Biodiversity of Arabian horses in Syria

Dissertation

zur Erlangung des akademischen Grades
Doctor rerum agriculturarum
(Dr. rer. agr.)

eingereicht an der Lebenswissenschaftlichen Fakultät der Humboldt Universität zu Berlin

von
M.Sc. Saria Almarzook

Präsidentin der Humboldt-Universität zu Berlin
Prof. Dr. Sabine Kunst

Dekan der Humboldt-Universität zu Berlin
Prof. Dr. Bernhard Grimm

Gutachterin/Gutachter

Prof. Dr. Gudrun Brockmann

Prof. Dr. Dirk Hinrichs

Prof. Dr. Armin Schmitt

Tag der mündlichen Prüfung:
17. September 2018

Dedication

This research is dedicated to my homeland

...Syria

Contents

Zusammenfassung

Die Ziele dieser Arbeit waren (1) die genetische Vielfalt der syrisch-arabischen Pferden zu untersuchen und (2) genetische Variationen zu identifizieren, die mit Ausdauerleistungen bei diesen Pferden in Verbindung gebracht werden können. Das Hauptziel der Untersuchung der Artenvielfalt syrischer Araber war die Identifizierung der Populationsstruktur in einer Reihe von syrischen Araberpferden, die die drei Hauptstämme Saglawi, Kahlawi und Hamdani repräsentieren. Für die Studie wurden molekulare Marker der Sequenzdaten der 353 bp der hypervariablen Region der mitochondrialen D-Schleife, sowie der Daten von 12 Pferde-Mikrosatelliten und 38.671 genomweite SNPs verwendet. Kenntnisse über die Variabilität der Ausdauer sowie über morphologische Eigenschaften sind in Zuchtprogrammen von Vorteil. Wir haben in einer genomweiten Assoziationsstudie (GWAS) getestet, ob es einen kausalen Zusammenhang zwischen morphologischen Merkmalen und Variabilität der Ausdauer mit genetischen Polymorphismen gibt. Darüber hinaus haben wir Kandidatengene näher charakterisiert, welche zu dem komplexen Merkmal der Ausdauerleistung beitragen könnten.

Für die in dieser Arbeit vorgestellten Studien wurden insgesamt 200 syrische Araber-Pferde (136 weibliche und 64 männliche) eingesetzt. Pferde stammen aus fünf Regierungsregionen Syriens. Diese Pferde stammen von 123 Muttertieren und 88 Vatertieren ab. Sie wurden zwischen 1986 und 2013 geboren. Für die Pferde standen folgende morphologische Daten zur Verfügung: Rückenlänge, Körperlänge, Brusttiefe, Brustumfang, Brustbreite, Kruppenhöhe, Kruppbreite, Länge des vorderen Röhrbeines, Länge des hinteren Röhrbeines, Nackenumfang, Kehlumfang und Widerristhöhe sowie Ausdaueraufzeichnungen von Pferderennen.

Arabische Pferderegister klassifizieren arabische Pferde, basierend auf ihren Mutterlinien, in fünf Hauptstämme. Um den mütterlichen Ursprung der syrischen Araberpferde zu testen, wurden 192 Pferde, die die drei Hauptstämme in Syrien Saglawi (n=57), Kahlawi (n=75) und Hamdani (n=60) repräsentierten, für 353 bp der Region der mitochondrialen Verdrängungsschleife (D-Schleife) sequenziert. Die vergleichende Sequenzanalyse ergab 38 Sequenzvarianten, die 28 Haplotypen bilden. Die Haplotyp-Diversitätswerte innerhalb der Stämme betrugen 0.95, 0.91 und 0.90 in Kahlawi, Hamdani bzw. Saglawi. Die paarweisen Populationsdifferenzierung Werte (F_{ST}) zwischen den Stämmen waren niedrig und lagen zwischen 0.098 und 0.205. Die Haplotyp-Diversität und die F_{ST} Werte zwischen den Stämmen zeigten eine hohe Diversität innerhalb der Individuen jedes Stammes und eine geringe Variation zwischen den drei Stämmen. Dieses Ergebnis zeigt auf, dass eine Zuordnung eines Pferdes auf Grund der Sequenzanalyse der D-Schleife nicht eindeutig zu einem Stamm vorgenommen werden kann. Zudem streuen die verschiedenen mitochondrialen Haplotypen über den gesamten benachbarten Stammbaum hinweg ohne klare Trennung der drei Stämme. Im Median-Verbindungsnetzwerk wurden die syrischen Pferde in sieben große Haplogruppen eingeteilt. Diese Ergebnisse legen nahe, dass mehr als fünf Vorfahren existieren, die gemeinsame mütterliche Haplotypen teilen.Die Bestimmung des mtDNA-Status allein durch den D-Loop könnte zu einer Fehlinterpretation der Variation zwischen den Stämmen führen, daher führten wir zusätzlich eine genomweite Analyse der Pferde mit zwei Arten von Markern durch.

