lira et al., 2009), and, although the studies of Cai et al., (2008) and Lei et al., (2009) revealed a richer genetic diversity within the Bronze Age samples, there is obviously evidence for private mtDNA types from the East Asia. On the basis of these findings, the decision was made to use aDNA approaches based on samples of ancient wild and domestic horses, in particular from the Eurasian steppe and the time when the domestication event is suspected. Nonetheless, this study also includes samples from the Late Pleistocene period up to the Middle Ages, which also come from regions outside of the Eurasian steppe. Thus, it would be possible to trace changes either of the geographic distribution from a maternal lineage or from an investigated coat colour phenotypic trait over time, to detect an independent domestication centre outside of the Eurasian steppe. Finally, this dissertation has produced the largest available database of ancient Eurasian horses for reconstructing a meaningful original Eurasian-wide phylogeographic pattern of the horse (Figure 4; Table S1 in APPENDIX, PAPER I, supporting material on *Science Online*; Table S2 in APPENDIX, PAPER II, SUPPORTING INFORMATION).

Furthermore, it is necessary here to point out that the investigation of the coat colour variation in ancient horse populations is based on nuclear DNA. This is quite remarkable, because to date most previous aDNA studies were done on the basis of mtDNA due to the much higher copy number of mtDNA contained in cells in comparison to nuclear DNA's two copies (Hofreiter, 2008). Consequently, it is not surprising that this kind of aDNA approach has generated much interest in the scientific community. While such an aDNA approach would have been deemed illusory a few years ago, recent progress in aDNA extraction and amplification have allowed the analysis of targeted nuclear DNA sequences even from non-permafrost specimens. Not surprisingly, the success rate of aDNA amplification depended on the climate but also on the type of DNA. Whereas the success rate of nuclear DNA was 91% in Siberia, 52% in East Europe and 35% in Spain, the success rate of mtDNA was 100% in Northeast Siberia, 86% in East Asia, 80% in west and south Siberia, 44% in Asia Minor/Armenia/Europe (Spain not included) and 42% in Spain.

It is also noteworthy that methods for extracting and amplifying sequences using aDNA can be prone to errors caused by post-mortem modifications of the DNA strand (Hoelzel, 2005). Therefore, the data authentication of this method required a considerable work-load. DNA sampling, extractions and pre-PCR preparations of this work were carried out in laboratories dedicated exclusively to aDNA following standard procedures to avoid contamination. Direct sequencing was performed multiple times for each fragment from two independent extractions to detect PCR errors and miscoding of aDNA lesions from authentic sequences. Also, independent replications were carried out in different laboratories for some bone samples. Dealing with aDNA further required a frugal extraction and amplification method, because of the small amount of aDNA in fossil records. To mention only the two-step multiplex approach (Römpler et al., 2006b), which has in this study proved to be a very powerful method for analyzing fragmented aDNA sequences of short (e.g. SNP typing) but also larger sequences (e.g. mitochondrial lineage classification). This method allowed obtaining a maximum of information from a minimum amount of sample material. A further advantage is in the enhanced authentication potential, because all amplifications simultaneously start from the same aliquot of target DNA.

DISCUSSION

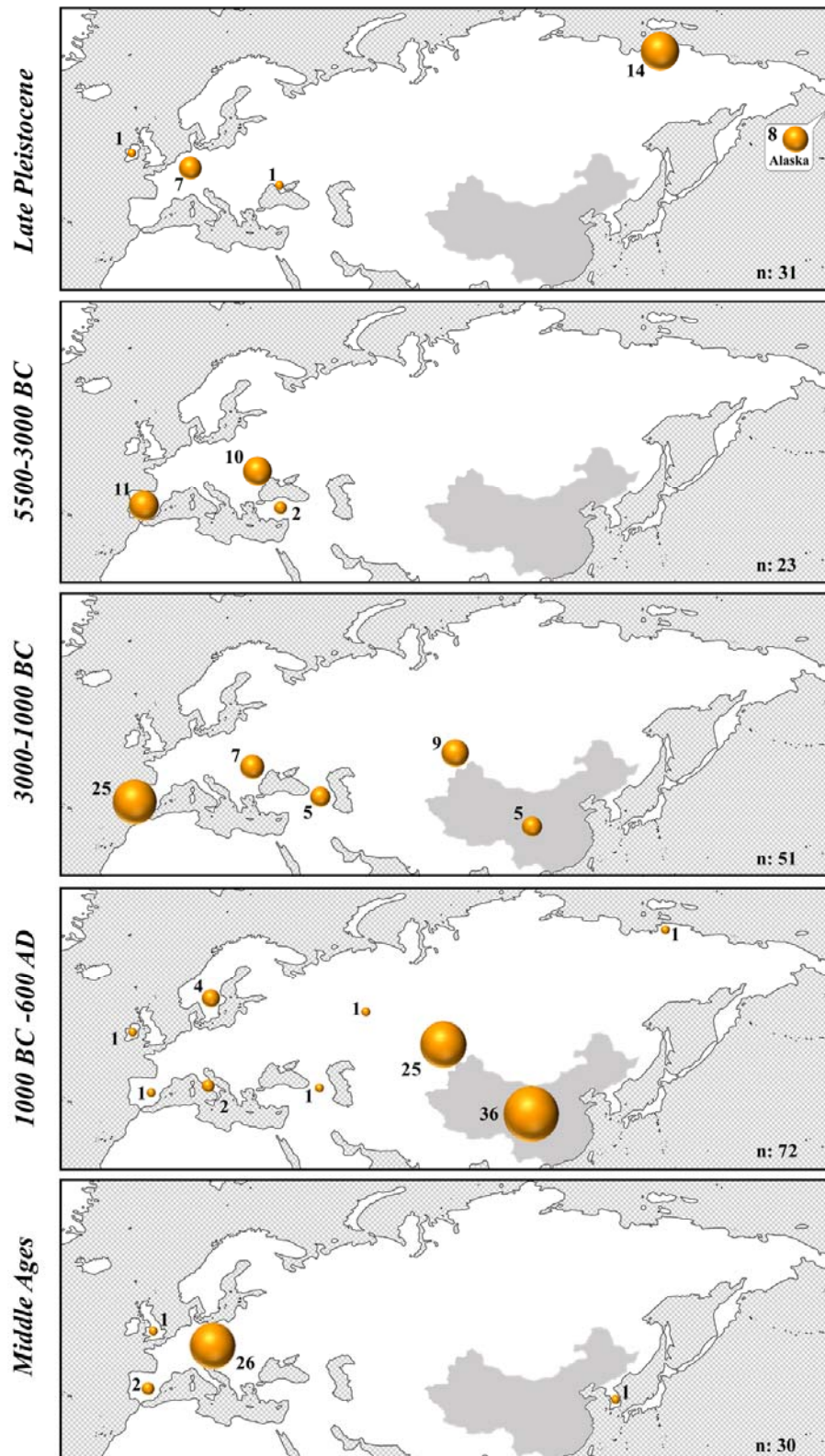


Figure 4. This figure illustrates the chronology of the geographic distribution of the ancient horse samples for the mitochondrial analysis. Size of circles and numerals represent sample frequencies (Table S2 in APPENDIX, PAPER II, SUPPORTING INFORMATION).

5.2 Mitochondrial HVR1 part for the equine lineage classification

An adequate lineage classification is the fundamental prerequisite of a meaningful phylogenetic population study, but also of discovering the wild progenitor of a domestic species. In particular, the lineage classification on the basis of the mitochondrial D-loop sequence has proven to be an important tool for understanding intraspecific relatedness and inferring the history and origin of domestic animals (e.g. Guo et al., 2006; Driscoll et al., 2009; Bonfiglio et al., 2010; Kimura et al., 2010; Larson et al., 2010). Consequently, most previous phylogenetic analyses on horses are done on the basis of the mitochondrial D-loop (e.g. Yang et al., 2002; Cozzi et al., 2004; Kakoi et al., 2007; Pérez-Gutiérrez et al., 2008), and in particular only on the HVR1 region of the D-loop when the approach includes aDNA (e.g. Jansen et al., 2002; McGahern et al., 200a,b; Lira et al., 2009).

First, the complete D-loop part (HVR1 and HVR2) was investigated in this study on its parsimony informative sites with the intent to improve the phylogenetic information. However, in contrast to other domestic species (Achilli et al., 2009), there was no enhancement on the phylogenetic resolution through the enlarged sequence part. Therefore, the sequences were finally truncated to nps 15,494–15,740 (HVR1) to accommodate most short published sequences. By doing so, the dataset of 87 new sequenced ancient horse samples of the phylogenetic part of this study could therefore be enlarged by 120 ancient and 1754 modern horse samples from public databases with the result that these 1961 horse samples can be subdivided into 434 different haplotypes.

However, in the run up of lineage classification, it can be advisable to identify strong hypervariable sites in order to exclude them from phylogenetic calculations and minimize effects of the homoplasy (e.g. parallel mutation or back mutation events). For instance, lineage classification of previous studies (Jansen et al., 2002; McGahern et al., 2006ab) did not use the complete HVR1 sequence for their lineage classification. Since Jansen et al., (2002), who calculated the hypervariable sites on the basis of 652 modern horse sequences and identified three mutational hotspots (the nucleotide positions 15585, 15597 and 15650; Table X), several subsequent studies also excluded these hotspots and thus took over their lineage classification.

However, more than half a decade has passed from the time when Jansen et al., (2002) established their lineage classification on the basis of 652 horse samples. In the meantime, a number of the 1961 mitochondrial HVR1 sequences of modern and ancient horses were available to verify the HVR1 sequence part for its mutational hotspots anew (1754 modern and 120 ancient samples of the GenBank, and 87 ancient samples of this study; July 19, 2010). Therefore, this study reviewed the mutational hotspots defined by Jansen et al., (2002) with the additional intent of detecting new strong hypervariable sites, the “maximum number of mutations at site”, within the HVR1 sequence. As a result, four strong mutational hotspots were identified. Three of them were the hotspots described by Jansen et al., (2002), but the fourth one was a novel hotspot at position 15604 and, therefore, first described by this study. Once these four mutational hotspots were eliminated from the dataset of 1961 horse samples, the number of haplotypes fell from 434 to 320. Due to the exclusion of one more mutational hotspot in comparison to the lineage classification of Jansen et al., (2002), it was necessary to redefine haplotypes and provide them with a new nomenclature (Figure 1 in APPENDIX, PAPER II). In fact, the naming of haplotypes presupposes the information either of their relationship among themselves or their spatial distribution, because, usually, meaningful names of haplotypes are given after phylogenetic and/or geographic criteria. However,

DISCUSSION

because of the situation that the maternal lineages' distribution did mainly not follow any geographic structure (Table 1, Figure 2,3,5 in APPENDIX, PAPER II), the new nomenclature was based on their phylogenetic criteria. The phylogenetic analysis of this study is based on a median network algorithm for the following reason. Frequently, the reconstruction of phylogeny from intraspecific mtDNA data is challenging because of small genetic distances between individuals. Whereas traditional tree-building methods are often unsatisfactory when applied to mtDNA data because of the high level of homoplasy, median network algorithms have been much more suitable so far. This is mainly due to the fact that networks can display alternative potential evolutionary paths in the form of cycles. In fact, the median network algorithm of this study is related to the networks of previous studies from Jansen et al., (2002) and McGahern et al., (2006b). They indicated seven major maternal haplogroups A–G on the basis of a reduced median network. These can be additionally grouped into 19 distinct phylogenetic clusters (A1–7, B1–2, C1–2, D1–3, E, F1–3, G). However, in contrast to the reduced median network algorithm, which requires binary data, the median joining network algorithm of this study calculates with the multi-state data adenine (A), cytosine (C), guanine (G), thymine (T) and the gap (–). This was intended to compute a more concise picture of the data. Unfortunately, the inclusion of all the sequences available for modern horses (n=1754) and ancient samples (n=207), which result in 320 maternal lineages, led to a strongly reticulated and unresolved network, even when the data was calculated by a reduced median network algorithm (cf. section 5.4). Therefore, the median joining network in this study was calculated focusing on 101 lineages from 46 primitive horse breeds (599 horse samples), two Przewalski's haplotypes and the 207 ancient samples (87 lineages), which yielded a total of 118 maternal lineages (Figure 1 in APPENDIX, PAPER II). This network was well-structured by formations which mainly branch and cluster (e.g. star-like pattern). Most of the 87 ancient haplotypes cluster within the major haplogroups (A–G; Figure 1 in APPENDIX, PAPER II) from previous studies (Jansen et al., 2002; McGahern et al., 2006a). However, some of the ancient haplotypes cluster in 5 newly defined haplogroups E, H1, X1, X4–5 (Figure 1 in APPENDIX, PAPER II).

5.3 Lineage diversity and geographic lineage distribution of ancient wild horses

Although population genetic analyses of a domestic species allow insight into the history of this species, the knowledge about the lineage diversity and the geographic lineage distribution of the ancestral wild population is extremely helpful, even crucial, for reconstructing the initial situation of a domestication event. Due to the fact that this study analyzed ancient horse remains of wild horses from the Late Pleistocene period up to the pre-domestic times, as well as early domestic horses, it could be possible to reconstruct the lineage diversities and geographic lineage distributions during the time when the ancestral wild horse was transformed to a domestic species.

Two key results of this study were the identification of evidence for large lineage diversity and the recognition of an absence of a clear phylogeographic structure of the pre-domestic horse population from the Eurasian steppe. Obviously, this wild horse population possessed no, or only a strongly reduced phylogeographic structure, because of a panmictic population organization and a high level of genetic exchange across the large steppe area. This assumption is based on the geographic distribution of Late Pleistocene maternal lineages, which indicates a panmictic horse population ranging from the Pyrenees to Alaska, and also

DISCUSSION

on the huge haplotype and nucleotide diversity of pre-domestic horses from the Ponto-Caspian region in the western Eurasian steppe. In fact, although the dataset of this study has a bias of Asian pre-domestic samples, because of a lack of horse remains in archaeological records, one can, in all probability, assume that the horse population of the entire Eurasian steppe generally had a large maternal lineage diversity without a phylogeographic structure. This could be due to the lack of geographic barriers within this large region, which could have restricted the gene flow of such a highly mobile species. This suggestion is additionally emphasized by the fact that, a panmictic population structure apparently seems to be common for a steppe species of genus *Equus*. Molecular research on the basis of nuclear markers and mtDNA led to similar results for the plains zebra population in Africa. Lorenzen et al., (2008) showed a high level of mtDNA genetic variation but no geographical structuring of mtDNA haplotypes within this species. Evidently, a high level of genetic exchange across large areas led toward the trend that a phylogeographic pattern consequently disappeared.

However, it is currently important to mention that a huge mitochondrial diversity without a phylogeographic structure in the wild horse population stands in contrast to the present opinion that is mainly based on the study of Vila` et al., (2001). They sequenced mtDNA of eight fossil wild horses from Alaska, with ages ranging between 12.000 and 28.000 years. Six of the eight ancient samples clustered monophyletically, which led to the hypothesis that such a degree of mitochondrial homogeneity was typical for a wild horse population. However, in consideration of the results from this study, it is more likely that this genetic uniformity was the result of the fact that a part of Alaska was a glacial refuge, which led to the genetic isolation of the Alaskan maternal lineages.

During the Pleistocene period, glaciers covered a large part of Eurasia. Therefore, horses were forced to populate an ice-free refugium, which led to wide-ranging horse migrations. Such refugia were the Iberian, Italian and Balkan Peninsulas, further parts of southern Europe, northern Siberia and western Alaska (Petit et al., 2003; Geml et al., 2010; Habel et al., 2010; Varga, 2010), which enabled the preservation of the horse population. However, the repeated alternation between glacial and interglacial periods split and fused horse populations over and over. Whereas glacial periods led to the temporal genetically isolated horse populations in different refugia, where individual lineage(s) thus evolved, in turn, interglacial periods led to the scenario in which the previously isolated populations could again rearrange themselves and fuse. Consequently, this increased the lineage diversity of such a newly arranged population. Most likely, the huge mitochondrial diversity of the pre-domestic horse population represents such a scenario. This is due to the situation that because of global warming at the end of the Late Glacial Time, the glacier ice sheets melted, resulting in profound ecological changes of the Eurasian landscape. Thus, the Eurasian grassland steppe was again accessible to horse populations from some glacial refugia, and the steppe became a gathering place for such former genetically isolated horses, probably though because of multiple ideal ecological conditions. In addition, this led to the mixing of these immigrant horses with those horses already living there. In this context, it is interesting to mention that many Paleolithic horse populations disappeared during this time from some Pleistocene refugia such as the Italian Peninsula (Bedetti et al., 2001), the Levant (Davis, 1980), Alaska (Vila` et al., 2001), the Central Balkans, as well as elsewhere in Europe, e.g. Serbia (Forsten and Dimitrijevic, 2004). However, the reason for their disappearance is not clear. Most likely, climate changes but also human impact could have been responsible.

In particular, previously genetically isolated populations from the Balkan Peninsula, southern Europe and northern Siberia could be such populations, which particularly contributed to the maternal diversity of the pre-domestic wild horse population from Eurasian steppe. However, in contrast, one can assume that horse populations from the Iberian and

DISCUSSION

Italian Peninsulas most likely were still geographically isolated by the Pyrenees and Alps. Whereas no statements can be made about lineages from Italy, because this study did not include samples from Late Pleistocene Italian wild horses, this assumption, however, can be confirmed by Iberian samples, because the Iberian pre-domestic wild horse population seems still to be genetically isolated. These data agree with the result of Lira et al., (2009) that the Iberian Peninsula was a Pleistocene-refuge and that the maternal lineages from this time period were private Iberian haplotypes, which survived there in the wild until at least the Bronze Age.

On the basis of this study, the Eurasian pre-domestic wild horse population can be geographically subdivided into at least two or even three populations: (1) the Iberian population, (2) another huge panmictic population of the Eurasian grassland steppe that stretched from the western borders of the steppes of Hungary to the eastern border of the steppes of Mongolia, and probably (3) a more or less isolated East Asian population. Whereas the Pyrenees effectively cut off the Iberian Peninsula from the rest of Eurasia, there was probably also an isolation of Chinese wild horses due to the Altai-Mountains and the Takla Makan and Gobi deserts. These geographical barriers could have effectively segregated them from their relatives in the Eurasian steppe.

5.4 Lineage diversity and geographic lineage distribution under domestication

One interesting result of this study was that the Eurasian early-domestic horse population already shows a huge haplotype and nucleotide diversity (Table 2. of APPENDIX, PAPER II). This indicates that a particularly large number of wild mares must have been domesticated. However, this stands in contrast to the common assumption that a domestic animal was most likely affected by a genetic bottleneck. Such a restricted origin hypothesis postulates that domestication started with a reproductive isolation of a small number of individuals from their wild ancestors (Vilà et al., 2001; Zeder, 2006). As a result, it must therefore be expected that there should be low lineage diversity within a newly established domestic animal stock. In fact, however, mitochondrial data of this study point to an already huge maternal diversity at the beginning of horse breeding, which can only be interpreted to mean that a huge number of wild mares were captured for domestication, which probably combined with introgressions events.

Until now, previous studies interpreted such huge maternal lineage diversity in the current modern domestic horse population as the result of the founder horses of the domestic stock being recruited from a widespread area. This in turn raises the question of a multiple origin scenario of horse domestication (e.g. Vilà et al., 2001; Jansen et al., 2002; McGahern et al., 2006a). Jansen et al., (2002) calculated that modern horses are descended from at least 77 different wild mares. In particular, the fact that there is an extensive genetic diversity among these 77 ancestral mares led to their conclusion that these maternal lineages were partly descended from genetically distinct horse populations. However, findings of this study could shed new light on the question as to whether there was a multiple origin scenario or not. Data of this study allow for establishing an alternative explanation to a multiple origin hypothesis, which is predicted to justify the huge maternal diversity in modern domestic horses. Making the assumption that at least three conceivable circumstances (all are indicated by the results of this study) were present at the same time: (1) a great number of founder mares came from the wild horse population of the Eurasian steppe; (2) this population was characterized by a huge number of maternal lineages without, however, a phylogeographic structure; and (3) a very large number of wild mares were introduced into the domestic animal stock, which scenario

DISCUSSION

could also explain why there is such huge mitochondrial diversity in the modern but also in the early-domestic horse population. Since the recent study of Outram et al., (2009), which pinpointed horse domestication to the northeast of the Eurasian steppe, there is more strong evidence than ever that most ancestral lineages of the modern domestic horse were recruited from there. Unfortunately, the dataset of this study possesses the already-mentioned bias of pre-domestic samples from the Asian part of the Eurasian steppe. As a result, it was not possible to clearly verify this hypothesis that the Eurasian steppe is the origin of horse domestication. In all probability, this bias was the main reason that, in fact, only three lineages were found in the pre-domestic horse population from the Ponto-Caspian region (out of eight lineages), which have survived in modern horses. Consequently, this question must be clarified in future investigations, which also include ancient Asian samples of pre-domestic horses.

However, assuming a huge panmictic pre-domestic horse population in the Eurasian steppe without a phylogeographic structure, it would be fatal to attempt to localize an exact domestication site within this vast area on the basis of the geographic mitochondrial lineage distribution. It would be only possible to draw the conclusion that a domestic lineage may come from this huge population, and was in all probability domesticated somewhere in this huge region; but the question of the precise geographic site of the domestication event or even the question of multiple domestication events will have to remain unanswered.

Nevertheless, data of this study of early-domestic horses also provide evidence for multiple introgressions or, alternatively, for a second wave of domestication during Iron Age, because many new haplotypes were discovered in the gene pool of domestic horses from this time period. Likewise, the mismatch distribution of the early-domestic European and Siberian horses produced evidence for a bottleneck 1500 years after the postulated origin of domestic horses in north Kazakhstan around 3500 BC and consequently indicates horse breeding in these regions (Outram et al., 2009). However, the trimodal character of the curve of the mismatch distribution from early-domestic Chinese samples could be an indication of local horse breeding, or it could also be evidence for mainly horse imports from different remote regions or the breeding of introgressed Chinese private lineages (Figure 4 in APPENDIX; PAPER II).

Although, most likely, the largest part of early-domestic lineages originated from the panmictic Eurasian steppe population, some lineages were probably domesticated and/or introgressed into other regions of Eurasia. In fact, some early-domestic haplotypes were obviously restricted to a geographic region. For example, the haplotype E and haplogroup K3 most likely had their roots in East Asia, so it is not surprising that both are also common in East Asian primitive breeds today. This agrees only partly with previous studies (McGahern et al., 2006a; Lei et al., 2009). Lei et al., (2009) support the idea that haplogroup K2 and K3 may have been taken into the domestic gene pool in the East Asia. However, the result of this study revealed that this is only true for the K3 haplogroup, whereas the K2 haplogroup most likely originated from the Eurasian steppe population. Furthermore, the B and X1 haplogroup, and the X13 and X16 haplotypes were obviously restricted to the European region and were therefore introgressed in European domestic horse stocks. These lineages predominate even in modern European horse breeds. Notably, the Iberian private haplotypes of the H1-group indicate introgression events on the Iberian Peninsula. This conclusion is based mainly on the fact that even though lineages of the H1-group were the dominant ancient Iberian lineages during pre-domestic and early-domestic times, today, they are only occasionally present in Iberian (e.g. Marismeño, Lusitano, Caballo de Corro) and in South American breeds (Argentinean Creole, Puerto Rican Paso Fino). However, in the case of an independent Iberian domestication centre, one should assume that the original dominant lineages of this

DISCUSSION

region were particularly used for domestication. Furthermore, another Iberian private haplotype, the B haplotype, supports the assumption that the Iberian lineages were only used for introgression. This singular Iberian haplotype is presently found in Percheron (France), Arabian and Wielkopolski (Poland) horses. Such presence in the Percheron horses could be an indication of trade between Iberian and French people in ancient times, whereas the presence in the Arabian breed is likely a relic from the Moorish occupation of south Iberia. The Wielkopolski horse is also a breed, which is well known for containing genes of the Arabian horse breed.

All these results agree with the previous study (Lira et al., 2009) about the lack of evidence for an independent horse domestication process in Iberia, at least not on a significant scale, and provide instead strong evidence for introgression of local wild mares. This assumption is additionally confirmed by the extremely ragged mismatch distribution of Iberian horse samples from the time of Chalcolithic period and the Bronze Age (Figure 4 in APPENDIX, PAPER II). Until the publication of Lira et al., (2009) and the results of this study, it was frequently assumed that the Iberian Peninsula was an independent domestication site as proposed in several previous studies (Jansen et al., Royo et al., 2005). Furthermore, Jansen et al., (2002) and Royo et al., (2005) assumed an Iberian origin of the D1 haplogroup (designation according to Jansen et al., 2002), due to the fact that the D1 mtDNA group appears to be associated mainly with breeds originating in the Iberian Peninsula. However, results of Lira et al., and the data of this study yielded no association between an Iberian origin and the D1 group (Seco-Morais et al., 2007; Lira et al., 2002). The sample set of this study clearly shows that the D1 group first appears during the Bronze Age in Europe, but not until the Middle Ages in Spain.

To sum up, it is therefore possible to draw the conclusion that the large diversity of mitochondrial lineages in the modern domestic horse population is probably not only a product of (1) multiple origins, but resulted from a combination with (2) an extremely large number of female founders from a wild horse population, (3) which was already characterized by a huge maternal lineage diversity, (4) large-scale introgression of local lineages into the domestic stock, (5) an extensive Eurasian-wide exchange of domestic horses soon after the domestication and, finally, (6) an extensive animal breeding. Consequently, it is not surprising that there is no breed specific maternal lineages distribution within current horse breeds and certainly not in so-called modern breeds, whose roots mostly do not predate the eighteenth century or later (Samraus, 1986). Most likely, one can assume that even primitive horse breeds, which have been known since ancient times (e.g. Akhal Teke, Arab, Barb), were characterized by a relatively large intrabreed mitochondrial diversity (e.g. Royo et al., 2005; Głażewska et al., 2007; Priskin et al., 2009). However, the inbreeding situation of many breeds could most likely be responsible that microsatellite studies indicate breed specific molecular markers (e.g. Vilà et al., 2001; Georgescu et al., 2008), even if these breeds are characterized by a huge mitochondrial diversity without clear breed specific maternal lineages.

After 5500 years of horse breeding, the world is populated by 58 million domestic horses (FAO Statistics Division 2010 for 2008; <http://faostat.fao.org>). The public databases (July 19, 2010) contain the huge number of 274 archived maternal lineages (according to the lineage classification of this work). Although this study documents that as much as 30% of the maternal lineages has been lost during 5500 years of horse breeding, most likely, extensive horse breeding and population growth led to the continuous increase of maternal lineages through mutations. Even though Jansen et al., (2002) calculated a relatively low mutation rate

DISCUSSION

of one mutation per 100.000 years within the 244bp long mitochondrial HVR1 part investigated here, certainly, the extensive breeding, in particular, of the modern horse breeds, contributed to this huge lineage diversity. This becomes even more evident in the following observation of this study. As mentioned in section 5.2, the inclusion of a large number of maternal lineages from modern horse breeds particularly introduced complex reticulated network formations, whereas the inclusion of lineages from primitive breeds were unproblematic (rarely led to reticulation). This could be a hint that certain modern breeds contribute to the largest part of mitochondrial diversity of the current domestic horse population. It is known that newly formatted modern breeds underwent a strong breed size growth through their popularity. On the contrary, it is supposed that primitive breeds were more or less excluded from modern interbreeding, and, in addition, these breeds were mainly kept small and sometimes shrank because of lack of contemporary attractiveness. Nevertheless, other possible explanations should not be discarded. Perhaps, modes of an mtDNA recombination mechanism in horses, to date unknown, or links between mtDNA polymorphism and economically important traits could be responsible for this phenomenon. Presently, little is known about mtDNA recombination, and although several studies have been done in farm animals trying to identify links between the mtDNA and productive traits, however, such associations still remain doubtful (Schutz et al., 1994; Boettcher et al., 1996; Mannen et al., 1998, 2003; Sutarno et al., 2002). The probability that regionally specific mitochondrial haplogroups in horses have been shaped by climatic selection (Mishmar et al., 2003) seems unlikely at first glance, because of their high mobility across large-sized areas. However, one should not leave neglect to consider a link between mtDNA polymorphism and an economically important trait, since a recent study concluded that the high-altitude environment has directed adaptive evolution of the mitochondrial ND6 gene in the plateau horse (Ning et al., 2010).

5.5 Coat colour variation of ancient wild horses

As this and the next section of the discussion are based to a certain extent on co-authored material, I would like, first of all, to give due credit to my co-author. I was responsible for the extraction, amplification and analysis of the ancient Iberian sample set. Although 45 horse remains were available, only 16 horse samples, ranging from the Mesolithic period up to the Middle Ages, were successfully typed for the coat colour SNPs (eight mutations in six genes of nuclear DNA). This low success rate was not surprising, because the rate of aDNA amplification depended on the climate (cf. section 1.5 and 5.1). However, in particular the Iberian samples played a substantial role in the context of the question of whether the Iberian Peninsula was a further independent ancient centre of horse domestication.

It is assumed that the wild ancestor of the domestic horse was uniform in coat colour phenotype, because wild species usually show species-specific colouration. However, until this study, the coat colour phenotype of the wild ancestor of the domestic horse was not known, because it became extinct. As mentioned in section 3, the Przewalski's horse most likely is not the ancestor of the domestic horse. Nevertheless, its uniform dun coat colour phenotype probably still reflects the coat colour of the ancestor of the domestic horse. In fact, prehistoric cave paintings of horses look very similar to the coat colour phenotype of the current Przewalski's horse. However, written records about coat colouration, drawings and even a photo of the extinct Tarpan paint a picture of a coat colouration, which varied little

DISCUSSION

(tan, Isabella, or mouse; Smith, 1866). In this context, it is important to mention that some doubt exists that the Tarpan population was a pure wild horse; it might, instead, be a feral horse population or at least a wild horse population, which was influenced by introgression of the domestic horses. Therefore, probably, it does not represent the true coat colouration of the ancestral wild horse.

This was the first genetic study to investigate coat colour variation in the extinct wild horse populations from the Late Pleistocene period up to pre-domestic times. A total of five genetic tests were available to detect the three basic coat colours (bay, black and chestnut), two kinds of dilution (silver and cream) and three spotted or painted colourations (overo, tobiano and sabino). Unfortunately, this study failed to establish a genetic test for the dun phenotype. However, this was not as daunting as it first sounds, because the primary objective was to detect the general state of coat colour variation within a horse population. There is no question that the genetic test for dun would be an additional helpful tool for verifying the distinction between a wild and domestic phenotype. Therefore, it remains important to continue developing such a diagnostic test for dun.

The results of this investigation clearly suggested that the Siberian and European Late Pleistocene horses were bay in colour (or probably bay-dun). Interestingly, in contrast, the Iberian Early Holocene horse population was a mixed population of 50% black and 50% bay horses. Equally, the east European samples (5th-4th millennia BC), dating to the millennia before archaeological evidence for domestication exists, had a similar mixed coat colour distribution. Unfortunately, this study has no information about the colour of Siberian Neolithic and Chalcolithic horses, because of a lack of bones in archaeological records during the time of the 6th to the 3rd millennia BC in this region. Since black horses are found neither in Siberia nor east/central Europe during the Late Pleistocene period, their occurrence in the Early Holocene period is probably the result of postglacial immigration or selection. However, human selection is less likely, as there were no colour phenotypes other than black and bay in this time period (Warmuth et al., 2011). In addition, a binomial test verifies that the chance of this study missing any other colour allele was statistically impossible (Table S7, supporting material on *Science Online*). Most likely, the occurrence of black coloured horses was caused by an adaptation event through climate changes, which in turn led to increased forest cover. Recently, similar results were shown for wolves (Anderson et al., 2009). Obviously, darker coat colour was better for adaptation to the forest, whereas the lighter brown colour was more suitable for the steppe land.

5.6 Coat colour variation under domestication and early breeding

Given the rapid change of coat colour variation, it can be concluded that horse domestication started in the Eurasian steppe region at the latest around 3000 BC. Whereas the coat colour analysis of the Iberian and east European horse populations of pre-domestic times suggested that these were either black or bay coloured, there was a significant increase in the coat colour variation in both Siberia and east Europe 500 years after the earliest known traces of horse domestication to date (Outram et al., 2009). The first observed coat colouration other than black or bay was the chestnut colouration, which was first present in a domestic Siberian horse from the northeastern border of the Eurasian steppe (3rd millennium BC). During the Bronze Age, the frequency of the chestnut colouration increased rapidly in Siberian and European domestic horses. As the first spotting phenotype, the sabino colouration also

DISCUSSION

appears during the 3rd millennium BC in Siberia, and later in Armenia and Moldavia during the middle Bronze Age. However, the tobiano spotting was only found in a single Eastern European sample from 1500–1000 BC, and the cream (buckskin) and (black) silver dilutions were observed only from ca. 800–600 BC in Siberia. On the basis of these results, the data are best explained by selective breeding, which supports domestication as the cause for the observed changes in colouration, and demonstrates that domestication started in the Eurasian steppe region at least by ca. 3000 BC. However, due to the fact that there was no colour change in Spanish samples until the Middle Ages, the data suggest that this region can most likely be excluded as a domestication or important breeding centre at least until the Iron Age. Therefore, the result of this study is clearly contrary to the conclusion of Warmuth et al., (2011), who obviously still believe that the genetic contribution of Iberian wild stock to local domestic horses may have been substantial.

In conclusion, this approach to localizing a domestication site with regard to an increase of coat colour variation can be considered suitable, and its practicability is hereby confirmed. It thereby displays a novel and very promising tool to decode especially the history of horse domestication, because the data make, for the first time, a very clear distinction between wild and domestic horses. Today, the majority of coat colour phenotypes are not confined to a single breed and the same mutation is responsible for a certain colour variant across breeds. This supports the idea that mutations producing the colour variants occurred prior to breed formation in the time or soon after the domestication process and also maintains the hypothesis that there was an extensive horse exchange soon after horse domestication. However, in the future it might be possible with a sufficient number of samples that data will display a more detailed pattern of the geographic distribution of certain coat colour phenotypic traits within the early domestic horse population. Thus, changes in the distribution of specific coat colourations soon after the domestication could help to reconstruct human mediated exchange events of domestic horses between widely-dispersed early breeding regions.

Although it is striking that the frequency of the black allele (a) for the *ASIP* gene decreases when comparing the dataset of pre-domestic wild horses with the early-domestic horses, the percentage of black coloured early-domestic horses from the Bronze Age is relative large. Of course, it may be that bay horses as carriers of the black allele (a) were domesticated, and later humans selected for the black coat colour. However, the assumption that horses, which were already black, were domesticated could be interesting in the following contexts. On the one hand, several studies discussed a molecular link between a pale coat colour and a lower aggressive behaviour in mammals. Actually, under this assumption, one would think that in particular horses with a pale coat colour were preferably captured for domestication. Nevertheless, this data reveals that black horses were most likely also used for domestication. The reason for such a correlation between the coat colour and the behaviour could be the common synthesis pathway of adrenalin and dopaquinone from dopa (Keeler et al., 1970; Laukner 1998). However, until now, there is no clear evidence for such a link. Nevertheless, the frequent pleiotropic effects on neuronal tissues through mutations in coat-colour-associated genes could explain such a correlation.

On the other hand, black coloured horses within an early-domestic horse population can indicate that some ancestors of these horses were captured from a temperate forested region or at least from a region of dry-woodland. This assumption results from the evidence that black coat colouration is an adaptation to an increased forest cover as recently shown for wolves (Anderson et al., 2009). Obviously, darker coat colour is better adapted to a forested environment, whereas the lighter brown colour is more suitable for camouflage within

DISCUSSION

grassland steppe and savannah. Today, it is known that after the close of the glacial period, formerly steppe regions became forested through climate changes. In particular, central Europe was largely temperate forested at the time when the first domestic horses appear (Warmuth et al., 2011), whereas the Iberian Peninsula was characterized by savannah and dry-woodland and the Eurasian steppe was also a region of savannah, dry-woodland and grassland. As noticed in the section 5.5, coat colour data of this study indicate such a scenario of coat colour adaptation. Whereas, the Siberian and European Pleistocene horses were bay in colour, the coat colouration of east European and Iberian early Holocene wild horse populations changed into a mix of black or bay. Probably, this mixed population of black or bay coloured wild horses resulted through a combination of climatic changes, an adaptation process, migration movements and their high mobility.

5.7 Chronological appearance of investigated coat colour phenotypes

Probably, the chronological order of the first appearance of the investigated coat colour phenotypes in this study was not only coincidental (spontaneous mutation) within the early-domestic horse population (and subsequent fixation by human selection) but also defeated by its allelic frequency within the ancestral wild horse population. This can be assumed, because a huge number of wild mares were domesticated or introgressed probably also from widely dispersed wild horse populations, or at least from the Eurasian steppe, which was most likely a Early Holocene gathering place for former genetically isolated horses (cf. section 5.3). The circumstance that many founder individuals were used for the domestic stock in turn increased the chance that a wild horse be introduced, which either carries two different alleles of a rare recessive coat-colour-associated gene or was even a homozygote of a very rare coat colour trait. Different frequencies of rare coat-colour-associated alleles (non-wild-type alleles) within a wild population can definitely depend on how strong the associated coat colour phenotype disadvantages the animal. It is known that an extremely divergent coat colour of an individual, which differs from its population, has, in all probability, existential disadvantages in e.g. survival and reproduction. This study shows that wild horses were either black or bay coloured, which was based on the colour's camouflaging quality (cf. section 5.5). In fact, the only identified non-wild-type allele, the chestnut allele (missense mutation in the *MC1R* gene), was found in an east European wild horse from the late 5th millennium BC. This allele causes the change of the basic coat colouration in a homozygote state (reddish coat colour). Obviously, the chance of reproduction of a chestnut coloured horse is limited in the wild, but natural selection was not strong enough to eliminate this allele completely from the wild horse gene pool.

However, there is most likely a further reason for different frequencies of alternative coat colour alleles within a wild horse population. The rate of an allele could also depend on its pleiotropic effect(s). Many coat colour associated genes are often involved with other developmental and cellular processes (cf. section 1.4.2). Thus, this can result in coat colour associated alleles being linked to critical pathogenic effects, which even extend to embryonic lethality. Depending on this degree of severity, pathogenic coat colour associated alleles could be individually subject to purifying selection under natural conditions. Therefore, it is not surprising that the *ASIP* and *MC1R* genes were the only ones with an allelic variability in wild horses, because they are considered to be less pleiotropic (Hoekstra, 2006). As shown in several studies, mutations in *KIT*, *EDNRB* and *SILV* can be associated with pathogenic effects to different extents. For instance, there is evidence of strong pleiotropic effects of the mutated

DISCUSSION

KIT allele especially in mice (e.g. within the gametogenesis and haematopoiesis). It should also be said at this juncture that currently, there is no evidence of such effects in sabino and tobiano horses. Furthermore, silver colourations caused by mutations in *SILV* are associated with impaired functions of the auditory and ophthalmological systems, as shown in dogs (e.g. Clark et al., 2006). However, a special case is the white foal syndrome in horses, which is caused by a mutation in the *EDNRB* gene. Heterozygous horses show the frame-overo phenotype for this mutation, and the homozygous state is even lethal in horses shortly after birth, as newborn foals die quickly through intestinal blockages (Metallinos et al., 1998; Santchi et al., 1998; Yang et al., 1998). The sustaining of this allele within a population requires the sufficient knowledge of a breeder regarding the inheritance of this phenotype. Consequently, it is not surprising that no frame-overo allele was identified in the ancient wild and or even the domestic horse samples. However, the frame-overo phenotype is presently very popular in many horse breeds. This example especially demonstrates that even phenotypes that are strongly linked to pathogenic effects can still be positively selected by breeders.

5.8 Conclusion about origin of horse domestication

Although this is the greatest dataset of aDNA, which has ever been analyzed to document horse domestication, there were still too many gaps in time and locations to make a precise temporal and geographical localization within the vast expanse of Eurasia and the extended time span when domestication potentially could have taken place. Nevertheless, given the rapid increase of coat colour variation within ancient horses, this study agrees with the apriori thesis (Outram et al., 2009) that horse domestication occurred in the Eurasian steppe at the latest by ca. 3000 BC. Although mitochondrial data suggest that the majority of maternal lineages of the current domestic horse population originated within the Eurasian steppe, mitochondrial data are probably very unsuitable to document a precise geographic site of horse domestication within this steppe region. This is due to the fact that during the time of horse domestication, there was obviously a panmictic population structure with a huge number of maternal lineages and a high level of genetic exchange across this very large area, which consequently prevented a phylogeographic population structure. However, such an ancestral wild horse population could explain the huge maternal diversity already existing among early-domestic horses. Therefore, an important outcome of this study is that the huge maternal diversity of the modern horse's mtDNA must not be the consequence of multiple origins of horse domestication only, but could also the result from a huge number of domesticated wild mares, which originated in the Eurasian steppe. Nonetheless, mitochondrial data of this study most likely indicate multiple introgressions or, alternatively, a second wave of domestication during the Iron Age somewhere in Eurasia, but certainly not on the Iberian Peninsula. Indeed, mitochondrial and coat colour data agree on the lack of evidence for an independent horse domestication process in Iberia, at least on a significant scale, and provide instead strong evidence for the rare introgression of Iberian wild mares.

5.9 Outlook

The result of this study verifies that coat colours were an early goal of ancient breeders. Therefore, coat colour alleles can be regarded as powerful molecular markers for documenting a domestication event. The practicability of this approach promises also to be usable for any other domesticated animal.

The next step in documenting horse domestication should be filling in the existing sample gaps, for instance, with more samples from the transition from the Chalcolithic period to the Bronze Age. In particular, samples from the Eneolithic sites of Botai (northern Kazakhstan) and Dereivka (Ukraine), as well as from East Asia (due to the lack of Asian samples within this study) seem to be very promising for defining the precise starting date of horse domestication. An increase of sensible chosen samples will in turn improve the geographic resolution of the first appearance of investigated coat colour phenotypes and also their traces of distribution through the expansion over time. Furthermore, there will be a better chance of detecting a human mediated coat colour variation in horses, when novel markers of newly discovered coat-colour-associated alleles are to be included. Certainly, in the near future, genome-wide association studies will shed more light onto the genetic background of even more genetically complex phenotypes. Probably, a more complex genetic architecture of the dun phenotype was the reason that the detection of its genetic background had failed so far. In fact, a genetic test for dun would be very helpful in defining whether it was really the phenotype of the ancestral wild horse and also as a diagnostic test for the better differentiation between ancient wild and domestic horses. Likewise, alleles of other early-breeding traits could be used as molecular markers (e.g. molecular marker for temperament and tameness). Without doubt, population genetic analyses will generally acquire a new dimension in the future, because the investigation of the allelic diversity of a trait not only of a recent population but also of their ancestral populations will become standard, which offers the unique chance to investigate the evolution of a trait over time.