Im ersten Ansatz wurden zwölf Mikrosatelliten zur Schätzung der genetischen Diversität innerhalb und zwischen den drei syrischen Araberpferden (Saglawi, Kahlawi und Hamdani) verwendet. Die Population umfasste insgesamt 84 Pferde, darunter die drei großen

syrischen Stämme Hamdani (n=26), Kahlawi (n=30) und Saglawi (n=28). Um die genetische Populationsstruktur innerhalb und zwischen den drei Stämmen zu bestimmen, führten wir genetische Diversitätsanalysen, die Analyse der molekularen Varianz, Wrights F-Statistik, konstruierte Hauptkoordinationsfiguren und einen phylogenetischen Baum mit der Neighbor-Joining (NJ) -Methode durch. Die Ergebnisse zeigen, dass Saglawi positiv und mehr als die anderen beiden Stämme zum gesamten polymorphen Inhalt beiträgt. Dennoch ergab die Analyse der molekularen Varianz, dass nur 0.045 der genetischen Variation auf Unterschiede in der Abstammungseinteilung durch das Zuchtbuch zurückzuführen waren, was durch niedrige paarweise F_{ST}-Werte (0.013 bis 0.015) unterstützt wurde. Diese Ergebnisse stimmten mit den Ergebnissen der wichtigsten Koordinierungszahlen überein. In ähnlicher Weise deutet der phylogenetische Baum auf eine geringe Populationsdifferenzierung zwischen den drei Stämmen hin. Wir mussten jedoch feststellen, dass die 12 Mikrosatelliten, die in dieser Studie verwendet wurden, nicht ausreichten, um das gesamte Genom darzustellen.

Im zweiten Ansatz verwendeten wir 38.671 genomweit verteilte SNPs. In dieser Studie untersuchten wir 48 Pferdemit dem Equine SNP70K BeadChip (Illumina), welche wieder die drei syrischen Mutterstämme Saglawi (n=18), Kahlawi (n=16) und Hamdani (n=14) repräsentierten. Zum Vergleich, wurden zusätzlich 24 arabische Pferde aus den USA und drei Przewalski-Pferde als Out-Gruppen hinzugefügt. Die beobachtete Heterozygotie (H_o) bei syrischen Araberpferden lag zwischen 0.30 und 0.32. Dies steht im Einklang mit der erwarteten Heterozygotie (H_e), die zwischen 0.30 und 0.31 lag. Die Inzuchtkoeffizienten (F_{IS}) lagen zwischen 0.02 und 0.05, was auf eine hohe genetische Diversität innerhalb der syrischen Stämme hindeutet. Ebenso war die paarweise genetische Differenzierung