One of the results of this study suggests that there was no original phylogeographic structure in the ancestral wild horse population of the Eurasian steppe because of a panmictic population structure with a huge number of maternal lineages and a high level of genetic exchange across this large area. This might well explain why it is, and will remain, largely impossible to discover the domestication origin on the basis of mtDNA within this steppe region, not to mention drawing conclusions about multiple origins. However, this again shows how important this novel approach of coat colour variation can be to define a domestication origin within such a region. This does not mean that analyses on the basis of mitochondrial DNA no longer make any sense, but instead, the outcome that the Iberian wild horse population was genetically isolated from the remaining Eurasian horses shows that mitochondrial studies can still be vital, when it comes to identifying private lineages from more or less genetically isolated regions outside of the Eurasian steppe (e.g. glacial refugia like Italy, Balkan Peninsula). Due to the bias of Chinese pre-domestic horse samples in this study, in the future, it will be interesting to focus on such ancient samples to determine if there were private Chinese lineages, which could indicate an isolated Chinese wild horse population from the Eurasian steppe population and probably even an independent Chinese domestication origin or simply introgression events.

Finally, this dissertation concludes that it will be more essential than ever to combine molecular and archaeological approaches in order to document ultimately the origin of horse domestication within the Eurasian steppe. The intensified cooperation between archaeologists and biologists is desirable, because this guarantees geneticists access to ancient bones in order to answer their biological questions, whereas archaeologists benefit from information about

DISCUSSION

the lineage and relationship, or characteristics of certain phenotypic traits of a dead organism. A final important point is that a successful collaboration can also minimize or even avoid pre-laboratory contaminations of a sample, which is of critical importance when analyzing aDNA.

REFERENCES

6. REFERENCES

- Achilli, A. Bonfiglio, S. Olivieri, A. Malusà, 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: e5753.
- Anderson, T.M. von Holdt, B.M. Candille, S.I. Musiani, M. Greco, C. Stahler, D.R. Smith, D.W. Padhukasahasram, B. Randi, E. Leonard, J.A. Bustamante, C.D. Ostrander, E.A. Tang, H. Wayne, R.K. & Barsh, G.S. (2009) Molecular and evolutionary history of melanism in North American gray wolves. *Science* 323: 1339-1343.
- Andersson, L. (2001) Genetic dissection of phenotypic diversity in farm animals. *Nature Reviews Genetics* 2, 130-138.
- Anthony, D. & Vinogradov, N.B. (1995) The Birth of the Chariot. *Archaeology* 48: 36-41.
- Anthony, D.W. (1986). The 'Kurgan Culture', Indo-European origins and the domestication of the horse: a reconsideration. *Current Anthropology* 27(4): 291-313.
- Araki, H. & Tachida, H. (1997) Bottleneck Effect on Evolutionary Rate in the Nearly Neutral Mutation Model. *Genetics* 147(2): 907-914.
- Avice, J.C. (2000) Phylogeography: The history and formation of species. First edition. *Harvard University Press*, Cambridge, Massachusetts.
- Ballard, J.W.O. & Rand, D.M. (2005) The population biology of mitochondrial DNA and its phylogenetic implications. *Annual Review of Ecology Evolution and Systematics* 36: 621-642.
- Bannikova, A.A. (2004) Molecular markers and modern phylogenetics of mammals. *Zhurnal Obshchei Biologii* 65(4): 278-305.
- Barton, L. Newsome, S.D. Chen, F.H. Wang, H. Guilderson, T.P. & Bettinger R.L. (2009) Agricultural origins and the isotopic identity of domestication in northern China. *Proceedings of the National Academy of Sciences of the USA* 106:5523-5528.
- Bedetti, C. Palombo, M.R. & Saredlla, R. (2001) Last occurrences of large mammals and birds in the Late Quaternary of the Italian peninsula. Pages 701-703 in Cavarretta, G. Gioia, P. Mussi, M. & Palombo, M.R. editors. The world of elephants. Consiglio Nazionale delle Ricerche, Rome.
- Bellone, R.R. Brooks, S.A. Sandmayer, L. Murphy, B.A., Forsyth, G. Archer, S. Bailey, E., Grahn, B. (2008) Differential gene expression of TRPM1, the potential cause of congenital stationary night blindness and coat spotting patterns (LP) in the Appaloosa horse (Equus caballus). *Genetics* 179: 1861-1870.
- Bellone, R.R. (2010) Pleiotropic effects of pigmentation genes in horses. *Animal Genetics* 41: 100-110.
- Besmer, P. Manova, K. Duttlinger, R. Huang, E.J. Packer, A. Gyssler, C. & Bachvarova, R.F. (1993) The kit-ligand (steel factor) and its receptor c-kit/W: pleiotropic roles in gametogenesis and melanogenesis. *Development Supplement* 125-137.
- Bettinger, R. Barton, L. & Morgan, C. (2010) The origins of food production in North China: A different kind of agricultural revolution. *Evolutionary Anthropology* 19:9-21.
- Binladen, J. Gilbert M.T.P. Willerslev, E. (2006) 800,000 year old mammoth DNA, modern elephant DNA or PCR artefact? *Biology Letters* 3: 55-56.
- Blanpain, C. & Sotiropoulou, P.A. (2010) A dominant role of the hair follicle stem cell niche in regulating melanocyte stemness. *Cell Stem Cell* 6(2): 95-6.
- Boettcher, P.J. Kuhn, M.T. & Freeman, A.E. (1996) Impacts of cytoplasmic inheritance on genetic evaluations. *Journal of Dairy Science* 79: 663-675.
- Bonfiglio, S. Achilli, A. Olivieri, A. Negrini, R. Colli, L. Liotta, L. Ajmone-Marsan, P. Torroni, A. & Ferretti, L. (2010) The Enigmatic Origin of Bovine mtDNA Haplogroup R: Sporadic Interbreeding or an Independent Event of Bos primigenius Domestication in Italy? *PLoS One* 5(12): e15760.
- Bosch, E. Calafell, F. Santos, F.R. Perez-Lezaun, A. Comas, D. Benchemsi, N. Tyler-Smith, C. & Bertranpetit, J. (1999) Variation in short tandem repeats is deeply structured by genetic background on the human Y chromosome. *The American Journal of Human Genetics* 65(6): 1623-1638.
- Boule, M. (1899) Observations sur quelques epuides fossils. *Soc. Geol. France* 27(2): 531-542.
- Brito, P.H. & Edwards, S.V. (2008) Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica* 135: 439-455.
- Brooks, S.A. Gabreski, N. Miller, D. Brisbin, A. Brown, H.E. Streeter, C., Mezey, J. Cook, D. & Antczak, D.F. (2010) Whole-genome SNP association in the horse: identification of a deletion in myosin Va responsible for Lavender Foal Syndrome. *PLoS Genetics* 6(4): e1000909.
- Cai, D. Tang, Z. Han, L. Speller, C.F. Yang, D.Y. Ma, X. Cao, J. Zhu, H. & Zhou H. (2009) Ancient DNA provides new insights into the origin of the Chinese domestic horse. *Journal of Archaeological Science* 36(3): 835-842.
- Calonje, M. Martín-Bravo, S. Dobeš, C.h. Gong, W. Jordon-Thaden, I. Kiefer, C.H. Kiefer, M. Paule, J. Schmickl, R. & Koch, M.A. (2008) Non-coding nuclear DNA marker in phylogenetic reconstructions. *Plant Systematics and Evolution* 282(3-4): 257-280.
- Cauvin, J. (2000) The Birth of the Gods and the Origins of Agriculture (*Cambridge Univ Press*, Cambridge, UK).
- Childe, V.G. (1950) The Urban Revolution. *Town Planning Review*. 21:3-17.
- Clark, L.A. Wahl, J.M. Rees, C.A. & Murphy, K.E. (2006) Retrotransposon insertion in SILV is responsible for merle patterning of the domestic dog. *Proceedings of the National Academy of Sciences* 103(5): 1376-1381.
- Cozzi, M.C. Strillacci, M.G. Valiati, P. Bighignoli, B. Cancedda, M. & Zanotti, M. (2004) Mitochondrial D-loop sequence variation among Italian horse breeds. *Genetics Selection Evolution* 36(6): 663-72.
- Crawford, D.C. Akey, D.T. & Nickerson, D.A. (2005) The patterns of natural variation in human genes. *Annual Review of Genomics and Human Genetics* 6:287-312.
- Darwin, C. (1868) The variation of animals and plants under domestication. John Murray, London.
- Davis, S.J. (1980) Late Pleistocene and Holocene equid remains from Israel. *Zoological Journal of the Linnean Society* 70: 289-312.
- Di Bernardo, G. Galderisi, U. Del Gaudio, S. D'Aniello, A. Lanave, C. De Robertis, M.T. Cascino, A. & Cipollaro, M. (2004) Genetic characterization of Pompeii and Herculaneum Equidae buried by Vesuvius in 79 A.D. *Journal of Cell Physiology* 199: 200-205.
- Diamond, J. & Bellwood, P. (2003) Farmers and their languages: the first expansions. *Science* 300: 597-603.
- Driscoll, C.A. Macdonald, D.W. & O'Brien, S.J. (2009) From wild animals to domestic pets, an evolutionary view of domestication *Proceedings of the National Academy of Sciences of the USA* 15.
- Drögemüller, C. Philipp, U. Haase, B. Günzel-Apel, A. & Leeb, T. (2007) A Noncoding Melanophilin Gene (MLPH) SNP at the Splice Donor of Exon 1 Represents a Candidate Causal Mutation for Coat Color Dilution in Dogs. *Journal of Heredity* 98(5): 468-473.
- Edwards, C.J. Gaillard, C. Bradley, D.G. & MacHugh, D.E. (2000) Y-specific microsatellite polymorphisms in a range of bovid species. *Animal Genetics* 31: 127-130.
- Excoffier, L. Smouse, P.E. & Quattro, J.M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131: 479-491.
- Flad, R. Yuan, J. & Li, S. (2007) Zooarcheological evidence of animal domestication in northwest China. In: Late Quaternary Climate Change and Human Adaptation in Arid China (Ed. by D.B. Madsen, F.H. Chen & X. Gao), pp. 167-203. Elsevier, Amsterdam, Holland.
- Forsten, A. & Dimitrijević, V.M. (2003) Pleistocene horses (genus Equus) in the central Balkans. *Geološki Anali Balkanskog Poluostrva* 65: 55-75.
- Forstén, A. (1992) Mitochondrial-DNA time-table and the evolution of Equus: comparison of molecular and paleontological evidence. *Ann. Zool. Fennici* 28: 301-309.
- Frankham, R. Ballou, J.D. & Briscoe, D.A. (2007) Introduction to Conservation Genetics. Cambridge.
- Fuller, D.Q. Qin, L. & Harvey, E. (2008) Past Human Migrations in East Asia. Matching Archaeology, Linguistics and Genetics, eds. Sanchez-Mazas, A. Blench, R. Ross, M.D. Peiros, I. Lin, M. (Routledge, London), pp 40-83.
- Gelderman, H. (1975) Investigation on inheritance of quantitative characters in animals by gene markers. I. Methods. *Theoretical and Applied Genetics* 46: 300-19.
- Geml, J. Tulloss, R.E. Laursen, G.A. Sazanova, N.A. & Taylor, D.L. (2010) Phylogeographic analyses of a boreal-temperate ectomycorrhizal basidiomycete, *Amanita muscaria*, suggest forest refugia in Alaska during the last glacial maximum. In: Relict Species: Phylogeography and Conservation Biology. Habel, J.C.; Assmann, T. (Eds.) *Springer-Verlag*, Berlin Heidelberg, pp. 173-186.

REFERENCES

- Georgescu, S.E. Manea, M.A. & Costache, M. (2008) The genetic structure of indigenous Romanian Hucul horse breed inferred from microsatellite data. *Romanian Biotechnological Letters* 13: 4030-4036.
- Głazewska, I. Wysocka, A. Gralak, B. Prus, R. & Sell, J. (2007) A new view on dam lines in Polish Arabian horses based on mtDNA analysis. *Genetics Selection Evolution* 39: 609-619.
- Guo, S. Savolainen, P. Su, J. Zhang, Q. Qi, D. Zhou, J. Zhong, Y. Zhao, & X. Liu, J. (2006) Origin of mitochondrial DNA diversity of domestic yaks. *BMC Evolutionary Biology* 22(6): 73.
- Habel, J.C. Drees, C. Schmitt, T. & Assmann, T. (2010) Review: Refugial Areas and Postglacial Colonizations in the Western Palearctic. In: Habel J.C., Assmann T. (Eds) *Relict Species: Phylogeography and Conservation Biology*. Springer, Heidelberg, 189-198.
- Hallsson, J. H. Favor, J. Hodgkinson, C. Glaser, T. Lamoreux, M.L. Magnúsdóttir, R. Gunnasson, G.J. Sweet, H.O. Copeland, N.G. Jenkins, N.A. & Steingrímsson, E. (2000) Genomic, transcriptional and mutational analysis of the mouse microphthalmia locus. *Genetics* 155(1): 291-300.
- Hansen, A. Mitchell, D.L. Wiuf, C. Paniker, L. Brand, T.B. Binladen, J. Gilichinsky, D.A. Ronn, R. & Willerslev, E. (2006) Crosslinks rather than strand breaks determine access to ancient DNA sequences from frozen sediments. *Genetics* 173: 1175-1179.
- Harpending, H.C. (1994) Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology* 66(4): 591-600.
- Harris, D.R. (1996) *The Origins and Spread of Agriculture and Pastoralism in Eurasia* (UCL, London).
- Hartl, D.L. & Clark, A.G. (2007) *Principles of Population Genetics*. Sunderland, MA: Sinauer Associates.
- Herrmann, B. & Hummel, S. (1994) Page 280 in *Ancient DNA: recovery and analysis of genetic material from paleontological, archaeological, museum, medical, and forensic specimens*. Springer-Verlag.
- Higuchi, R. Bowman, B. Freiberger, M. Ryder, O.A. & Wilson, A.C. (1984) DNA sequences from the quagga, and extinct member of the horse family. *Nature* 312: 282-284.
- Hoekstra, H.E. (2006) Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity* 97: 222-234.
- Hoelzel, A.R. (2005) Ancient genomes. *Genome Biology*, 6, p. 239.
- Hofreiter, M. (2008) Palaeogenomics. *Série Palévol des Comptes Rendus de l'Académie des Sciences* (invited review) 7: 113-124.
- Hofreiter, M. Jaenicke, V. Serre, D. v. Haeseler, A. & Pääbo, S. (2001) DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Research* 29(23): 4793-4799.
- Hofreiter, M. Serre, D. Poinar, H.N., Kuch, M. and Pääbo, S. (2001) Ancient DNA. *Nature Reviews Genetics* 2(5): 353-359.
- Hopwood, A.T. (1936) The former distribution of caballine and zebline horses in Europe and Asia. *Proc. Zool. Soc. London* 1936: 897-912.
- Hudson, R.R. (1990) Gene genealogies and the coalescent process. In *Oxford surveys in evolutionary biology* (eds. D.J. Futuyma and J. Antonovics), pp. 1-44. Oxford University Press, Oxford, UK.
- Igea, J. Juste, J. & Castresana, J. (2010) Novel intron markers to study the phylogeny of closely related mammalian species. *BMC Evolutionary Biology* 10: 369.
- Ishida, Y. David, V.A. Eizirik, E. Schäffer, A.A. Neelam, B.A. Roelke, M.E. Hannah, S. S., O'Bbrien, S.J. & Menotti-Raymond, M. (2006) A homozygous single-base deletion in MLPH causes the dilute coat color phenotype in the domestic cat. *Genomics* 88: 698-705.
- Jansen, T. Forster, P. Levine, M.A. Oelke, H. Hurler, 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 USA* 99: 10905-10910.
- Jobling, M.A. & Tyler-Smith, C. (2003) The human Y chromosome: an evolutionary marker comes of age. *Nature Reviews Genetic* 4: 598-612.
- Kakoi, H. Tozaki, T. Gawahara, H. (2007) Molecular analysis using mitochondrial DNA and microsatellites to infer the formation process of Japanese native horse populations. *Biochemical Genetics* (3-4): 375-95.
- Kantanen, J. Edwards, C.J. Bradley, D.G. Viinalass, H. Thessler, S. Ivanova, Z. Kiselyova, T. Kinkulov, M. Popov, R. Stojanovik, S. Ammosov, I. & Vilkki, J. (2009) Maternal and paternal genealogy of Eurasian taurine cattle (*Bos taurus*). *Heredity* 103,404-415.
- Kawaguchi, N. Ono, T. Mochii, M. & Noda, M. (2001) Spontaneous mutation in Mitf gene causes osteopetrosis in silver homozygote quail. *Developmental Dynamics* 220: 133-140.
- Keeler, C. Mellinger, T. Fromm, E. Wade, L. (1970) Melanin, adrenalin and the legacy of fear. *Journal of Heredity* 61: 81-88.
- Kelekna, P. (2009) The Politico-Economic Impact of the Horse on Old World Cultures: An Overview, SINO-PLATONIC PAPERS, Number 190.
- Kimura, B. Marshall, F.B. Chen, S. Rosenbom, S. Moehlman, P.D. Tuross, N. Sabin, R.C. Peters, J. Barich, B. Yohannes, H. Kebede, F. Teclai, R. Beja-Pereira, A. & Mulligan C.J. (2010) Ancient DNA from Nubian and Somali wild ass provides insights into donkey ancestry and domestication. *Proceedings of the Royal Society B* 278(1702): 50-7.
- Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature* 217 (5129): 624-626.
- Kimura, M. (1983) *The Neutral Theory of Molecular Evolution*. Cambridge University Press.
- Kruska, D. (2005) On the Evolutionary Significance of Encephalization in Some Eutherian Mammals: Effects of Adaptive Radiation, Domestication, and Feralization Brain. *Behavior and Evolution* 65 (2), 73-108.
- Lalueza-Fox, C. Römpler, H. Caramelli, D. Stäubert, C. Catalano, G. Hughes, D. Rohland, N. Pilli, E. Longo, L. Condemi, S. de la Rasilla, M. Fortea, J. Rosas, A. Stoneking, M. Schöneberg, T. Bertranpetit, J. & Hofreiter, M. (2007) A melanocortin 1 receptor variant suggests light skin and red hair in some Neanderthals. *Science* 318 (5855): 1453-5.
- Larson, G. Albarella, U. Dobney, K. Rowley-Conwy, P. Schibler, J. Tresset, A. Vigne, J. Edwards, C. Schlumbaum, A. Dinu, A. Balacsescu, A. Dolman, G. Tagliacozzo, A. Manaseryan, N. Miracle, P. Van Wijngaarden-Bakker, L. Masseti, M. Bradley, D. & Cooper, A. (2007). From the Cover: Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *Proceedings of the National Academy of Sciences of the USA* 104(39): 15276-15281.
- Larson, G. Dobney, K. Albarella, U. Fang, M. Matisoo-Smith, E. Robins, J. Lowden, S. Rowley-Conwy, P. Andersson, L. & Cooper A. (2005) Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* 307: 1618-1621.
- 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 of the USA* 107(17): 7686-91.
- 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. (2007) Patterns of East Asian pig domestication, migration, and turnover revealed by modern and ancient DNA. *Proceedings of the National Academy of Sciences of the USA* 107(17): 7686-7691.
- Lau, A.N. Peng, L. Goto, H. Chemnick, L. Ryder, O.A. & Makova, K.D. (2009) Horse domestication and conservation genetics of Przewalski's horse inferred from sex chromosomal and autosomal sequences. *Molecular Biology and Evolution* 26: 199-208.
- Laukner, A. (1998) Die Fellfarbe beim Hund. Teil 1: Grundlagen der Farbgenese. *Tierärztliche Praxis* 26(K): 49-54.
- Lei, C.Z. Chen, H. Zhang, H.C. Cai, X. Liu, R.Y. Luo, L.Y. Wang, C.F. Zhang, R.F. Lan, X.Y. & Sun, W.B. (2006) Origin and phylogeographical structure of Chinese cattle. *Animal Genetics* 37: 579-582.
- Lei, C.Z. Su, R. Xie, W.M. Bower, M.A. Edwards, C.J. Weining, S. Wang, X.B. Li, F. Liu, R.Y. Zhang, Y.S. Zhang, C.M. Chen, H. & Liu, L. (2009) Multiple maternal origins of living native horse and ancient horse populations in China. *Animal Genetics* 40,933-944.
- Levine, M.A. (1999) The origin of the horse husbandry on the Eurasian Steppe. In: *Late Prehistoric Exploitation of the Eurasian Steppe*. (Ed. by M.A. Levine, J. Rassarrerin, A.M. Kislenco, N.S. Tatarintreva), pp. 5-58. McDonald Institute, Cambridge, UK.
- Li, M.H. Zerabruk, M. Vangen, O. Olsaker, I. & Kantanen, J. (2007) Reduced genetic structure of north Ethiopian cattle revealed by Y-chromosome analysis. *Heredity* 98: 214-221.
- Ling, Y. Ma, Y. Guan, W. Cheng, Y. Wang, Y. Han, J. Jin, D. Mang, L. & Mahmut, H. (2010) Identification of Y chromosome genetic variations in Chinese indigenous horse breeds. *Journal of Heredity* 101: 639-643.
- Lira, J. Linderholm, A. Olaria, C. Brandström, M.D. Gilbert, M.T.P. Ellegren, H. Willerslev, E. Liden, 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.
- Liu, X.Y. Hunt, H.V. & Jones, M.K. (2009) River valleys and foothills: Changing archaeological perceptions of North China's earliest farms. *Antiquity* 83:82-95.
- Liu, Y.P. Wu, G.S. Yao, Y.G. Miao, Y.W. Luikart, G. Baig, M. Beja-Pereira, A. Ding, Z.L. Palanichamy M.G. & Zhang, Y.P. (2006) Multiple maternal origins of chickens: out of the Asian jungles. *Molecular Phylogenetics and Evolution* 38:112-119.
- Lorenzen, E.D. Arctander, P. & Siegmund, H.R. (2008) High variation and very low differentiation in wide ranging plains zebra (*Equus quagga*): insights from mtDNA and microsatellites. *Molecular Ecology* 17(12): 2812-2824.