zwischen den drei syrischen Stämmen sehr gering (F_{ST} <0.05). Hierarchisches Clustering zeigte eine klare Unterscheidung zwischen Pferden von Araber und Przewalski. Unter arabischen Pferden fanden wir drei Gruppen, die entweder Pferde aus den USA, Pferde aus Syrien oder Pferde aus Syrien und den USA zusammenhielten. Individuen aus demselben syrischen Araberstamm wurden auf verschiedene Untergruppen verteilt. Bei der Analyse syrischer Araber allein wurde die beste Populationsdifferenzierung mit drei verschiedenen Clustern gefunden. Im Gegensatz zu den Erwartungen aus dem Zuchtbuch stimmten diese Cluster nicht mit der auf mütterlichen Linien beruhenden Abstammungseinteilung überein. Obwohl diese Befunde die Hypothese von drei Gründern unterstützen, stimmt die genetische Information nicht mit dem gegenwärtig verwendeten Stammbezeichnungssystem überein. Die in dieser Studie gewonnenen Informationen können zur Überarbeitung der aktuellen Zuchtpraxis verwendet werden. Darüber hinaus belegten die Ergebnisse, dass syrische Araberpferde ein wichtiges Reservoir für die genetische Vielfalt darstellen.

In der ebenfalls durchgeführten genomweiten Assoziationsstudie war unser Ziel, genetische Marker zu identifizieren, die mit morphologischen Merkmalen bei syrischen Arabern als auch der Ausdauerleistung assoziiert sind. Wir führten eine genomweite Assoziationsstudie (GWAS) mit 14.920 informativen SNPs des Equine SNP70 BeadChip (Illumina) und Merkmalen für Wachstum und Körperkonformation durch. Von 48 genotypisierten Tieren konnten 37 Pferde (8 männliche, 29 weibliche) mit normalisierten Phänotypen verwendet werden. Die Populationsgröße war für eine genomweite Assoziationsstudie sehr klein, deshalb konnten wir wahrscheinlich keine signifikanten Assoziationen finden (LOD> 5.5). Dennoch zeigten zwei morphologische Merkmale

suggestive Assoziationen (LOD> 5). Der Halsumfang (NG) war suggestiv mit SNP BIEC2_772752 auf Chromosom 3 assoziiert (LOD = 5.18). Das Hauptallel scheint einen dominant wirkenden Effekt auf den Halsumfang von 6,0 cm zu haben. Zusätzlich wurde die Brustbreite (ChW) suggestiv mit SNP BIEC2_444806 auf Chromosom 19 assoziiert (LOD = 5.10). Bei diesem SNP ist der heterozygote Genotyp im Vergleich zu den beiden homozygoten Genotypklassen mit einer geringeren Brustbreite assoziiert. Die flankierende genomische Region um die suggestiven SNPs enthält keine bekannten, funktionell relevanten Gene. Daher sind zukünftige Studien erforderlich, um die nachgewiesenen suggestiven Assoziationen unter Verwendung einer entweder größeren oder einer unabhängigen Population zu validieren.

Ein weiteres Ziel der Arbeit war die erstmalige Untersuchung von ausdauerbezogenen Genen bei in Syrien aufgezogenen arabischen Pferden. Es sollten, genetische Variation in Kandidatengenen identifiziert werden, die möglicherweise Ausdauerleistungsmerkmale beeinflussen um diese gegebenenfalls mit Ausdauerphänotypen zu assoziieren. Die drei mitochondrialen Gene *MT-COX3*, *MT-CYB* und *MT-ND4* und die beiden autosomalen Gene *ACTN3* und *MSTN* wurden vergleichend sequenziert. Leistungsmerkmale wurden für 42 arabische Pferde aus Aufzeichnungen zwischen 2001 und 2010 bei Ausdauerveranstaltungen von über 40, 80 und 120 km Entfernung zur Verfügung gestellt. Die Pferde wurden nach ihren Leistungen in Low und High Performer eingeteilt. Vergleichende Sequenzierung führte zur Identifizierung von 13 Varianten in den mitochondrialen Genen und zusätzlichen 13 Varianten in den autosomalen Genen. Die Variante mt: 9280T> C in *MT-COX3* wurde nur bei den Pferden der High Performer Gruppe

gefunden. Es wurde kein Unterschied zwischen den beiden Leistungsgruppen für Allelfrequenzen von DNA-Varianten in *ACTN3*- und *MSTN*-Genen gefunden.