REFERENCES

- LPP, LIFE Network, IUCN–WISP & FAO (2010) Adding value to livestock diversity – marketing to promote local breeds and improve livelihoods. *FAO Animal Production and Health Paper* 168.
- Lu, H.Y., Zhang, J.-P., Liu, K.B., Wu, N.-Q., Li, Y.-M., Zhou, K.-S., Ye, M.-L., Zhang, T.-Y., Zhang, H.-J., Yang, X.-Y., Shen, L.-C., Xu, D. & Li, Q. (2009) Earliest domestication of common millet (*Panicum miliaceum*) in East Asia extended to 10,000 years ago. *Proceedings of the National Academy of Sciences of the USA* 106:7367-7372.
- Lui's, C., Bastos-silveira, C., Cothran, E.G. & Do Mar Oom, M. (2006) Iberian Origins of New World Horse Breeds. *Journal of Heredity* 97(2): 107-113.
- Luikart, G., England, P.R., Tallmon, D., Jordan, S. & Taberlet, P. (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics* 4: 981-994.
- Lynch, M., Koskella, B. & Schaack, S. (2006) Mutation pressure and the evolution of organelle genomic architecture. *Scienc*. 311: 1727-1730.
- Mannen H., Morimoto ML, Oyamat K, Mukai F, & Tsuji, S. (2003) Identification of mitochondrial DNA substitutions related to meat quality in Japanese Black cattle. *Journal of Animal Science* 81: 68-73.
- Mannen, H., Kojima, T., Oyama, K., Mukai, F., Ishida, T. & Tsuji S. (1998) Effect of mitochondrial DNA variation on carcass traits of Japanese Black cattle. *Journal of Dairy Science* 76: 36-41.
- Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research* 27: 209-220.
- Mathias, E., Mundy, P. & Köhler-Rollefson, I. (2010) Marketing products from local livestock breeds: an analysis of eight cases. *Animal Genetic Resources* 47: 59-71.
- Matsushima, Y., Shinkai, Y., Kobayashi, Y., Sakamoto, M., Kunieda, T. & Tachibana, M. (2002) A mouse model of Waardenburg syndrome type 4 with a new spontaneous mutation of the endothelin-B receptor gene. *Mammalian Genome* 13: 30-35.
- McCallion, A.S. & Chakravarti, A. (2001) EDNRB/EDN3 and Hirschsprung disease type II. *Pigment Cell Research* 14: 161-169.
- McGahern, A.M., Edwards, C.J., Bower, M.A., Heffernan, A., Park, S.D., Brophy, P.O., Bradley, D.G., MacHugh, D.E. & Hill, E.W. (2006a) Mitochondrial DNA sequence diversity in extant Irish horse populations and in ancient horses. *Animal Genetics* 37(5):498-502.
- McGahern, A.M., Edwards, C.J., Bower, M.A., Heffernan, A., Park, S.D.E., Brophy, P.O., Bradley, D.G., MacHugh, D.E. & Hill, E.W. (2006b) Mitochondrial DNA sequence diversity in Irish extant populations and ancient horses. *Animal Genetics* 37: 498-502.
- McGrew, P. (1944) An early Pleistocene (Blancan) fauna from Nebraska. *Geol. Ser., Field Mus. Nat. Hist.* 9(2): 33-66.
- Metallinos, D.L., Bowling, A.T. & Rine, J. (1998) A missense mutation in the endothelin-B receptor gene is associated with lethal white foal syndrome: an equine version of Hirschsprung disease. *Mammalian Genome* 9(6): 426-431.
- Minvielle, F., Bed'Hom, B., Coville, J.-L., Ito, S., Inoue-Murayama, M. & Gourichon, D. (2010) The "silver" Japanese quail and the MITF gene: causal mutation, associated traits and homology with the "blue" chicken plumage. *BMC Genetics* 11: 15.
- Montoliu, L., Oetting, W.S. & Bennett, D.C. (2010) Color Genes. European Society for Pigment Cell Research, World Wide Web (URL: <http://www.espcr.org/micemut>, 1st July).
- Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. & Erlich, H. (1986) Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor Symposia on Quantitative Biology* 51: 263-273.
- Naderi, S., Rezaei, H.R., Taberlet, P., Zundel, S., Rafat, S.A., Naghash, H.R., Elbarody, 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.
- Navascués, M., Depaulis, F. & Emerson, B.C. (2010) Combining contemporary and ancient DNA in population genetic and phylogeographic studies. *Molecular Ecology Resources* 10: 760-772.
- Ning, T., Xiao, H., Li, J., Hua, S. & Zhang, Y.P. (2010) Adaptive evolution of the mitochondrial ND6 gene in the domestic horse. *Genetics and Molecular Research* 9: 144-150.
- Nordlund, J.J., Boissy, R.E., Hearing, V.J., Oetting, W., King, R.A. & Ortonne, J.-P. (2006) The pigmented system: physiology and pathophysiology. Wiley-Blackwell.
- Oakenfull, E.A. & Clegg, J.B. (1998) Phylogenetic relationships within the genus *Equus* and the evolution of the α and θ globin genes. *Journal of Molecular Evolution* 47: 772-783.
- Ogawa, H., Tu, C.H., Kagamizono, H., Soki, K., Inoue, Y., Akatsuka, H., Nagata, S., Wada, T., Ikeya, M., Makimura, S., Uchida, K., Yamaguchi, R. & Otsuka, H. (1997) Clinical, morphologic, and biochemical characteristics of Chediak-Higashi syndrome in fifty-six Japanese black cattle. *American Journal of Veterinary Research* 58: 1221-1226.
- O'Rourke, D.H., Hayes, M.G. & Carlyle S.W. (2000) Spatial and temporal stability of mtDNA haplogroup frequencies in Native North America. *Human Biology* 72(1): 15-34.
- Outram, A.K., Stear, N.A., Bendrey, R., Olsen, S., Kasparov, A., Zaubert, V., Thorpe, N. & Evershed, R.P. (2009). The earliest horse harnessing and milking. *Science* 323: 1332-1335.
- Owen, D.I. (1991) The 'First Equestrian': An Ur III Glyptic Scene. *Acta Sumerologica* 13: 259-71.
- Pääbo, S. (1989) Ancient DNA: Extraction, Characterization, Molecular Cloning, and Enzymatic Amplification. *Proceedings of the National Academy of Sciences of the USA* 86: 1939-43.
- Pang, J.-F., Kluebsch, C., Zou, X.-J., Zhang, A.-B., Luo, L.-Y., Angleby, H., Ardalan, A., Ekström, C., Sköllermo, S., Lundeberg, J., Matsumura, S., Leitner, T., Zhang, Y.-P. & Savolainen, P. (2009) MTDNA Data Indicate a Single Origin for Dogs South of Yangtze River, Less Than 16,300 Years Ago, from Numerous Wolves. *Molecular Biology and Evolution* 26(12): 2849-2864.
- Pérez-Gutiérrez, L.M., De la Peña, A. & Arana, P. (2008) Genetic analysis of the Hispano-Breton heavy horse. *Animal Genetics* 39(5): 506-14.
- Pérez-Pardal, L., Ginja, C., Royo, L.J., Álvarez, I., Fernández, I., del Valle, A., Traoré, A., Ponce de León, F.A., Beja-Pereira, A., Penedo, M.C.T. & Goyache F. (2011) Genetic structure of the bovine Y-specific microsatellite UMN0103 reflects the genetic history of the species. *Animal Genetics* DOI: 10.1111/j.1365-2052.2011.02177.x
- Petit, R.J., Aguinagalde, I., de Beaulieu, J.-L., Bittkau, C., Brewer, S., Cheddadi, R., Ennos, R., Fineschi, S., Grivet, D., Lascoux, M., Mohanty, A., Müller-Starck, G., Demesure-Musch, B., Palmé, A., Martin, J.P., Rendell, S. & Vendramin, G.G. (2003) Glacial Refugia: Hotspots But Not Melting Pots of Genetic Diversity. *Science* 300(5625): 1563-1565.
- Philipp, U., Hamann, H., Mecklenburg, L., Nishino, S., Mignot, E., Günzel-Apel, A.R., Schmutz, S.M. & Leeb, T. (2005) Polymorphisms within the canine MLPH gene are associated with diluted coat color in dogs. *BMC Genetics* 6: 34.
- Pielberg, G.R., Golovko, A., Sundström, E., Curik, I., Lennartsson, J., Seltenthammer, M., H. Druml, T., Binns, M., Fitzsimmons, C., Lindgren, G., Sandberg, K., Baumung, R., Vetterlein, M., Strömberg, S., Grabherr, M., Wade, C., Lindblad-Toh, K., Ponten, F., Heldin, C., Sölkner, J. & Andersson, L. (2008). A cis-acting regulatory mutation causes premature hair graying and susceptibility to melanoma in the horse. *Nature Genetics* 40: 1004-1009.
- Pilot, M., Branicki, W., Jędrzejewski, W., Goszczyński, J., Jędrzejewska, B., Dykyy, I., Shkvryva, M. & Tsingarska, E. (2010) Phylogeographic history of grey wolves in Europe. *BMC Evolutionary Biology* 10:104 doi:10.1186/1471-2148-10-104.
- Prasolova, L.A., Os'kina, I.N., Plysnina, I.Z. & Trut, L.N. (2009) Maternal methyl supplements affect the phenotypic variation of the agouti gene in the offspring of rats with different behavioral types. *Russian Journal of Genetics* 45: 587-592.
- Price, E.O. & King, J.A. (1968) Domestication and adaptation. In: Haféz, E.S.E. (Ed.), *Adaptation of Domestic Animals*. Lea and Febiger, Philadelphia, pp. 34-45.
- Priskin, K., Szabó, K., Tömöry, G., Bogácsi-Szabó, E., Csányi, B., Fördög, R., Downes, C.S. & Raskó, I. (2010) Mitochondrial sequence variation in ancient horses from the Carpathian Basin and possible modern relatives. *Genetica* 138(2): 211-8.
- Randi, E. (2008) Detecting hybridization between wild species and their domesticated relatives. *Molecular Ecology* 17: 285-293.
- Rieder, S. (2009) Molecular tests for coat colours in horses. *Journal of Animal Breeding and Genetics* 126(6): 415-24.
- Rogers, A. & Harpending, H. (1992) Population growth makes waves in the distribution of pairwise differences. *Molecular Biology and Evolution* 9:552-569.
- Römpler, H., Dear, P.H., Krause, J., Meyer, M., Rohland, N., Schöneberg, T., Spriggs, H., Stiller, M. & Hofreiter, M. (2006b) Multiplex amplification of ancient DNA. *Nature Protocol* 1: 720-728.
- Römpler, H., Rohland, N., Lalueza-Fox, C., Willerslev, E., Kuznetsova, T., Rabeder, G., Bertranpetit, J., Schöneberg, T. & Hofreiter, M. (2006a) Nuclear gene indicates coat-color polymorphism in mammoths. *Science* 313: 62.
- Roots, C. (2007) *Domestication*. Greenwood Guides to the Animal World. Greenwood Pub Group. pp134.
- Roulin, A., Ducrest, A.L. (2011) Association between melanism, physiology and behaviour: A role for the melanocortin system. *European Journal of Pharmacology* [Epub ahead of print].

REFERENCES

- 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(6): 663-669.
- Samraus, H.H. (1986) Endangered domestic animal species in German speaking countries. I: General aspects. *Tierärztliche Praxis* 14: 297-307.
- Santchi, E.M. Purdy, A.K. Valberg, S.J. Vrotsos, P.D. Kaese, H. & Mickelson, N.R. (1998) Endothelin receptor B polymorphism associated with lethal white foal syndrome in horses. *Mammalian Genome* 9: 306-309.
- Savolainen, P. Zhang, Y.P. Luo, J. Lundeberg, J. & Leitner, T. (2002) Genetic evidence for an East Asian origin of domestic dogs. *Science* 298: 1610-1613.
- Schaffner, S.F. Foo, C. Gabriel, S. Reich, D. Daly, M.J. & Altshuler, D. (2005) Calibrating a Coalescent Simulation of Human Genome Sequence Variation. *Genome Research*, 15(11): 1576-83.
- Schena, M. (1999) DNA Microarrays: A Practical Approach (Practical Approach Series, 205) (Oxford).
- Schmutz, S.M. Berryere, T.G. (2007) Genes affecting coat colour and pattern in domestic dogs: a review. *Animal Genetics* 38(6): 539-49.
- Schouwey, K. & Beermann, F. (2008). The notch pathway: Hair graying and pigment cell homeostasis. *Histology and Histopathology* 23(5): 609-619.
- Schutz, M.M. Freeman, A.E. Lindberg, G.L. Koehler, C.M. & Beitz, D.C. (1994) The effect of mitochondrial DNA on milk production and health of dairy cattle. *Livestock Production Science* 37: 283-295.
- Seco-Morais, J. Oom, M.M. Quesada, F. & Matheson, C.D. (2007) Ancient Iberian horses: a method to recover DNA from archaeological samples buried under sub-optimal conditions for preservation. *Journal of Archaeological Science* 34: 1713-1719.
- Shibanuma, K. Tong, Z-B. Vanderhoof, V.H. Vanevski, K. & Nelson, L.N. (2002) Investigation of KIT gene mutations in women with 46 XX spontaneous premature ovarian failure. *BioMed Central Womens Health* 2: 8.
- Silver, D. Hou, L. & Pavan, W.J. (2006) The genetic regulation of pigment cell development; Saint-Jeanet, editor. Georgetown: Landes Biosciences. pp. 155-169.
- Slatkin, M. & Hudson, R.R. (1991) Pairwise comparison of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 129: 555-562.
- Smith, C.H. (1866) The natural history of horses, with Memoir of Gesner. in: Ders., Mammalia. Horses, London 17-58.
- Sponenberg, P. (2009) Equine Color Genetics, 3rd Edition. Wiley-Blackwell 296 pp.
- Stoneking, M. (2000) Hypervariable sites in the mtDNA control region are mutational hotspots. *American Journal of Human Genetics* 67: 1029-1032.
- Stritzel, S. Wöhlke, A. & Distel, O. (2009) A role of the microphthalmia-associated transcription factor in congenital sensorineural deafness and eye pigmentation in Dalmatian dogs. *Journal of Animal Breeding and Genetics* 126: 59-62.
- Sutarno, C.G. Cummins, J.M. Greeff, J. & Lymbery, A.J. (2002) Mitochondrial DNA polymorphisms and fertility in beef cattle. *Theriogenology* 57: 1603-1610.
- Tapio, M. Marzanov, M. Ozerov, M. Cinkulov, M. Gonzarenko, G. Kiselyova, T. Murawski, M. Viinalass, H. & Kantanen, J. (2006) Sheep mitochondrial DNA variation in European, Caucasian and Central Asian areas. *Molecular Biology and Evolution* 23:1776-1783.
- Tennessen, J.A. (2008) Positive selection drives a correlation between non-synonymous/synonymous divergence and functional divergence. *Bioinformatics* 24: 1421-1425.
- Trut, L.N. (1999) Early canid domestication: the farm-fox experiment. *American Scientist* 87: 160-169.
- Trut, L.N. Ruvinsky, A. & Sampson, J. (2001) The genetics of the dog, experimental studies in early canid domestication. eds Ruvinsky, A., Sampson, J., CABI Publishing, New York, pp 15-41.
- Uerpmann, H-P. (1987) The ancient distribution of ungulate mammals of the Middle East. Beihefte zum Tübinger Atlas des Vorderen Orients. Reihe A 9 (Naturwissenschaften). L. Reichert, Wiesbaden, 27: 1-173.
- Uerpmann, H-P. (2005) Domestication of the Horse - When, Where, and Why?. in: Bodson, Liliane (ed.): Le cheval et les autres équidés: aspects de l'histoire de leur insertion dans les activités humaines, Colloques d'histoire des connaissances zoologiques 6: 15-29.
- Vaez, M. Follet, S.A. Bed'hom, B. Gourichon, D. Tixier-Boichard, M. Burke, T. (2008) A single point-mutation within the melanophilin gene causes the lavender plumage colour dilution phenotype in the chicken. *BMC Genetics* 9: 7.
- Van Dyke, F. (2003) Conservation Biology: Foundations, Concepts, Applications. McGraw-Hill, New York.
- Van Gele, M. Dynodt, P. & Lambert, J. (2009) Griscelli syndrome: a model system to study vesicular trafficking. *Pigment Cell Melanoma Research* 22: 268-282.
- Varga, Z. (2010) Extra-Mediterranean Refugia, Post-Glacial Vegetation History and Area Dynamics in Eastern Central Europe. Relict Species, Part I, 57-87.
- Vilà, C. Leonard, J.A. Götherstrom, A. Marklund, S. Sandberg, K. Liden, K. Wayne, R.K. & Ellergren, H. (2001) Widespread origins of domestic horse lineages. *Science* 291: 474-477.
- Vilà, C. Savolainen, P. Maldonado J.E. Amorim, I.R. Rice, J.E. Honeycutt, R.L. Crandall, K.A. Lundeberg, J. & Wayne R.K. (1997) Multiple and ancient origins of the domestic dog. *Science* 276(5319):1687-1689.
- Vilà, C. Savolainen, P. Maldonado, J.E. Amorim, I.R. Rice, J.E. Honeycutt, R.L. Crandall, K.A. Lundeberg, J. & Wayne, R.K. (1997) Multiple and ancient origins of the domestic dog. *Science* 276: 1687-1689.
- Visscher, P.M. Hill, W.G. & Wray, N.R. (2008) Heritability in the genomics era--concepts and misconceptions. *Nature Reviews Genetics* 9(4): 255-66.
- Volf, J. Kus, E. & Prokopova, L. (1991) General studbook of the Przewalski horse. Prague (Czech Republic): Zoological Garden Prague.
- Wallner, B. Brem, G. Müller, M. & Achmann, R. (2003) Fixed nucleotide differences on the Y chromosome indicate clear divergence between Equus przewalskii and Equus caballus. *Animal Genetics* 34: 453-456.
- Wallner, B. Piumi, F. Brem, G. Müller, 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(2):158-164.
- Walsh, B. (2008) Using molecular markers for detecting domestication, improvement, and adaptation genes. *Euphytica* 161: 1-17.
- Warmuth, V. Eriksson, A. Bower, M.A. Cañon, J. Cothran, G. Distl, O. Glowatzki-Mullis, M-L. Hunt, H. Luis, C. do Mar Oom, M. Yupanqui, I.T. Ząbek, T & Manica, A. (2011) European Domestic Horses Originated in Two Holocene Refugia. *PLoS ONE* 6(3): e18194.
- Weatherford, J. (2004) Genghis Khan and the Making of the Modern World. New York: Crown.
- Weinstock, J. Willerslev, E. Sher, A. Tong, W. Ho, S.Y.W. Rubenstein, D. Storer, J. Burns, J. Martin, L. Bravi, C. Prieto, A. Froese, D. Scott, E. Xulong, L. & Cooper, A. (2005) Evolution, systematics, and phylogeography of Pleistocene horses in the New World: a molecular perspective. *PLoS Biology* 3(8): 0001-0007.
- Weir, B.S. & Cockerham, C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38(6): 1358-1370.
- Welle, M. Philipp, U. Rüfenacht, S. Roosje, P. Scharfenstein, M. Schütz, E. Brenig, B. Linek, M. Mecklenburg, L. Grest, P. Drögemüller, M. Haase, B. Leeb, T. Drögemüller, C. (2009) MLPD Genotype-Melanin Phenotype Correlation in Dilute Dogs. *Journal of Heredity* 100: S75-S79.
- Westbroek, W. Adams, D. Huizing, M. Koshoffer, A. Dorward, H. Tinloy, B. Parkes, J. Helip-Wooley, A. Kleta, R. Tsilou, E. Duvernay, P. Digre, K.B. Creel, D.J. White, J.G. Boissy, R.E. & Gahl, W.A. (2007) Cellular defects in Chediak-Higashi syndrome correlate with the molecular genotype and clinical phenotype. *Journal of Investigative Dermatology* 127: 2674-2677.
- White, D.J. Wolff, J.N. Pierson, M. & Gemmill, N.J. (2008) Revealing the hidden complexities of mtDNA inheritance. *Molecular Ecology* 17: 4925-4942.
- Yang, D.Y. Eng, B. & Saunders, S.R. (2003) Hypersensitive PCR, ancient human mtDNA and contamination. *Human Biology* 75: 355-364.
- Yang, G.C. Croacker, D. Zhang, A.L. Manglick, P. Cartmill, T. & Cass, D. (1998) A dinucleotide mutation in the endothelin-B receptor gene is associated with lethal white foal syndrome (LWFS), a horse variant of Hirschsprung disease. *Human Molecular Genetics* 7: 1047-1052.
- Yang, Y.H. Kim, K.I. Cothran, E.G. & Flannery, A.R. (2002) Genetic diversity of Cheju horses (Equus caballus) determined by using mitochondrial DNA D-loop polymorphism. *Biochemical Genetics* 40(5-6): 175-86.
- Yang, Z. (2006) Computational molecular evolution. UK: Oxford University Press.
- Yindee, M. (2010) Y-chromosomal variation confirms independent domestications of swamp and river buffalo. *Animal Genetics* 41(4): 433-435.
- Zeder, M.A. Emshwiller, E. Bradley, D.G. & Smith, B.D. (2006) Documenting domestication: The intersection of genetics and archaeology. *Trends in Genetics* 22(3):139-155.
- Zhu, M.J. Li, X. & Zhao, S.H. (2010) Digital candidate gene approach (DigiCGA) for identification of cancer genes. *Methods Molecular Biology* 653: 105-29.
- Zuckerkind, E. & Pauling, L.B. (1962) "Molecular disease, evolution, and genetic heterogeneity". In Kasha, M. and Pullman, B (editors). Horizons in Biochemistry. Academic Press, New York. pp. 189-225.

7. APPENDIX

PAPER I

TITLE:

Coat color variation at the beginning of horse domestication

AUTHORS:

Arne Ludwig, Melanie Pruvost, Monika Reissmann, Norbert Benecke, Gudrun A. Brockmann, Pedro Castaños, Michael Cieslak, Sebastian Lippold, Laura Llorente, Anna-Sapfo Malaspinas, Montgomery Slatkin & Michael Hofreiter

JOURNAL:

American Association for the Advancement of Science © 2011
Received: 25 February 2009; Accepted: 1 April 2009; Published: 24 April 2009

LINK (join AAAS to view full text):

<http://dx.doi.org/10.1126/science.1172750>

CITATION:

*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. **American Association for the Advancement of Science** 324, 485.*

PAPER II

TITLE:

Origin and History of Mitochondrial DNA Lineages in Domestic Horses

AUTHORS:

Michael Cieslak, Melanie Pruvost, Norbert Benecke, Michael Hofreiter, Arturo Morales, Monika Reissmann, Arne Ludwig

JOURNAL:

Public Library of Science ONE (PLoS ONE)

Received: August 18, 2010; Accepted: November 8, 2010; Published: December 20, 2010

LINK:

<http://dx.doi.org/10.1371/journal.pone.0015311>

CITATION:

*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. **Public Library of Science ONE** 5(12), e15311.*

Origin and History of Mitochondrial DNA Lineages in Domestic Horses

Michael Cieslak¹, Melanie Pruvost^{1,2*}, Norbert Benecke², Michael Hofreiter^{3,4}, Arturo Morales⁵, Monika Reissmann⁶, Arne Ludwig^{1*}

1 Leibniz Institute for Zoo and Wildlife Research, Evolutionary Genetics, Berlin, Germany, **2** German Archaeological Institute, Berlin, Germany, **3** Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, **4** Department of Biology, The University of York, Heslington, York, United Kingdom, **5** Laboratory of Archaeozoology, Universidad Autonoma Madrid, Madrid, Spain, **6** Department of Breeding Biology and Molecular Genetics, Humboldt University Berlin, Berlin, Germany

Abstract

Domestic horses represent a genetic paradox: although they have the greatest number of maternal lineages (mtDNA) of all domestic species, their paternal lineages are extremely homogeneous on the Y-chromosome. In order to address their huge mtDNA variation and the origin and history of maternal lineages in domestic horses, we analyzed 1961 partial d-loop sequences from 207 ancient remains and 1754 modern horses. The sample set ranged from Alaska and North East Siberia to the Iberian Peninsula and from the Late Pleistocene to modern times. We found a panmictic Late Pleistocene horse population ranging from Alaska to the Pyrenees. Later, during the Early Holocene and the Copper Age, more or less separated sub-populations are indicated for the Eurasian steppe region and Iberia. Our data suggest multiple domestications and introgressions of females especially during the Iron Age. Although all Eurasian regions contributed to the genetic pedigree of modern breeds, most haplotypes had their roots in Eastern Europe and Siberia. We found 87 ancient haplotypes (Pleistocene to Mediaeval Times); 56 of these haplotypes were also observed in domestic horses, although thus far only 39 haplotypes have been confirmed to survive in modern breeds. Thus, at least seventeen haplotypes of early domestic horses have become extinct during the last 5,500 years. It is concluded that the large diversity of mtDNA lineages is not a product of animal breeding but, in fact, represents ancestral variability.

Citation: Cieslak M, Pruvost M, Benecke N, Hofreiter M, Morales A, et al. (2010) Origin and History of Mitochondrial DNA Lineages in Domestic Horses. PLoS ONE 5(12): e15311. doi:10.1371/journal.pone.0015311

Editor: Manfred Kayser, Erasmus University Medical Center, The Netherlands

Received: August 18, 2010; **Accepted:** November 8, 2010; **Published:** December 20, 2010

Copyright: © 2010 Cieslak et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was funded by grants provided by the Deutsche Forschungsgemeinschaft (LU852-7/2-3) and by the Max Planck Society. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: ludwig@izw-berlin.de (AL); melpruvost@googlemail.com (MP)

Introduction

The domestication of animals and plants was a key innovation in the emergence of modern urban societies. Deciphering the spatial and temporal origin of the domestic species seems crucial for understanding the cultural changes of civilizations and has consequently attracted considerable scientific as well as public interest [1,2]. Among the early domesticated animals, two species of major importance differ from all others by having been bred not mainly as food sources: the dog and the horse. Whereas the dog may arguably have been the first animal to be domesticated [3,4 – commonly used for hunting, guard and military actions], horses attained a prominent role as animals of transport and warfare, changing societies on a continent-wide scale. As M. Levine so aptly remarked: “*Before the development of firearms, the horse was crucial to warfare and before the invention of the steam engine, it was the fastest and most reliable form of land transport*” [5]. Despite the pivotal role horses have played in the development of human societies, the origin and history of modern breeds remain unclear in several aspects. For example, it is sometimes difficult to set apart the remains of wild and domestic horses. Contrary to dogs or pigs, the plasticity of the skeleton of the genus *Equus* is almost nil. As a result, even species as different as zebras and asses are almost indistinguishable based on their bones. Structural stability explains the problems that

paleontologists encounter when attempting to work out the evolutionary history of horses [6]. To complicate matters further, the domestication of the horse did not result in any relevant changes in body size, as was the case with the remaining domesticates. This lack of diagnostic anatomical and biometrical criteria lies at the root of assigning status to archaeozoological remains [5]. It was for this reason that faunal analysts welcomed the introduction of molecular analyses some years ago.

Due to the high copy number of mtDNA in cells, their comparatively fast molecular evolution and their strictly maternal inheritance, over the last two decades mtDNA has become a valuable tool for phylogenetic and phylogeographic studies using ancient DNA analyses. However, both the lack of ancient sequences from pre-domestic and early domestic horses as well as the absence of a phylogeographic structure in modern breeds have prevented the establishment of any firm conclusions thus far. Previous studies have demonstrated the difficulty of restricting mitochondrial lineages to a particular geographic area or specific breed. To further complicate matters, the wild ancestor of domestic horses seems to be extinct since the agriotypic status of the few surviving Przewalski's horses is under debate [7]. Moreover, and in contrast with many other domesticated species [8–11], whereas an impressive amount of variability was discovered in modern horses in the mitochondrial control region

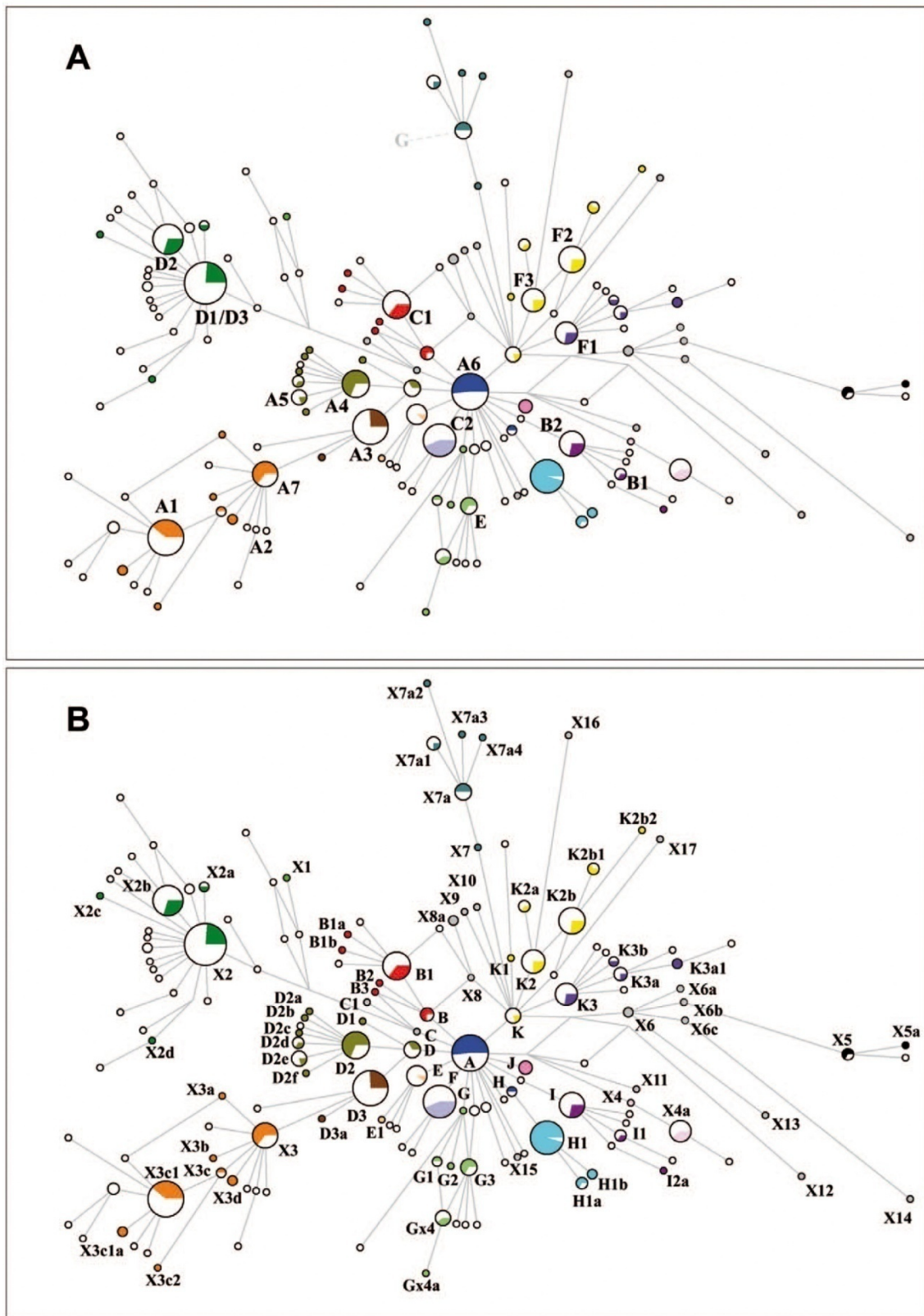


Figure 1. Median-joining network based on 247 bp of the mitochondrial D-Loop from 207 ancient samples (colored) and 601 modern sequences of primitive breeds (white). Eighty-seven haplotypes were found among the ancient samples; 103 different haplotypes among the primitive breeds and 36 haplotypes are shared (Table S5). Figure 1A shows haplotypes named after [13–14], whereas Figure 1B shows the new nomenclature (Table S4). Haplogroups are signed by a color code.
doi:10.1371/journal.pone.0015311.g001

[12–15], the species exhibits extremely low nucleotide diversity on the Y-chromosome. Recently the first description of two different Y-chromosomal haplotypes in Chinese indigenous horse breeds was reported [16]. Consequently, a sex bias has been assumed for the founders of domestic horses. It is likely that only a few stallions but many females were domesticated [17].

Even though several recent studies focused on the mtDNA variability of horses from restricted geographic regions and time periods [15,18], some important issues remain to be solved, e.g. whether the large mtDNA variability was also present in the pre-domestic population and how this variation was distributed geographically. In this study, we attempt to shed new light on these issues by combining a large set of pre-domestic, early domestic and modern horse mtDNA sequences.

Results

Haplogroups and haplotypes

We successfully sequenced DNA from 85 out of 157 ancient samples resulting in a success rate of 54% (Table S1), i.e. 6 North East Siberian, 6 East Asian, 24 West and South Siberian, 6 Armenian, and 22 European samples (Iberia excluded), as well as 2 samples from Asia Minor and 19 Iberian remains. All sequences were deposited in Genbank (see Table S2 for accession numbers). In an attempt to improve the genetic information and the resolution of horse phylogeography we amplified the complete control region (D-loop) of the mitochondrial DNA (722 bp from 15,468–16,190 nps). But in contrast to other domestic species [19], we found that sequence length has no influence on phylogenetic resolution in horses and we decided to focus our analyses on the HVR1 region including the majority of Genbank sequences (247 bp from 15,494–15,740 nps). After their incorporation in our data set (85 sequences) of 122 previously published ancient sequences, the alignment of the 207 ancient sequences produced 78 variable sites (4 of which are indels) defining 113 haplotypes. Using all sequences (1754 modern horses and 207 ancient samples), we identified 4 strong mutational hotspots, 3 which were described in previous studies [13,18] at nucleotide positions 15585, 15597 and 15650 and a new hotspot at position 15604. These positions were not included in the phylogenetic calculations. In addition, a few more mutational hotspots were individually down-weighted (Table S3). Once these mutational hotspots were eliminated from the dataset, the number of haplotypes fell to 87. Therefore, we introduced a new nomenclature for the haplotypes/groups (Figure 1 A, B; Table S4).

Strikingly, the inclusion of all the sequences available for modern horses ($n = 1754$) and ancient samples ($n = 207$) led to an unstructured and unresolved network. For this reason, the network (Figure 1B) was calculated focusing on 599 sequences from 46 primitive breeds, two Przewalski haplotypes and the 207 ancient samples. We divided the sequences into 19 haplogroups from A to K and from X1 to X7 based on the structure of the network (Figure 1B).

Geographical origin of the mitochondrial lineages

Table 1 and Figure 2 give an overview of the geographic origin of mitochondrial lineages.

North and East Siberia. Only 2 out of the 12 Late Pleistocene lineages identified from 14 individuals are still present in modern breeds. Among these, basal haplotype A was recorded from the Late Pleistocene up to Mediaeval Times in several regions of Eurasia whereas K1 was only found in modern types. Only X8a was found in both Alaska and Siberia in the Late Pleistocene, although X8a is extinct today. X2d was only found in Iron Age wild horses but never in domestic horses, but haplogroup X2 is one of the most common haplogroups in modern breeds.

East Asia (China, Mongolia, and Korea). No pre-domestic horse samples were available for this area. Haplotype D3a was found five times in Bronze Age samples. Haplotype K3 first occurred in China in the Bronze Age. Seven out of the 20 East Asian Iron Age lineages ($n = 36$) were only observed in this region, whereas 13 lineages were also present in other regions and times.

West and South Siberia, Kazakhstan. As was the case for East Asia, no data were available from pre-domestication times in this region. X3 appeared first in West Siberia during the Copper Age and, later, in China during the Bronze Age. X5 was confined to this region. The other eight Bronze Age lineages were also found in other regions. Six out of 18 Iron Age lineages ($n = 26$) were found only in this region.

Europe and Asia Minor (westwards from the Ural Mountains and the Caspian Sea to the Pyrenees, including Armenia). We profited from the larger number of samples available for this region dating from the Late Pleistocene to Mediaeval Times. Interestingly, we found haplotype B1 before and after domestication in this region. Moreover, we found that 4 out of 7 lineages which were present among the Late Pleistocene horses ($n = 9$ samples), 6 out of 8 pre-domestic lineages ($n = 12$), 4 out of 11 Bronze Age lineages ($n = 12$), 5 out of 8 Iron Age lineages ($n = 8$) and 9 out of 18 Mediaeval lineages ($n = 27$) were confined to this region. Furthermore, 3 Late Pleistocene lineages, 2 pre-domestic lineages, 8 Bronze Age lineages, 3 Iron Age lineages, and 9 Mediaeval lineages were also found in other regions and times.

Iberian Peninsula. Three pre-domestic lineages (B, H1, J) ($n = 13$) were confined to this region. They were present during the Copper Age and Early Bronze Age. Haplotype B was also found in Iron Age horses. The dominating haplogroup during the Copper Age and Bronze Age was H1. However, the three Copper/Bronze Age haplotypes A, D2 and X4a were also found in other regions. Likewise, both Mediaeval lineages (X2, X2b) were found throughout Eurasia, probably due to the exchange of domestic animals.

Primitive horse breeds (Table S5). The haplogroup distribution of primitive breeds in Eurasia is shown in Figure 3. We found 39 out of 87 ancient lineages in our sample set of modern breeds (Table S6). Eleven out of these 39 haplotypes were lineages that were confined to a single primitive breed (B/Arabian; D2d/Cheju; G1/Akhal Teke; H/Garrano; H1/Marismeno; H1a/Lusitano; K2b1/Sicilian Oriental Purebred; K3b/ Yakut; X1/Pottoka; X2a/Debao; X3c/Lusitano; X5/Fulani). The haplotypes D, D2, E, K2 and X7a are rare or non-existent in Europe and Asia Minor; however, they are rather common in mainland Asia (Figure 3). Similarly, the A haplotype was not present in North West Europe or on the Iberian Peninsula. It is conspicuous that the K3 haplogroup is mainly present in Mongolia, Yakutia, Korea, Africa and Hungary. In contrast, haplotypes D3 and F are mainly

Table 1. Haplotype distribution.

		LATE PLEISTOCENE			COPPER AGE			BRONZE AGE			IRON AGE			MEDIÉVAL			
	<i>n</i>	Alaska	North East Siberia	Europe	Europe Asia Minor	Iberia	China	West Siberia	Europe ARM	Iberia	North East Siberia	China	West South Siberia KZ	Europe Iberia	Korea	Iberia	Europe (no Iberia)
LATE PLEISTOCENE	8	B2, X6, X6a-c, X14															
	14	X8a	B3, C, G2, K1, X3d, X8, X10-12, X15														
	9		A	B1a, C1, X3c2, H													
COPPER AGE	12			B1	D2b, D2f, G1, X4, X9, X17												
	13					B, H1, J											
BRONZE AGE	5						K3, D3a										
	9				F		X3	X5									
	12			D	F		X7a		D, D1, X3c1a, X5a								
	23		A			B, H1, J			D2	B, H1, H1a							
IRON AGE	1										X2d						
	36		A	F		K3	D2d, G3, K2, X3c1		D2, X4a, Gx4a, I, X2b								E, K2a, K2b1, K2b2, K3a1, X7a3, X7a4

Table 1. Cont.

	LATE PLEISTOCENE			COPPER AGE			BRONZE AGE			IRON AGE			MEDIEVAL						
	<i>n</i>	Alaska	North East Siberia	Europe	Minor/ARM	Europe Asia	Iberia	China	West Siberia	Europe ARM	Iberia	North East Siberia	China	West South Siberia	Europe	Iberia	Korea	Iberia	Europe (no Iberia)
West South Siberia, Kazakhstan	26		A	B1					D3, X3c1, X5, X7a	D2, I, K2b, X2b	X4a	X2	E1, K, K3a, K3b, X3a, X3c						
Europe	8			B1				D3				X2	B1b, D2e, X1, X7a2, X13						
Iberia	1						B						B						
Korea	1															G			
Iberia	2											X2					X2, X2b		
Europe (no Iberia)	27		A	B1	F			D3, X3c1	I, K2b, X2b			X2							D2a, D2c, I1, X2a, X2c, X3b, X7, X7a1, X16

Diagonal elements: private haplotypes; below diagonal: shared haplotypes. Grey-color code shows the frequency of shared haplotypes.
 doi:10.1371/journal.pone.0015311.t001

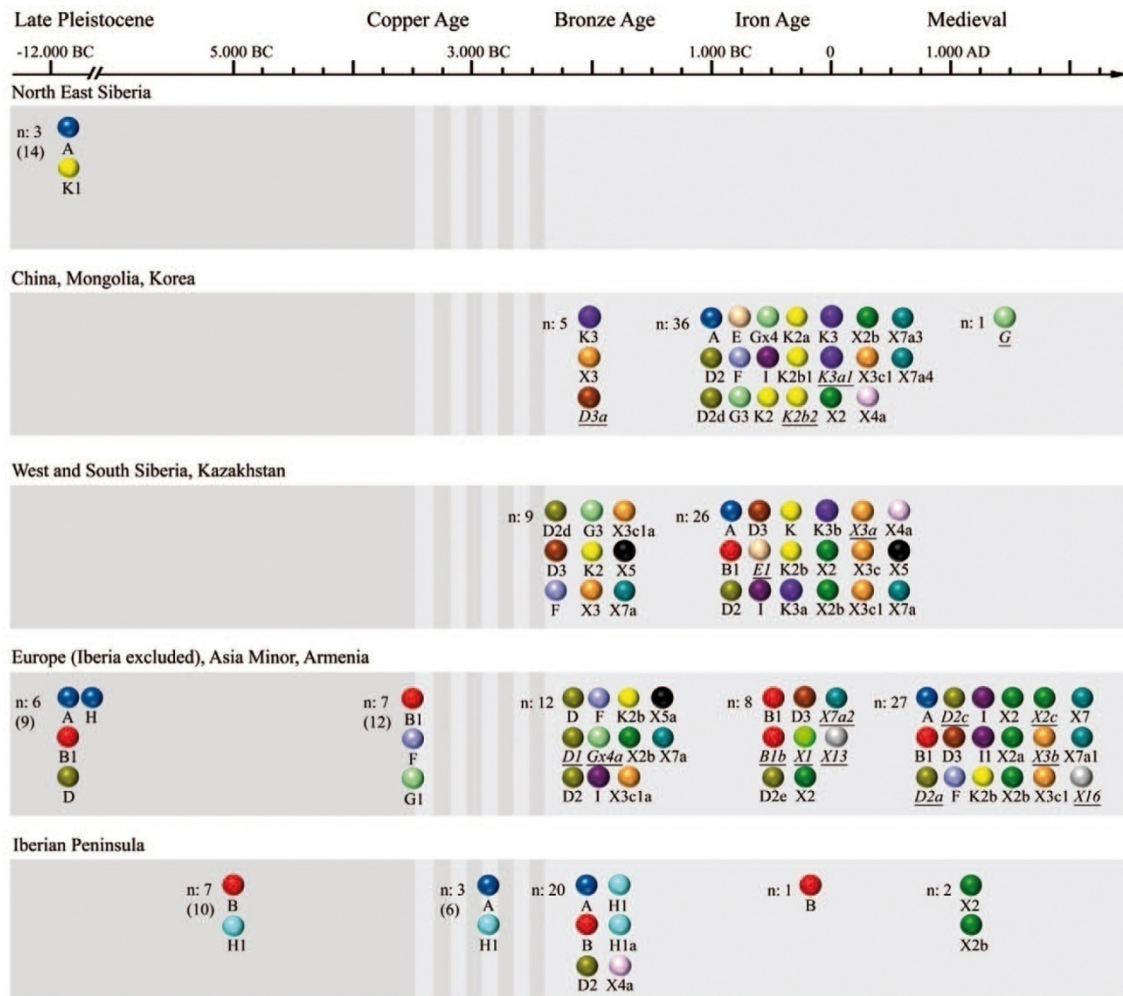


Figure 2. Origin of mitochondrial lineages of modern horses. Only mitochondrial lineages are shown which were/are present in domestic horses. Haplotypes in italic/underlined were so far only found in early-domestic remains. The striped zone indicates the postulated period of horse domestication.
 doi:10.1371/journal.pone.0015311.g002

present in Europe/Asia Minor, whereas B1 is particularly common in the Nordic region of Europe, and haplotype I appears especially on the Iberian Peninsula, Africa and Central Asia. K2b, X2, X2b, X3 and X3c1 are frequent and common haplotypes, and are more or less equally distributed across Eurasia.