Zusammenfassend zeigt die vorliegende Studie die Diversität der großen syrisch-arabischen Pferde-Stämme: Saglawi, Kahlawi und Hamdani, anhand der Sequenzdaten der hypervariablen Region der mitochondrialen D-Schleife, sowie der Daten von 12 genutzten Mikrosatelliten und 38.671 genomweite SNPs. In der GWAS, die in dieser Studie mit 14.920 informativen SNPs und 12 morphologischen Merkmalen von syrischen Arabern durchgeführt wurde, zeigten zwei morphologische Merkmale (Halsumfang und Brustbreite) suggestive Assoziationen (LOD> 5) mit zwei SNPs auf Chromosom 3 bzw. 19. Desweiteren führte die vergleichende Sequenzanalyse der fünf ausdauerbezogenen Kandidatengene zur Identifizierung von 13 Polymorphismen in den mitochondrialen Genen und 13 Polymorphismen in den autosomalen Genen. Nur eine mitochondriale Variante war in der Hochleistungsgruppe vorhanden.

Weitere anwendungsorientierte Forschung, gut gemessene Daten und gut geplante Zuchtstrategien sind erforderlich, um die vielfältigen genetischen Ressourcen der arabischen Pferde und deren wünschenswerte Eigenschaften zu erhalten.

Summary

The goals of this thesis were (1) to study the genetic diversity among Syrian Arabian horses and (2) to identify genetic variation that could be linked to endurance racing performance in these horses. The major objective of studying the biodiversity of Syrian Arabian horses was to identify the population structure in a set of Syrian Arabian horses representing the three major strains Saglawi, Kahlawi, and Hamdani. For the study, we used different

genetic markers. The variability of endurance and morphological traits are beneficial in breeding programs. We tested genome-wide associations (GWAS) to find a causal relationship between morphological traits and genetic variants. Furthermore, we used the candidate gene approach to investigate the variability of genes contributing to endurance performance.

For the studies presented in this thesis, a total number of 200 Syrian Arabian horses (136 females and 64 males) were used. Horses originate from five governance regions of Syria. These horses are descendant from 123 dams and 88 sires. They were born between 1986 and 2013. For these horses the following morphological data were available and used: back line length (BLL), body length (BL), chest depth (ChD), chest girth (ChG), chest width (ChW), croup height (CH), croup width (CW), fore cannon length (FCL), hind cannon length (HCL), neck girth (NG), throat girth (TG), and withers height (WH), as well as endurance racing records.

Arabian horse registries classify Arabian horses based on their dam lineages into five main strains. To test the maternal origin of Syrian Arabian horses, 192 horses representing the three major strains in Syria Saglawi (n=57), Kahlawi (n=75), and Hamdani (n=60) were sequenced for 353 bp of the mitochondrial displacement loop (D-loop) region. Comparative sequence analysis revealed 38 sequence variants forming 28 haplotypes. The haplotype diversity values within strains were 0.95, 0.91, and 0.90 in Kahlawi, Hamdani, and Saglawi, respectively. The pair-wise population differentiation estimates (F_{ST}) between strains were low, ranging between 0.098 and 0.205. The haplotype diversity and the pair-wise population differentiation estimates (F_{ST}) between strains showed high diversity within individuals of each strain and low variation between the three strains. The different

mitochondrial haplotypes scatter across the whole neighbor joining tree without clear separation of the three strains. In the median-joining network, the Syrian horses were grouped into seven major haplogroups. These results suggest that more than five ancestors exist that share common maternal haplotypes.

Determination of mitochondrial DNA status solely by D-loop could lead to misinterpretation of variation between strains, therefore, we additionally performed a genome-wide analysis of the horses using two types of markers.