Presence of ancient haplotypes in modern horses

Thirty-nine of the 87 ancient haplotypes were also found in modern horses (Table S5). Today, X2 is the most common ancient haplotype in modern horses. X2 was found 326 times out of 1754 modern horse sequences. The second most common haplotype D3 was found 133 times followed by X2b (103), X3c1 (98), I (85), F (80), B1 (79), and A (75). The following 48 ancient lineages became extinct: B1a, B1b, B2, B3, C, C1, D1, D2a, D2b, D2c, D2f, D3a, E1, G, G2, Gx4a, H1b, I2a, J, K2b2, K3a1, X1, X2d, X2c, X3a, X3b, X3c1a, X3c2, X3d, X4, X5a, X6, X6a, X6b, X6c, X7, X7a2, X8, X8a, X9–12, X13, X14–15, X16, X17; italicized haplotypes were present in early domestic samples, but are extinct in modern horses today.

Demography

The mismatch distributions of the West/South Siberian and European sample sets from early-domestic times show clear unimodal curves (Figure 4, Table S7). In contrast, the curves of the West/South Siberian sample sets were marked by a weak bimodal peak and the East Asian sample set from early domestic times has a ragged, bimodal or trimodal character. Finally, the curves of the Iberian pre-domestic and Copper/Bronze Age population, and the European samples of pre-domestic times were extremely ragged.

Genetic variability

Nucleotide diversity varied between 0.034 (Europe 1000 BC–600 AD) and 0.005 (Iberia 5500–3000 BC), whereas haplotype diversity varied between 1.00 (West Siberia 3000–1000 BC; Europe 1000 BC–600 AD) and 0.60 (Iberia 5500–3000 BC). Altogether, the Iberian sets had the lowest values of both nucleotide and haplotype diversity (Table 2). The average

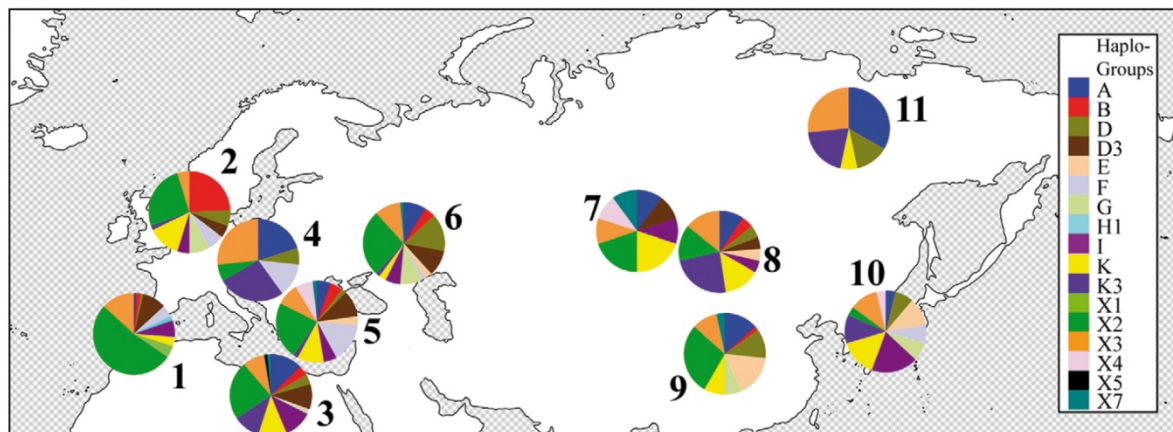


Figure 3. Haplogroup distribution of Eurasian primitive breeds (n = 601 animals; Table S13). (1: Asturcon (n = 12), Cartujano (n = 8), Garrano (n = 5), Losino (n = 10), Lusitano (n = 10), Marismeno (n = 12), Merens (n = 10), Pottoka (n = 13), Sorraia (n = 27); 2: Duelmener (n = 10), Exmoor (n = 17), Gotland (n = 3), Icelandic (n = 6), Norwegian Fjord (n = 11), Shetland (n = 18), Welsh (n = 1); 3: Arabian (n = 99), Barb (n = 37), Fulani (n = 9), Egyptian (n = 7); 4: Hucul (n = 11), Koonik (n = 5); 5: Anatolian (n = 17), Giara (n = 2), Mallorquina (n = 2), Pindos (n = 7), Sanfrantellano (n = 10), Sicilian Oriental (n = 1), Sicilian Ind. (n = 13), Skyros (n = 5); 6: Akhal Teke (n = 37), Caspian (n = 13), Vyatskaya (n = 18); 7: Tuva (n = 10); 8: Mongolian (n = 20), Przewalski (n = 2); 9: Debao (n = 24), Guan Mountains (n = 10), Guanzhong (n = 2), Wenshan (n = 2), XiNeHe (n = 1), Dali (n = 6), Yunnan (n = 1), Tibetan (n = 16); 10: Cheju (n = 25), Taishu (n = 2); 11: Yakut (n = 15)).
 doi:10.1371/journal.pone.0015311.g003

nucleotide diversity of wild horses (Late Pleistocene and pre-domestic samples) was about 25% lower than the average of the domestic horses (Iberia was not included; for details see Table 2). A sub-division in the following three pre-domestic sub-populations (Table S8, Table S9) was supported by their F_{st} values ranging between 0.4 and 0.1: Alaskan horses, Iberian horses and Eurasian steppe horses. Nevertheless, some haplotypes were shared, indicating a common ancestry from a panmictic Late Pleistocene population or at least substantial gene flow between the regions. Furthermore, F_{st} values and haplotype composition did not support differentiation between horses from the Late Pleistocene and the Copper Age in Europe, as recently described based on shifts in their coat color [20], although, due to a lack of samples, this issue could not be addressed further. However, some lineages (e.g. E, E1) appeared later during the Bronze Age and Iron Age; others increased in frequency (e.g. X2, X3, D3), and still others became extinct soon after the glaciations (e.g. C, J, X8) (Figure 5).

Discussion

A recent study pinpointed horse domestication in North Kazakhstan to around 3500 BC [21]. As demonstrated by our data, a large amount of genetic diversity was already present in pre-domestic horses before that time. Under normal circumstances, the bottleneck caused by domestication should have significantly reduced that genetic variation. As indicated by our data and previous studies [12,13,20], this has only partly been the case for horses. Altogether, 87 haplotypes were found in the archaeological remains (Pleistocene to Mediaeval Times). Overall, 32% (i.e., 9 out of 28 haplotypes) of the pre-domestic lineages (Pleistocene to Copper Age) have survived in the gene pool of domestic horses up to now. Seven of these pre-domestic lineages were also found in domestic horses from the Bronze Age and Iron Age and the remaining two were recorded in samples from the Mediaeval Ages. Considering all 87 of the ancient haplotypes, 29 haplotypes were found in Bronze Age samples, 47 haplotypes in Iron Age remains, and 21 haplotypes in Mediaeval bones. Our data provide evidence

for multiple introgressions or, alternatively, for a second wave of domestication during the Iron Age, given that new 22 haplotypes were discovered in the gene pool of domestic horses for this period. In addition, ten haplotypes have their appearance during Mediaeval Times. However, we would like to raise a note of caution regarding the fact that, given the huge variation of mtDNA in horses, more ancient samples are necessary to determine the first occurrence of haplotypes in space and time. As is the case for all ancient DNA studies, we have a sample bias favoring modern sequences. Public databases contain 274 archived haplotypes of modern horses (19th July 2010). Thus far, apparently only 45% of the mtDNA variation found in our study has survived in the modern domestic horses (i.e., 39 out of 87 haplotypes), and in terms of domestic samples between the Bronze Age and today, 30% (17 out of 56) of the maternal lineages have been lost during 5,500 years of horse breeding. This is still a large amount of genetic variation for a domestic animal. The reasons for the presence of such a large amount of genetic variation in the modern domestic horse may include one or a combination of the following: i) multiple origins; ii) an extremely large number of female founders and iii) large-scale introgression of local lineages into the domestic stock. In our opinion, it is not unlikely that all three of these reasons contributed evenly to dictate the present day state of affairs. Taken together, one major outcome of this study is that the huge diversity of the horse's mtDNA is not a consequence of animal breeding but instead a feature that was already present in the wild horse populations.

Our dataset has a bias of pre-domestic samples from Asia. For this reason one has to be careful when proposing putative centers of domestication. Notwithstanding, our results postulating local introgressions of wild mares and multiple domestication episodes in Europe and East Asia are not speculative. Both of these regions contributed significantly to determine the genetic portfolio of modern breeds during the Bronze and Iron Ages. Likewise, the mismatch distribution of the early-domestic European and Siberian horses produced evidence for a bottleneck 1500 years after the postulated origin of domestic horses in North Kazakhstan

Demographic Expansion

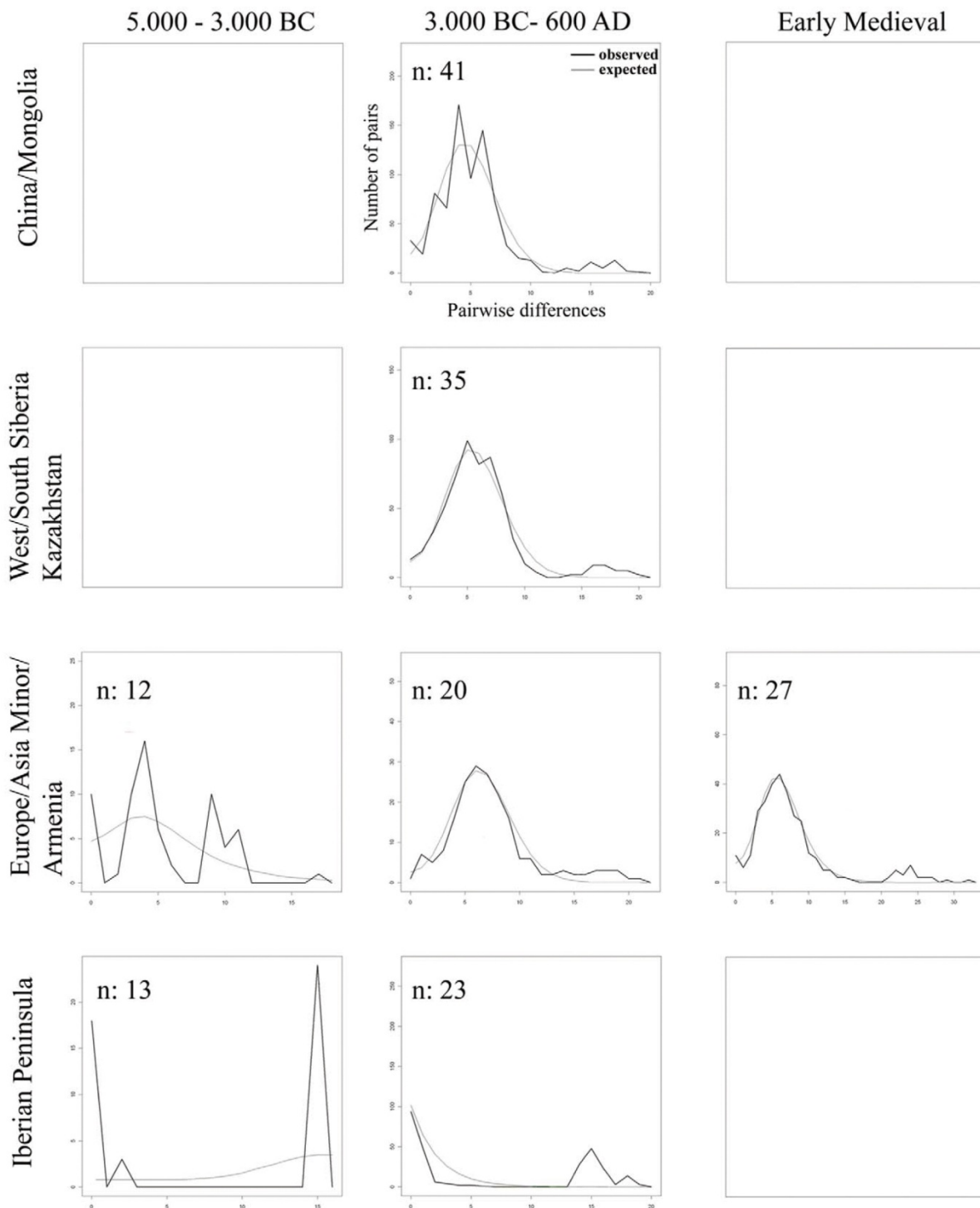


Figure 4. Mismatch distribution of mitochondrial haplotypes indicating population expansion (domestication) for all regions excluding Iberia during 3000 BC–600 AD.
doi:10.1371/journal.pone.0015311.g004

Table 2. Nucleotide- and haplotype-diversity.

		<i>n</i>	No. of haplotypes	Haplotype diversity	No. of polymorphic sites	Nucleotide diversity
LATE PLEISTOCENE	<i>Alaska</i>	8	7	0.964+/-0.077	13	0.018+/-0.011
	<i>North East Siberia</i>	14	12	0.978+/-0.035	17	0.017+/-0.010
	<i>Europe</i>	9	7	0.917+/-0.092	11	0.013+/-0.008
COPPER AGE	<i>Europe/ Asia Minor Armenia</i>	12	8	0.849+/-0.104	18	0.016+/-0.010
	<i>Iberia</i>	10	3	0.600+/-0.131	4	0.005+/-0.004
BRONZE AGE	<i>China</i>	5	3	0.700+/-0.218	7	0.014+/-0.010
	<i>West Siberia</i>	9	9	1.000+/-0.052	17	0.021+/-0.013
	<i>Europe Armenia</i>	12	11	0.985+/-0.040	27	0.023+/-0.014
	<i>Iberia</i>	38	18	0.819+/-0.062	35	0.012+/-0.007
IRON AGE	<i>China</i>	35	20	0.952+/-0.019	31	0.021+/-0.011
	<i>West South Siberia, Kazakhstan</i>	26	18	0.969+/-0.018	30	0.025+/-0.014
	<i>Europe</i>	8	8	1.000+/-0.063	22	0.034+/-0.020
MEDIEVAL	<i>Europe</i>	26	19	0.966+/-0.022	32	0.027+/-0.015

doi:10.1371/journal.pone.0015311.t002

around 3500 BC [21]. Furthermore, our data on Bronze and Iron Age horses indicate local domestication and/or introgression of East Asian haplotypes. The evolution of these private East Asian haplotypes was enhanced by the isolation of Mongolian and Chinese wild horses due to the Altai-Mountains and the Takla Makan and Gobi deserts. These geographical barriers effectively segregated them from their cousins in the Eurasian steppe.

Our data have produced evidence for a panmictic horse population ranging from the Siberian Lowlands to the Pyrenees during the Late Pleistocene. The widespread distribution of haplogroups B1 (Europe, North East Siberia and Alaska), X8 (North East Siberia and Alaska), and X3 (North East Siberia and Ireland) as well as that of haplotype A (Germany, Ukraine and North East Siberia) lends support to the idea of widespread gene-flow across Eurasia. Later, in the Early Holocene and during the Copper Age, we find evidence for a genetic sub-division of Iberian and Eurasian steppe horses.

In Europe, we found three pre-domestic lineages (out of 8) that have survived in modern horses (Table S6). Two of them, B1 are still frequently present in modern European breeds, whereas G1 is now rare. In contrast, haplotype E and haplogroup K3 had their roots in East Asia, and both are common in East Asian primitive breeds today.

Separated from the rest of Europe by the Pyrenees, the history of the Iberian horses has to be considered independently. The Pyrenees effectively cut off the Iberian Peninsula from the rest of Eurasia and turned it into a glacial refuge. Pre-domestic to Copper Age Iberian horses were revealed as geographically isolated by our data. Interestingly, the haplotypes of the H1-group that dominated the ancient Iberian lineages during pre-domestic and early-domestic times are still occasionally present in Iberian breeds (e.g. Marismefio, Lusitano, Caballo de Corro) and also in the South American breeds (Argentinean Creole, Puerto Rican Paso Fino) today [18]. Remarkably, the ancient and singular Iberian haplotype B is nowadays found in Percheron (France), Arabian and Wielkopolski (Poland) horses. Such presence in the Percheron horses could be an indication of trade between Iberian and French people in ancient times, whereas the presence in the Arabian breed is likely a relic from the Moorish occupation of South Iberia. The Wielkopolski horse is a breed that also contains Arabian genes. We therefore agree with previous studies [18,20] about the

lack of evidence for an independent horse domestication process in Iberia, at least not on a significant scale, and provide instead strong evidence for introgression of local mares. Nowadays, all of the primitive Iberian breeds are dominated by haplotypes of the X3- and X2-haplogroups.

Considering our data from the Bronze Age horses, it appears that domestic horses spread soon after their initial domestication. Horse-riding was probably the catalyst for such fast spreading, starting an unprecedented process of gene flow 5,500 years ago. After 5,500 years of horse husbandry, the world is populated by 58 million domestic horses [FAO Statistics Division 2010 for 2008; <http://faostat.fao.org>]. The precise number of breeds is impossible to determine, but ranges from 600 up to 1000 breeds. Some of them are primitive breeds known since ancient times (e.g. Akhal Teke, Arab, Barb); others were founded a few decades ago (e.g. Lewitzer, Nederlands Mini Paarden, Pony of the Americas, Rocky Mountain Horse). However, many of the so-called "modern breeds" have their roots at the end of the eighteenth century or later [22]. Although inbreeding is sometimes used in animal husbandry for breed improvement, breed formation is often an open process. Crosses between different breeds or introgression of foreign specimens as well as introgression of wild horses have been reported [23]. For these reasons, mitochondrial lineage diversity changed over time by breeding or hybridization and introgression. As a result, a breed does not necessarily have a straightforward history and only rarely is it genetically isolated from the rest of the remaining population. All these processes have left their footprints in the genetic pedigree of the domestic horses. In fact, no other domestic animal today is less structured phylogenetically and phylogeographically speaking, or has an equivalent amount of mtDNA variation.