In the first approach, twelve equine microsatellites were used to estimate the genetic diversity within and between the three Syrian Arabian horse strains (Saglawi, Kahlawi and Hamdani). The population included in total 84 horses comprising the major Syrian strains Hamdani (n = 26), Kahlawi (n = 30) and Saglawi (n = 28). To determine the genetic population structure within and between the three strains, we performed standard genetic diversity analyses, the analysis of molecular variance and Wright´s F statistics. Additionally, we constructed genetic distance-based principle coordination figures and a phylogenetic tree using the neighbor-joining (NJ) method. The results show that Saglawi contributes positively and more than the other two strains to the total polymorphic content. Nevertheless, the analysis of molecular variance revealed that only 0.045 of the genetic variation was attributed to strain differences, which was further supported by low pairwise F_{ST} values (0.013 to 0.015). These results were consistent with the results of the principle coordination figures. Likewise, the phylogenetic tree suggests low level of population differentiation between the three strains. However, we have to state that the 12 microsatellites used in this study are not sufficient to display the whole genome.

In the second approach, we used 38.671 genome wide-distributed SNPs. In this study, we examined 48 horses representing Saglawi (n=18), Kahlawi (n=16) and Hamdani (n=14) strains using the Equine SNP70K BeadChip (Illumina). For comparison, additional 24 Arabian horses from the USA and three Przewalski's horses were added as out groups. The observed heterozygosis (H_o) within Syrian Arabian horse strains ranged between 0.30 and 0.32. This is consistent with the expected heterozygosity (H_e), which ranged between 0.30 and 0.31. The inbreeding coefficients (F_{IS}) were between 0.02 and 0.05, indicating high genetic diversity within Syrian strains. Likewise, the pairwise genetic differentiation F_{ST} between the three Syrian strains was very low ($F_{ST} < 0.05$). Hierarchical clustering showed a clear distinction between Arabian and Przewalski's horses. Among Arabian horses, we found three clusters containing either horses from the USA or horses from Syria or horses from Syria and the USA together. Individuals from the same Syrian Arabian horse strain were spread across different sub-clusters. When analyzing Syrian Arabian horses alone, the best population differentiation was found with three distinct clusters. In contrast to expectations from the studbook, these clusters did not coincide with strains affiliation, which is based on maternal linages. Although this finding supports the hypothesis of three founders, the genetic information is not consistent with the currently used strain designation system. Information obtained in this study can be used to revise the current breeding practice. Beyond that, the results provided evidence that Syrian Arabian horses are an important reservoir for genetic diversity.

In the genome-wide association study, our goal was to identify genetic markers associated with morphological traits in Syrian Arabian horses. We performed a genome wide association study (GWAS) using 14.920 informative SNPs of the Equine SNP70 BeadChip

(Illumina) and traits for growth and body conformation. Out of 48 genotyped animals (see above), we could use 37 horses (8 males, 29 females) which had normalized phenotypes. Since the population size was very small, we were not surprised, not to find significant associations (LOD > 5.5). Nevertheless, two morphological traits showed suggestive associations (LOD > 5). Neck girth (NG) was suggestively associated with SNP BIEC2_772752 on chromosome 3 (LOD = 5.18). The major allele seems to have a dominant increasing effect on neck girth by 6.0 cm. In addition, chest width (ChW) was suggestively associated with SNP BIEC2_444806 on chromosome 19 (LOD = 5.10). At this SNP, the heterozygous genotype is associated with a smaller chest width compared to the two homozygous genotype classes. The flanking genomic region around the suggestive SNPs does not contain known functionally relevant genes. Therefore, future studies are required to validate the detected suggestive associations using a larger Arabian horse population.

This study is the first examining endurance-related genes in Arabian horses raised in Syria. The aim was to identify genetic variation in candidate genes that could potentially affect endurance traits and to associate them with endurance phenotypes. The three mitochondrial genes *MT-COX3*, *MT-CYB* and *MT-ND4* and the two autosomal genes *ACTN3* and *MSTN* were comparatively sequenced. Performance traits were available for 42 Arabian horses from records in endurance events between 2001 and 2010, over 40, 80, and 120 km distances. Horses were grouped according to their performance into low and high performers groups.