Materials and Methods

Samples

Our samples range from Alaska/North East Siberia to the Iberian Peninsula and cover a temporal window from the Late Pleistocene to Mediaeval Times (Table S1; ¹⁴C data are listed in Table S10). We successfully analyzed a total of 85 ancient Eurasian samples (out of 157). The sequence dataset obtained from those samples that yielded ancient DNA was supplemented by 122 ancient DNA sequences archived in GenBank (Table S2). In addition, 1754 GenBank

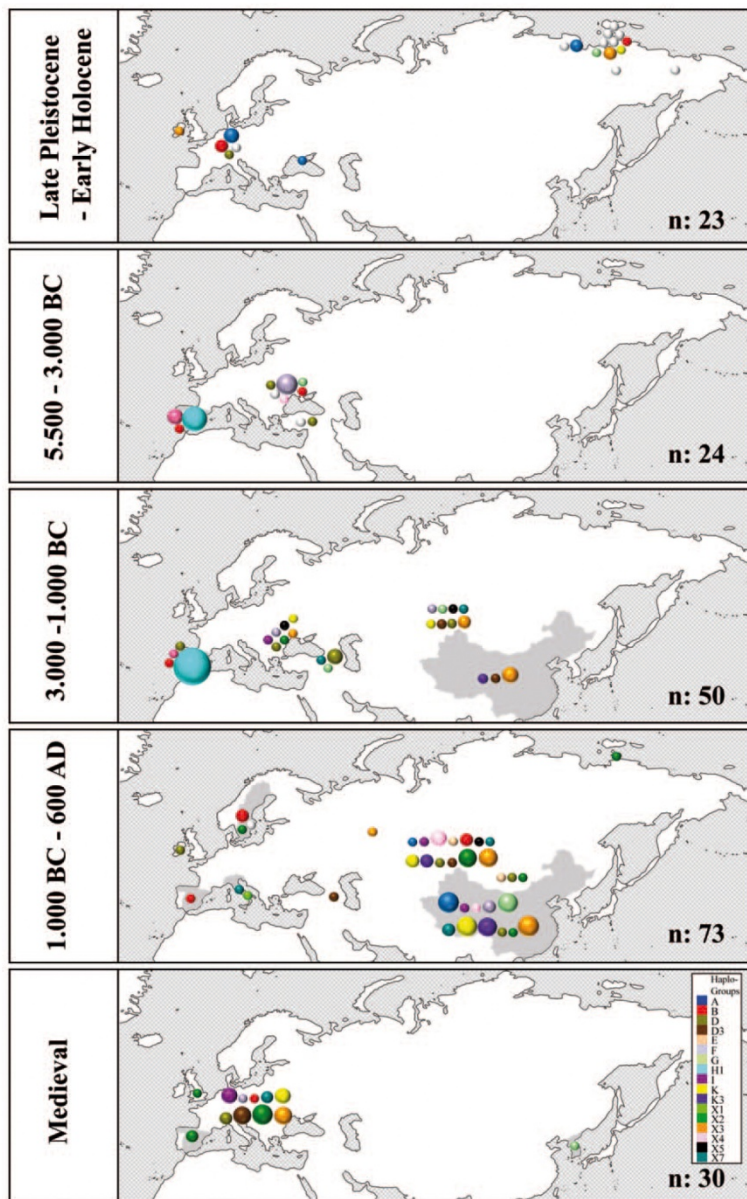


Figure 5. This figure illustrates the chronology of geographic haplogroup distribution. Size of circles represents haplogroup frequencies. doi:10.1371/journal.pone.0015311.g005

sequences were incorporated from modern horses including the Przewalski horse (Table S11). We divided our dataset based on the geographic origin of the samples into: i) North East Siberia; ii) East Asia (e.g. China, Mongolia and Korea); iii) West and South Siberia (West Siberian Plain and Kazakhstan); iv) Europe and Asia Minor (westwards from the Ural Mountains and the Caspian Sea to the Pyrenees and Armenia); and v) Iberian Peninsula (south of the Pyrenees). Additionally, we considered the following time (cultural) periods: i) Late Pleistocene/Early Holocene; ii) Copper Age; iii) Bronze Age; iv) Iron Age; v) Early Mediaeval. It should be stressed that the range of dates of the chrono-cultural periods occasionally differs for each of the geographical regions.

Ancient DNA extraction, amplification and sequencing

DNA was extracted from 250–400 mg samples of bone or tooth. The surfaces of these tissues were removed by abrasion to minimize contamination. Samples were ground to powder with a freezer mill and incubated in 0.45 M EDTA (pH 8.0) and 0.25 mg/ml Proteinase K overnight under rotation at room temperature. Afterwards, the pellets were collected by centrifugation for 5 min at 4,000 rpm in a Universal 320 centrifuge (Hettich). DNA was purified from the supernatant using a silica-based method as previously described [24–25].

Two different overlapping sets of primer were designed (Table S12). Using these primers, 722 bp (15.468–16.190 nps; primer set

1) or 713 bp (15,468–16,181 nps; primer set 2) were respectively amplified in a two-step multiplex PCR [26–27]. Overlapping PCR products including primers varied in length between 178 bp and 108 bp. We used 4 µl of extract in each multiplex PCR. The initial PCR was performed in a 20 µl reaction volume containing 1× AmpliTaq Gold PCR buffer II (ABI), 4 mM MgCl₂, 1 mg/ml Bovine Serum Albumin (BSA), 250 µM each of dATP, dCTP and dGTP, 500 µM of dUTP, 150 nM of each primer and 2 U of AmpliTaq Gold (ABI). PCR was performed by adding 1 U of heat-labile Uracil-DNA Glycosylase (USB) and an initial incubation step of 15 min at 37°C to control for carry-over contamination. We used 5 µl of the diluted (1/30) PCR products for the next step (total reaction volume 20 µl). Singleplex PCRs contained 1× AmpliTaq Gold PCR Buffer II, 4 mM MgCl₂, 1 mg/ml bovine serum albumin (BSA), 250 µM each of dATP, dCTP and dGTP, 500 µM of dUTP, 1.5 µM of each primer and 0.5 U of AmpliTaq Gold DNA polymerase. In both cases, PCR was run under the following conditions: denaturation and Taq activation at 94°C for 9 min followed by 30 to 35 cycles consisting of denaturation at 95°C for 20 sec, annealing temperature depending on the primer set (set I: 58°C; set II: 47°C) for 30 sec, elongation at 72°C for 30 sec and final extension at 72°C for 4 min. Negative extraction controls and negative PCR controls were performed for each PCR. Amplification products were visualized on agarose gel. Fragments were sequenced on an ABI PRISM 3730 capillary sequencer using the BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems).

Authentication

DNA sampling, extractions and pre-PCR preparations were carried out in laboratories dedicated to ancient DNA following the standard procedures to avoid contamination. Direct sequencing was performed at least three times for each fragment from two independent extractions to detect PCR errors and miscoding of ancient DNA lesions from authentic sequences [28]. Independent replications were carried out in different laboratories for 21 bone samples (Humboldt University Berlin, Max Planck Institute Leipzig).

Phylogenetic reconstructions

Sequences were truncated to nucleotide positions 15,494–15,740 (part of the mitochondrial HVR I) to accommodate published short sequences. Ancient sequences of the control region were aligned using BioEdit 7.0.9 (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>) based on a reference sequence [29]. Haplotypes were determined on the basis of informative substitutions excluding mutational hotspots [13] which were detected in a median joining network containing 1754 modern horses and 207 ancient samples. NETWORK 4.5.1.0 (<http://www.fluxus-engineering.com>) was used for calculation of the median joining network. Calculations were performed under the following conditions: deletions or insertions were double weighted; default setting of Epsilon (0) was chosen; transition/transversion ratio (TI/TV = 6.5). Variable sites were down- or up-weighted according to the maximal number of mutations at each position (Table S3). In addition to the three mutational hotspots identified in previous studies [13,18], we excluded a fourth mutational hotspot. Thus, we defined new haplotypes and named them following a new nomenclature. Ancient haplotypes are named according to the letter of their haplogroup supplemented by a number or a number and a letter depending on how far they are from the basal node (for example: haplotype D3 is separated from D by 1 mutation, and haplotype D3a by 2 mutations). X indicates questionable haplogroups (X1–X8) and

haplotypes (X9–X17), respectively. The phylogenetic position of these haplotypes depends strongly on the defaults of network reconstruction.

mtDNA diversity and population structure

The genetic variability among ancient sequences was quantified as both nucleotide diversity and haplotype diversity. In addition, the average number of pairwise differences within and between populations and the population pairwise *F*_{st} values among populations (based on the Tamura & Nei distance; 10,000 permutations) were computed using Arlequin (<http://cmpg.unibe.ch/software/arlequin3/>). Demographic expansions (mismatch distribution) for certain time points and cultural horizons were estimated using Arlequin (demographic expansion) and figured using an Arlequin integrated R script (<http://www.r-project.org/>). The nucleotide diversity and mismatch distribution calculations were performed under the following setting: gamma shape parameter ($\alpha=0.57$) estimated in PAUP 4.0 (Sinauer Associates) using maximum likelihood.

Supporting Information

Table S1 Samples analyzed for this study. The samples marked by a Y (Typing) gave a complete genotype. The extraction and amplification of the samples (Ext/Amp) and the reproduction (Rep) were performed in two institutes by Melanie Pruvost (MP) and Michael Cieslak (MC) at the Humboldt University in Berlin and Sebastian Lippold (SL) and Melanie Pruvost (MP) at the Max Planck Institute in Leipzig. EBA = Early Bronze Age; MED = Medieval Times. (DOC)

Table S2 Samples analyzed for this study and samples from the Genbank. The samples highlighted in grey are the extended samples from the Genbank. The table includes the sample name, accession number, location, date and also an assigned haplogroup/haplotype name. (DOC)

Table S3 Mitochondrial DNA control region (HVRI; position: 15,494–15,740) and the nucleotide characteristic data from the sample set of table S2 with additionally 207 ancient samples (Table S11). (DOC)

Table S4 Variable positions of the d-loop sequences of ancient haplotypes. Nps less 15000 are shown. D2e corresponds to mtDNA reference, X79547 [23]. Mutational hotspots (15585, 15597, 15604 and 15650) are not included, because these positions were excluded in phylogenetic reconstructions. Gaps were signed by “-”. Previous nomenclature [13–14] is given in brackets. (DOC)

Table S5 Ancient haplotypes occurrence in primitive breeds. (DOC)

Table S6 Occurrence of ancient haplotypes in modern horses based on their Genbank entries. (DOC)

Table S7 Sum of squared deviation and raggedness index (mismatch distribution). (DOC)

Table S8 Population pairwise Fst values. Above diagonal: Matrix of significant Fst P values ($p=0.05$). Below diagonal: Population pairwise Fst values. (DOC)

Table S9 Average number of pairwise differences between and within populations. Above diagonal: Average number of pairwise differences between populations (PiXY); (grey = significant; $p=0.05$). Diagonal elements: Average number of pairwise differences within population (PiX). Below diagonal: Corrected average pairwise difference $(PiXY - (PiX + PiY)/2)$. (DOC)

Table S10 ^{14}C Dates of samples analyzed in this study. For details of samples see table 1. The Pleistocene samples are not directly dated, but estimated from context should be around 20,000 years old. The remaining dates are either calibrated carbon dates or derived from archaeological context. (DOC)

Table S11 GenBank Accession no of modern horses (Przewalski horses are bold). (DOC)

Table S12 Primer sequences. (DOC)

Table S13 Primitive horse breeds and the Przewalski horse. (DOC)

Acknowledgments

We are grateful to the numerous archaeologists and curators who provided material and information for our study. In particular, we thank Kurt Alt, Soner Aksu, Nikola Hensel, Christine Weber and Sebastian Lippold for the support prior to and during the course of the ancient DNA analyses. Dietmar Lieckfeldt is gratefully acknowledged for his technical assistance. This article results from the project of horse domestication, which was originally initiated by AL and NB in 2004, and is a part of MC's dissertation.

Author Contributions

Conceived and designed the experiments: MP NB MR MH AL. Performed the experiments: MC MP. Analyzed the data: MC MP AL. Contributed reagents/materials/analysis tools: NB AM MH. Wrote the paper: AL MH MP. Responsible for the lab work: MC MP. Provided samples and archaeological knowledge: AM NB. Responsible for research design and statistics: AL MP MH NB. Designed the figures: MC.

References

1. Anthony DW, Bogucki P, Comşa E, Gimbutas M, Jovanović B, et al. (1986) The "Kurgan Culture," Indo-European origins, and the domestication of the horse: A reconsideration. *Current Anthropology* 27(4): 291–313. Available: <http://www.jstor.org/pss/2743045>.
2. Levine M (1999) The origins of horse husbandry on the Eurasian steppe. In: Levine M, Rassamkin Y, Kislenco AM, Tatarintseva NS, eds. *Late Prehistoric Exploitation of the Eurasian Steppe*. Cambridge McDonald Institute. pp 5–58.
3. Vilà C, Savolainen P, Maldonado JE, Amorim IR, Rice JE, et al. (1997) Multiple and ancient origins of the domestic dog. *Science* 276: 1687–9. Available: <http://www.sciencemag.org/cgi/content/abstract/276/5319/1687>.
4. Savolainen P, Zhang Y, Luo J, Lundeberg J, Leitner T, et al. (2002) Genetic evidence for an East Asian origin of domestic dogs. *Science* 298: 1610. Available: <http://www.sciencemag.org/cgi/content/abstract/298/5598/1610>.
5. Levine MA (2005) Domestication and early history of the horse. In: Mills DS, McDonnell SM, eds. *The Domestication of horse: the origins, development and management of its behaviour*. Cambridge: Cambridge University Press. pp 5–22.
6. Mashkour M (2006) *Equids in Time and Space*. Oxford, Great Britain: Oxbow Books.
7. Lau AN, Peng L, Goto H, Chemnick L, Ryder OA, et al. (2009) Horse domestication and conservation genetics of Przewalski's horse inferred from sex chromosomal and autosomal sequences. *Mol Biol Evol* 26: 199–208. Available: <http://mbe.oxfordjournals.org/cgi/content/abstract/msn239>.
8. Giuffra E, Kijas JMH, Amarger V, Carlborg O, Jeon J-T, et al. (2000) The origin of the domestic pig: Independent domestication and subsequent introgression. *Genetics* 154: 1785–1791. Available: <http://www.genetics.org/cgi/content/short/154/4/1785>.
9. Edwards CJ, Bollongino R, Scheu A, Chamberlain A, Tresset A, et al. (2007) Mitochondrial DNA analysis shows a Near Eastern Neolithic origin for domestic cattle and no indication of domestication of European aurochs. *Proceedings Royal Society* 274: 1377–1385. Available: <http://rspb.royalsocietypublishing.org/content/274/1616/1377.full.pdf+html>.
10. Tapio M, Marzanov N, Ozerov M, Cinkulov M, Gonzarenko G, et al. (2006) Sheep mitochondrial DNA variation in European, Caucasian and Central Asian areas. *Mol Biol Evol* 23: 1776–1783. DOI 10.1093/molbev/msl043. Available: <http://mbe.oxfordjournals.org/cgi/content/short/msl043v1>.
11. Naderi S, Rezaei H-R, Taberlet P, Zundel S, Rafat S-A, et al. (2007) Large-scale mitochondrial DNA analysis of the domestic goat reveals six haplogroups with high diversity. *PLoS ONE* 2(10): e1012. doi:10.1371/journal.pone.0001012. Available: <http://www.plosone.org/article/info:doi%2F10.1371%2Fjournal.pone.0001012>.
12. Vilà C, Leonard JA, Gotherstrom A, Marklund S, Sandberg K, et al. (2001) Widespread origins of domestic horse lineages. *Science* 291: 474–477. Available: <http://www.sciencemag.org/cgi/content/abstract/291/5503/474>.
13. Jansen T, Forster P, Levine MA, Oelke H, Hurler M, et al. (2002) Mitochondrial DNA and the origins of the domestic horse. *Proceedings of the National Academy of Sciences of the United States of America* 99: 10905–10910. Available: <http://www.pnas.org/content/99/16/10905>.
14. McGahern AM, Bower MA, Edwards CJ, Brophy PO, Sulimova G, et al. (2006) Evidence for biogeographic patterning of mitochondrial DNA sequences in Eastern horse populations. *Animal Genetics* 37: 494–497. Available: <http://www3.interscience.wiley.com/journal/118596251/abstract>.
15. Lei CZ, Su R, Bower MA, Edwards CJ, Wang XB, et al. (2009) Multiple maternal origins of native modern and ancient horse populations in China. *Animal Genetics*. pp 933–944. Available: <http://www3.interscience.wiley.com/journal/122594301/abstract>.
16. Ling Y, Ma Y, Guan W, Cheng Y, Wang Y, et al. (2010) Identification of Y chromosome genetic variations in Chinese indigenous horse breeds. *Journal of Heredity* 101: 639–643. Available: <http://jhered.oxfordjournals.org/cgi/content/abstract/esq047>.
17. 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: 158–164. Available: <http://jhered.oxfordjournals.org/cgi/content/short/95/2/158>.
18. Lira J, Linderholm A, Olaria C, Brandström Durling M, Gilbert MTP, et al. (2010) Ancient DNA reveals traces of Iberian Neolithic and Bronze Age lineages in modern Iberian horses. *Molecular Ecology* 19: 64–78. Available: <http://www3.interscience.wiley.com/journal/123191525/abstract>.
19. Achilli A, Bonfiglio S, Olivieri A, Malusà A, Ajmone-Marsan P, et al. (2009) The multifaceted origin of taurine cattle reflected by the mitochondrial genome. *PLoS ONE* 4: e5753. Available: <http://www.plosone.org/article/info:doi%2F10.1371%2Fjournal.pone.0005753>.
20. Ludwig A, Pruvost M, Reissmann M, Benecke N, Brockmann GA, et al. (2009) Coat color variation at the beginning of horse domestication. *Science* 324: 485. Available: <http://www.sciencemag.org/cgi/content/short/324/5926/485>.
21. Outram AK, Stear NA, Bendrey R, Olsen S, Kasparov A, et al. (2009) The earliest horse harnessing and milking. *Science* 323: 1332–1335. Available: <http://www.sciencemag.org/cgi/content/abstract/323/5919/1332>.
22. Sambaugh HH (1986) Endangered domestic animal species in German speaking countries. I: General aspects. *Tierärztliche Praxis* 14: 297–307. Available: <http://www.ncbi.nlm.nih.gov/pubmed/3764877>.
23. Waring GH (2002) *Horse behaviour: The behavioural traits and adaptations of domestic and wild horses, including ponies*. Park Ridge, NJ: Noyes Publishing.
24. Rohland N, Hofreiter M (2007a) Ancient DNA extraction from bones and teeth. *Nature Protocols* 2(7): 1756–62. Available: http://www.natureprotocols.com/2007/07/12/ancient_dna_extraction_from_bo.php.
25. Rohland N, Hofreiter M (2007b) Comparison and optimization of ancient DNA extraction. *Biotechniques* 42: 343–352. Available: <http://www.biotechniques.com/BiotechniquesJournal/2007/March/Comparison-and-optimization-of-ancient-DNA-extraction/biotechniques-41414.html>.
26. Krause J, Dear PH, Pollack JL, Slatkin M, Barnes I, et al. (2006) Multiplex amplification of the mammoth mitochondrial genome and the evolution of Elephantidae. *Nature* 439: 724–727. Available: <http://www.nature.com/nature/journal/v439/n7077/abs/nature04432.html>.
27. Römpler H, Dear PH, Krause J, Meyer M, Rohland N, et al. (2006) Multiplex amplification of ancient DNA. *Nature Protocols* 1: 720–728.

PAPER III

TITLE:

Colours of domestication

AUTHORS:

Michael Cieslak, Monika Reissmann, Michael Hofreiter and Arne Ludwig

JOURNAL:

Biological Reviews © 2011 Cambridge Philosophical Society
*Received: 28 July 2010; Revised: 28 February 2011; Accepted: 4 March 2011; Published
online: 28 March 2011*

LINK (The definitive version is available at www3.interscience.wiley.com):

<http://dx.doi.org/10.1111/j.1469-185X.2011.00177.x>

CITATION:

Cieslak, M., Reissmann, M., Hofreiter, M. and Ludwig, A. (2011) Colours of domestication. Biological Reviews 86, 885-899.

PAPER IV

Michael Cieslak^{1*}, Monika Reißmann² and Arne Ludwig^{1*}

Examination of *melanophilin* as a candidate gene for the equine dun phenotype

Received: 21 October 2010; Revised: 01 November 2010

¹Leibniz Institute for Zoo and Wildlife Research, Evolutionary Genetics, 12561 Berlin, Germany; ²Humboldt-University Berlin, Faculty of Agriculture and Horticulture, Invalidenstraße 42, 10115 Berlin, Germany.

*Corresponding authors: Michael Cieslak, Leibniz Institute for Zoo and Wildlife Research, D-10115 Berlin, Germany, cieslak@izw-berlin.de; Arne Ludwig, Leibniz Institute for Zoo and Wildlife Research, D-10115 Berlin, Germany, Tel.: +49 30 5168312, ludwig@izw-berlin.de

Abstract

Background

During the last decade, the underlying molecular background was discovered for many coat color phenotypes in horses, but the genetics of the equine dun phenotype is still unknown. It is a widely distributed coat color phenotype in horse breeds probably resulting from its dominant mode of inheritance. Furthermore, there is evidence that dun was the original wild-type phenotype of the wild ancestor of the domestic horse. Here we investigate the 16 exons of the *melanophilin* gene as candidate gene for the equine dun phenotype.

Findings

We found seven substitutions (four of them were synonymous and the other three were non-synonymous) within the 16 exons of seven non-dun horses and nine dun horses including two Przewalski's horses. However, whether alone or in combination, none of these SNPs were ever linked to the dun phenotype.

Conclusions

Although the *melanophilin* gene is responsible for dilution phenotypes in human, mice, and other domesticated species, the detected SNPs of this investigation revealed no facts showing that there is any obvious association with the dun phenotype. Nevertheless, it could not be excluded that it is still involved in dun, because this research did not include further important sequence regions of the *melanophilin* gene like the UTR regions, the promoter region and the introns (potential splice site locations).

Findings

Background

The dun coloration is probably the wild phenotype in horses depending on its occurrence in Przewalski horses as well as in some primitive horse breeds (e.g. Exmoor Pony, Konik, Icelandic Horses, Norwegian Fjords, Sorraia). But the proof of this assumption is still open [1]. Dun horses are characterized by the dilution of their basic coat color, whereas the extent of dilution at the lower legs and head is less obvious. In addition, dun horses possess a dorsal stripe, while the presence of other patterns, especially darker stripes on the lower legs and the shoulders can vary [2; Figure 1]. Breeding has shown that the dun coloration is inherited as a dominant trait. Dn^+ has been suggested as the dominant allele, and Dn^{nd} has been suggested as the recessive allele [2]. Obviously, it is inherited independently of other known coat-color-associated genes and can occur in combination with any other genes that modify the basic coat colors (black, bay, chestnut) [2]. While responsible mutations of several equine phenotypes are known [e.g. 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14], the dun gene still has not been discovered.

Recently a DNA test (<http://www.vgl.ucdavis.edu/services/dunhorse.ph>) became available testing the zygosity of dun in a few breeds. However, this is not a direct DNA test, but this analysis includes information of DNA markers, and requires additional information of the coat color phenotype and the pedigree of three generations of the subject horse. Therefore, the aim of this study is to establish a direct genetic test for the dun phenotype. Such diagnostic tests are very important for breeders, because of breeding selection to preserve offspring with the desired coat colour phenotype. This can be vital for a successful breeding program, because certain coat colourations can substantially improve the marketability of an animal. In fact, in a linkage map analysis, the dun phenotype was mapped on the p-arm of ECA8 [15]. But until now, no promising candidate gene was discovered in this chromosome region [15].

	Exon 2	Exon 7	Exon 8	Exon 8	Exon 9	Exon 9	Exon 10
SNP substitution	c.177G>A	c.766A>G	c.918C>T	c.989G>T	c.1031G>A	c.1050C>A	c.1254T>G
AA substitution	no	p.N256D	no	p.G330V	p.R344H	no	no

Breed	Coat color						
Fjord	<i>Bay dun</i>		R		K		
Fjord	<i>Bay dun</i>						
Fjord	<i>Bay dun</i>		R				
Fjord	<i>Light bay dun</i>	R	R		R		
Fjord	<i>Bay dun</i>		G				
Liebenthaler	<i>Bay dun</i>		R		R		
Liebenthaler	<i>Black dun</i>		G			M	K
Przewalski	<i>Bay dun</i>			T			
Przewalski	<i>Bay dun</i>			T			
Ardenner	<i>Bay</i>						
Ardenner	<i>Chestnut</i>		G				
Kinsky	<i>Palomino</i>	R					
Riding horse	<i>Bay</i>						
Riding horse	<i>Chestnut</i>						
Riding horse	<i>Gray</i>		R				
Riding horse	<i>Black</i>	R	G				

Table 1. The table shows the position of the identified mutations in the equine melanophilin gene and the genotypes of each of the horses in the study (reference sequence: ENSECAG00000024624).

Although, *melanophilin (MLPH)* gene is located on chromosome 6, the decision was made to investigate this gene as possible candidate gene for the dun phenotype. The basic prerequisite for this decision was in particular that this gene plays a crucial role in the pigmentation process, because it acts after melanin synthesis and dilutes the basic coat coloration. These basic coat colorations (black, bay, chestnut) are defined by the ratio of the two pigments eumelanin and pheomelanin. This is an important prerequisite, because it is obvious that the causal gene for the dun phenotype affects both types of melanins equally. *MLPH* create together with *MYO5A* and *RAB27A* the protein complex, which has a decisive role in the intracellular melanosome (cellular compartment loaded with melanin/pigment) transport from the perinuclear to the peripheral region within the melanocytes (pigment cell). These proteins are organized as a motor protein complex, which transport maturing melanosomes to their final destination at the dendrite tips [16]. Today we know from mutative disruptions of this motor protein complex of myosin Va (*MYO5A*), *RAB27A* and *MLPH* that it leads to large clumps of pigment in hair shafts and an accumulation of melanosomes in melanocytes, and consequently to a visible hypopigmentation of the basic coat color [17, 18]. In mice, the following hypopigmented phenotypes are described for

these genes: *dilute (MYO5A)* [19], *ashen (RAB27A)* [20] and *leaden (MLPH)* [21]. Furthermore, mutations in the *MLPH* gene cause a diluted coat color in cats [22], as well as the lavender plumage dilution in chicken [23]. Moreover, a regulatory *MLPH* mutation might represent a causal mutation for coat color dilution in dogs [24, 25, 26]. Furthermore, recently, a SNP study identified a mutation in the *MYO5A*, which is responsible for the lavender foal syndrome [27]. Those foals have also diluted phenotypes, but this syndrome is lethal. Based on this current state of knowledge, the *MLPH* gene was a promising candidate gene as we started these analyses.

Methods

Horse samples: We focused our study on breeds which are known as homozygote dun horses and vice versa (Table 1).

DNA analysis: DNA was extracted from hairs or blood using standard protocols. The equine *MLPH* gene is located on chromosome 6 and includes 16 exons. The transcript length of *MLPH* (ENSECAT00000026561) amounts 1800 bps and is a product of gene ENSECAG00000024624.

		Primer sequence	Temp.
Exon 1	Exon 1 F	TTTGTGGATGAGCCAACATC	56 °C
	Exon 1 R	TTCAGAGGACATGGGAGGAC	
Exons 2/ 3	Exon 2 F	GGTTCACTGGTTTGGATGCT	55 °C
	Exon 3 R	ACATTAGCATGCCCTGGGAAC	
Exon 4	Exon 4 F	GCTGAGACAGGAGCTGGAAC	57 °C
	Exon 4R	GAGCTAGAAGCCCACTGCAT	
Exon 5	Exon 5 F	GAGGATGGATGGGTCGTAGA	58 °C
	Exon 5 R	GCCAGAGGAGTCATGGAGAG	
Exon 6	Exon 6 F	GGCATCTGGTCTCTCTGTC	58 °C
	Exon 6 R	CCAGAGCACTGTCCCTTACC	
Exon 7	Exon 7 F	GCGGATGGGTGTCTAGTGT	58 °C
	Exon 7 R	GCCTCAGTCTCTACTTCC	
Exon 8	Exon 8 F	CCCCACAACATCAGAGGAAT	58 °C
	Exon 8 R	TGTGCATTTTGTGACAGCTC	
Exon 9	Exon 9 F	CATGGACAGTGGCCTTACCT	58 °C
	Exon 9 R	GACATTCTTCCCTGCAAAGC	
Exon 10	Exon 10 F	TCATGTGGAGAATGGATGGA	55 °C
	Exon 10 R	GGCTGGTGACCCAGTTACAA	
Exon 11	Exon 11 F	ACACTGACAGGCTCAACG	58 °C
	Exon 11 R	GTCCACTGTTGCCAAACCTT	
Exon 12	Exon 12 F	AAAATGAGACCCCAGGAAG	58 °C
	Exon 12 R	CTGCAGTTCAGGAGAAAGG	
Exon 13	Exon 13 F	CTAACATGCCTTCCCAGCAT	58 °C
	Exon 13 R	ACCATCCGTCTGCTCAGTGT	
Exon 14	Exon 14 F	CCAAATCAGCTCCAGTGTT	56 °C
	Exon 14 R	TGCTGCATCCACAGATTAC	
Exon 15	Exon 15 F	GCCCTTTGTGTGCTTCTTA	58 °C
	Exon 15 R	TCTCATGGTCACGAGGCTGTG	
Exon 16	Exon 16 F	GCACACTTGTGTGTCATGT	55 °C
	Exon 16 R	AGTGAAACGTGGGAGCAG	

Table 2. PCR Primers used for the melanophilin amplification. The primers are located in flanking intron regions.

Primers were designed based on the horse genome assembly (EquCab2.0) using the Primer3 software (<http://www.simgene.com/Primer3>). They are located in flanking intron regions. Primer sequences and their optimal annealing conditions are listed in Table 2. The PCR was performed in a 25 µl reaction volume containing 1-2µl (20-100ng) of the genomic DNA, 50µM of each primer, 0.2 µM dNTP-mix, 2.5µl 10 X PCR-buffer, 1.5 mM MgCl₂-solution, 1 U bioTaqTM DNA-polymerase. The thermocycling protocol was: DNA and primers were initially incubated at 94°C for 3 minutes, cycled for 30 times at 94°C for 30 seconds, optimal annealing temperature for the primer system for 30 seconds (Table 2), 72°C for 1 minute and the final extension: 72°C for 7 minutes. PCR products varied

in length between 238bp and 781bp. Amplification products were visualized on an agarose gel. Finally, fragments were sequenced on an ABI PRISM 3730 capillary sequencer using the BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems) following the supplied protocol.

Alignments were performed using reference sequence: ENSECAG00000024624 in BioEdit version 7.0.9 (<http://www.mbio.ncsu.edu/bioedit/bioedit.html>).

Sampling and all analyses comply with the current law of Germany.

Results and discussion

We sequenced the full de novo sequences of the 16 exons of the *MLPH* gene. Thereby, we identified seven SNPs (Table 1). Whereas, four of them were synonymous (c.177 G>A, synonymous (heterozygote); c.918 C>T, synonymous (homozygote); c.1050 C>A, synonymous (heterozygote); c.1254 T>G, synonymous (heterozygote)), and the other three were non-synonymous substitutions (c.766 A>G non-synonymous (p.N256D; heterozygote and homozygote); c.989 G>T non-synonymous (p.G330V; heterozygote); c.1031 G>A, non-synonymous (p.R344H; heterozygote)).

However, none of these substitutions showed any obvious association with the dun phenotype, because there was no correlation between one or more substitutions and dun. Nevertheless, two amino acid substitutions (p.G330V and p.R344H) were solely found in dun horses but never in all of them. The five synonymous substitutions also show no association with the dun phenotype. But interestingly, five out of seven SNPs were only found in dun horses; however, none of these SNPs was exclusively present in dun horses. Therefore, one can say that this investigation revealed whether one or a combination of SNPs, which was the causal dun mutation. Otherwise, there is still the likelihood that *MLPH* is the causative gene for the dun phenotype, because this investigation did not include additional important sequence regions of this gene like the UTR regions, the promoter region and the intron sequences (potential splice site locations). Finally, we cannot exclude that one or a combination of these mutations are involved in another dilution phenotype; however, the dataset of this research allowed only focusing on the dun phenotype.

Until now, in human, laboratory mouse and 3 domestic species (dog, cat, chicken), mutant *MLPH* alleles are known for leading to hypopigmentation of the basic coat color [21, 22, 23, 26, 28]. In total, 4 different SNPs were determined, resulting in 3 different mechanisms that lead to hypo-

pigmentation: (a) disruption of the RAB27A binding domain; (b) early stopping of the *MLPH* translation; (c) affection on the splicing efficiency of *MLPH*. Except for the SNP at the splice donor of exon 1 as a candidate causal mutation for coat color dilution in dogs, the other 3 SNPs all have a common characteristic: they are all located within the coding region of the *MLPH* gene responsible for the binding of RAB27A. On the contrary, neither of the substitutions detected in this investigation are located in this sequence region.

As noticed above, a linkage map analysis mapped the dun phenotype on the p-arm of ECA8. But Bricker [15] did not find a promising candidate gene for dun within this region. Addressing further studies, in our opinion, there are several genes (within and close to the defined region), which can be considered as candidate genes, for example, the *D-dopachrome tautomerase gene (DDT)*. This gene codes for an enzyme, which is probably involved in the melanin biosynthesis [29] and strong related to the migration inhibitory factor (MIF). Obviously, MIF may play an important role in the expression of dorsal-ventral pigment pattern of amphibians [30]. Another gene could be the *rabphilin (RPH3A)* gene. It is a RAB3A effector [31] that in turn, in collaboration with SNARE (small noncoding RNA) plays a possible role in targeting melanosomes to the plasma membrane and melanosome transfer to surrounding cells [32]. Interestingly, we found also an albinistic gene, the *hermansky-pudlak syndrome 4 (HPS4)* gene. It leads to abnormalities of melanosomes in the so-called light-ear mutant mouse [33].



Figure 1. Dun horses are characterized by the dilution of their basic coat color, whereas the extent of dilution at the lower legs and head is less obvious. In addition, dun horses possess a dorsal stripe, while the presence of other patterns, especially darker stripes on the lower legs and the shoulders can vary.

Abbreviations

AA: amino acid; CDS: coding sequence for protein; MLPH: Melanophilin; MYO5A: Myosin5A; SNP: Single Nucleotide Polymorphism.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

AL, MC and MR designed the research. MC and MR conducted the lab work. MC analyzed the data. All authors wrote the paper. All authors read and approved the manuscript.

Acknowledgements and funding

We thank all horse breeders for their help in arranging sample collection. Dietmar Lieckfeldt is gratefully acknowledged for his technical assistance. This work was funded by the Deutsche Forschungsgemeinschaft (LU 852/6-2 & LU 852/7-3).

References

- Ludwig A, Pruvost M, Reissmann M, Benecke N, Brockmann GA, Castanos P, Cieslak M, Lippold S, Llorente L, Malaspinas AS, Slatkin M, Hofreiter M: **Coat color variation at the beginning of horse domestication.** *Science* 2009, **324**: 485.
- Sponenberg P: **Equine Color Genetics**, 3rd Edition. *Wiley-Blackwell* 2009, 296 pp.
- Marklund L, Moller MJ, Sandberg K, Andersson L: **A missense mutation in the gene for melanocyte-stimulating hormone receptor (MC1R) is associated with the chestnut coat color in horses.** *Mamm Genome* 1996, **7**: 895-899.
- Metallinos DL, Bowling AT, Rine J: **A missense mutation in the endothelin-B receptor gene is associated with lethal white foal syndrome: an equine version of Hirschsprung disease.** *Mamm Genome* 1998, **9**: 426-431.
- Santchi EM, Purdy AK, Valberg SJ, Vrotsos PD, Kaese H, Mickelson NR: **Endothelin receptor B polymorphism associated with lethal white foal syndrome in horses.** *Mamm Genome* 1998, **9**: 306-309.
- Rieder S, Taourit S, Mariat D, Langlois B, Guerin G: **Mutations in the agouti (ASIP), the extension (MC1R), and the brown (TYRP1) loci and their association to coat colour phenotypes in horses (Equus caballus).** *Mamm Genome* 2001, **12**: 450-455.
- Mariat D, Taourit S, Guerin G: **A mutation in the MATP gene causes the cream coat colour in the horse.** *Genet Select Evol* 2003, **35**: 119-133.
- Brooks SA and Bailey E: **Exon skipping in the KIT gene causes a sabino spotting pattern in horses.** *Mamm Genome* 2005, **16**: 893-902.
- Brunberg E, Andersson L, Cothran G, Sandberg K, Mikko S, Lindgren G: **A missense mutation in PMEL17 is associated with the Silver coat color in the horse.** *BMC Genet* 2006, **7**: 46.
- Brooks SA, Lear TL, Adelson DL, Bailey E: **A chromosome inversion near the KIT gene and the tobiano spotting pattern in horses.** *Cytogenet Genome Res* 2007, **119**: 225-230.
- Haase B, Brooks SA, Schlumbaum A, Azor PJ, Bailey E, Alaeddine F, Mevissen M, Burger D, Poncet P, Rieder S, Leeb T: **Allelic heterogeneity at the equine KIT Locus in dominant white (W) horses.** *PLoS Genet* 2007, **3**(11), e195.
- Reissmann M, Bierwolf J, Brockmann GA: **Two SNPs in the SILV gene are associated with silver coat colour in ponies.** *Anim Genet* 2007, **38**: 1-6.
- Pielberg GR, Golovko A, Sundström E, Curik I, Lennartsson J, Seltenthammer MH, Druml T, Binns M, Fitzsimmons C, Lindgren G, Sandberg K, Baumung R, Vetterlein M, Strömberg S, Grabherr M, Wade C, Lindblad-Toh K, Ponten F, Heldin C, Sölkner J, Andersson L: **A cis-acting regulatory mutation causes premature hair graying and susceptibility to melanoma in the horse.** *Nat Genetics* 2008, **40**: 1004-1009.
- Haase B, Brooks SA, Tozaki T, Burger D, Poncet PA, Rieder S, Hasegawa T, Penedo C, Leeb T: **Seven novel KIT mutations in horses with white coat colour phenotypes.** *Anim Genet* 2009, **40**: 623-629.
- Bricker SJ: **Mapping of the Dun locus in Horses (Equus caballus) to Equine Chromosome 8 by Linkage to Microsatellite Markers.** *Master Thesis* 2001 University of California, Davis, 60 pp.
- Hume AN, Ushakov DS, Tarafder AK, Ferenczi MA, Seabra MC: **Rab27a and MyoVa are the primary Mlph interactors regulating melanosome transport in melanocytes.** *J. Cell Sci.* 2007, **120**: 3111-3122.
- Van Gele M, Dynoodt P, Lambert J: **Griscelli syndrome: a model system to study vesicular trafficking.** *Pigment Cell Melanoma Res* 2009, **22**: 268-282.
- O'Sullivan TN, Wu XS, Rachel RA, Huang JD, Swing DA, Matesic LE, Hammer JA 3rd, Copeland NG, Jenkins NA: **dsu functions in a MYO5A-independent pathway to suppress the coat color of dilute mice.** *Proc Natl Acad Sci USA* 2004, **101**: 16831-16836.
- Jenkins NA, Copeland NG, Taylor BA, Lee BK: **Dilute (d) coat colour mutation of DBA/2J mice is associated with the site of integration of an ecotropic MuLV genome.** *Nature* 1981, **293**: 370-374.
- Wilson SM, Yip R, Swing DA, O'Sullivan TN, Zhang Y, Novak EK, Swank RT, Russell B, Copeland NG, Jenkins NA: **A mutation in Rab27a causes the vesicle transport defects observed in ashken mice.** *Proc. Natl. Acad. Sci. U. S. A.* 2000, **97**: 7933-7938.

21. Matesic LE, Yip R, Reuss AE, Swing DA, O'Sullivan TN, Fletcher CF, Copeland NG, Jenkins NA: **Mutations in *Mlph*, encoding a member of the Rab effector family, cause the melanosome transport defects observed in leaden mice.** *Proc Natl Acad Sci U S A* 2001, **98**: 10238–10243.
 22. Ishida Y, David VA, Eizirik E, Schäfer AA, Neelam BA, Roelke ME, Hannah SS, O'Brien SJ, Menotti-Raymond M: **A homozygous single-base deletion in *MLPH* causes the dilute coat color phenotype in the domestic cat.** *Genomics* 2006, **88**: 698-705.
 23. Vaez M, Follet SA, Bed'hom B, Gourichon D, Tixier-Boichard M, Burke T: **A single point-mutation within the melanophilin gene causes the lavender plumage colour dilution phenotype in the chicken.** *BMC Genet* 2008, **9**: 7.
 24. Philipp U, Hamann H, Mecklenburg L, Nishino S, Mignot E, Günzel-Apel AR, Schmutz SM, Leeb T: **Polymorphisms within the canine *MLPH* gene are associated with diluted coat color in dogs.** *BMC Genet.* 2005, **6**: 34.
 25. Drögemüller C, Philipp U, Haase B, Günzel-Apel A, Leeb T: **A Noncoding Melanophilin Gene (*MLPH*) SNP at the Splice Donor of Exon 1 Represents a Candidate Causal Mutation for Coat Color Dilution in Dogs.** *Journal of Heredity* 2007, **98**(5):468-473.
 26. Welle M, Philipp, U, Rüfenacht S, Roosje P, Scharfenstein M, Schütz E, Brenig B, Linek M, Mecklenburg L, Grest P, Drögemüller M, Haase B, Leeb T, Drögemüller C: ***MLPH* Genotype-Melanin Phenotype Correlation in Dilute Dogs.** *J Heredity* 2009, **100**: S75-S79.
 27. Brooks SA, Gabreski N, Miller D, Brisbin A, Brown HE, Streeter C, Mezey J, Cook D, Antczak DF: **Whole-Genome SNP Association in the Horse: Identification of a Deletion in Myosin Va Responsible for Lavender Foal Syndrome.** *PLoS Genetics* 2010, **6**(4):e1000909.
 28. Ménasché G, Ho CH, Sanal O, Feldmann J, Tezcan I, Ersoy F, Houdusse A, Fischer A, de Saint Basile G: **Griscelli syndrome restricted to hypopigmentation results from a melanophilin defect (*GS3*) or a *MYO5A* F-exon deletion (*GS1*).** *J Clin Invest.* 2003, **112**(3): 450-6.
 29. Slominski A, Tobin DJ, Shibahara S, Wortsman J: **Melanin Pigmentation in Mammalian Skin and Its Hormonal Regulation.** *Physiol. Rev.* 2004, **84**: 1155-1228.
 30. Fukuzawa T, Samaraweera P, Mangano FT, Law JH, Bagnara JT: **Evidence that *MIF* plays a role in the development of pigmentation patterns in the frog.** *Dev Biol.* 1995, **167**(1): 148-58.
 31. Lin CC, Huang CC, Lin KH, Cheng KH, Yang DM, Tsai YS, Ong RY, Huang YN, Kao LS: **Visualization of *Rab3A* dissociation during exocytosis: a study by total internal reflection microscopy.** *J Cell Physiol.* 2007, **211**(2): 316-26.
 32. Scott G & Zhao Q: ***Rab3a* and SNARE proteins: potential regulators of melanosome movement.** *J Invest Dermatol.* 2001, **116**(2): 296-304.
 33. Suzuki T, Li W, Zhang Q, Karim A, Novak EK, Sviderskaya EV, Hill SP, Bennett DC, Levin AV, Nieuwenhuis HK, Fong CT, Castellán C, Mitterski B, Swank RT, Spritz RA: **"Hermansky-Pudlak syndrome is caused by mutations in *HPS4*, the human homolog of the mouse light-ear gene".** *Nat Genet* 2002, **30**(3): 321-4.
-

ERKLÄRUNG

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen oder indirekt übernommenen Daten sind unter Angabe der Quelle gekennzeichnet. Die Arbeit wurde bisher weder im In- noch im Ausland an einer anderen Hochschule vorgelegt.

Unterschrift:

Michael Cieslak