Comparative sequencing led to identify 13 variants in the mitochondrial genes and additional 13 variants in the autosomal genes. The variant mt:9280T>C in *MT-COX3* was found only in the high performing endurance horses. No difference between the two

performance groups was found for allele frequencies of DNA variants in *ACTN3* and *MSTN* genes.

In conclusion, the present study illustrates the diversity of the major Syrian Arabian horse strains: Saglawi, Kahlawi and Hamdani using 353 bp of mitochondrial D-loop hypervariable region, 12 equine microsatellites, and 38.671 genome-wide SNPs. In the GWAS performed in this study using 14.920 informative SNPs and 12 morphological traits information of Syrian Arabian horses, two morphological traits (neck girth and chest width) showed suggestive associations with two SNPs on chromosome 3 and 19, respectively. Finally, the comparative sequences analysis of the five endurance-related genes of the candidate genes led to the identification of 13 polymorphisms in the mitochondrial genes and 13 polymorphisms in the autosomal genes. Only one mitochondrial variant was present only in the high performance group.

Continued application-oriented research, appropriate well-measured data, and well planned breeding strategies are required to preserve genetic diversity resources, and to maintain the desirable traits in the Arabian horse breed.

List of publications and presentations

Almarzook S., Reissmann M., Arends D. & Brockmann G. (2017) Genetic diversity of Syrian Arabian horses. Animal genetics 48, 486–9.ISSN 1365-2052. DOI:10.1111/age.12568

Almarzook S., Reissmann M. & Brockmann G. (2017) Diversity of mitochondrial DNA in three Arabian horse strains. Journal of applied genetics 58, 273–6. ISSN 1234-1983. DOI:10.1007/s13353-016-0384-z

Almarzook S., Reissmann M., Arends D. & Brockmann G. (September 20-21, 2016) Genetic Diversity of Syrian Arabian horses. DGFZ conference. Hannover, Germany

Almarzook S., Reissmann M. & Brockmann G. (Juni 24, 2016) Poster: Maternal lineages diversity in Arabian horse breed. Agrosnet Doktorandentages. Rostock, Germany.

Almarzook S., Reissmann M., Arends D. & Brockmann G. (Mai 3-6, 2016) Genetic diversity of Arabian horses in Syria using SNP70 array. ConGenomics conference. Vairão, Portugal.

Almarzook S., Reissmann M. & Brockmann G. (September 16-17, 2015) Mitochondrial DNA lineages in Arabian horses. Vortragstagung der DGfZ und der GFT. Berlin, Germany.

Almarzook S., Reissmann M. & Brockmann G. (March 10-11, 2015) Variations in the mitochondrial D-Loop region sequences among three Strains of the Arabian horse breed. AgrosNet-Doktorandentag. Berlin, Germany.

Almarzook S., Reissmann M., Arends D. & Brockmann G. (Mai 19-22, 2014) Poster: Comparison of morphometric traits in Arabian horse strains. Complex Trait Community (CTC) international conference. Berlin, Germany.

List of abbreviations

µl	Microliter	**K**	Kahlawi strain
ACE	angiotensin I converting enzyme	**Km**	Kilometer
ACTN3	Actinin alpha 3	**$MgCl_2$**	Magnesium chloride
ADRB2	β2-adrenergic receptor	**MJ**	Median-joining
AMOVA	Analysis of molecular variance	**mRNA**	Messenger ribonucleic acid
ATP	Adenosine triphosphate	**mtDNA**	Mitochondrial DNA
BL	Body length	**NG**	Neck girth
BLL	Back line length	**N_{HT}**	Number of haplotypes
bp	Base pair	**N_J**	Neighbor-joining
ChD	Chest depth	**N_T**	Total number of animals
ChG	Chest girth	**N_U**	Number of unrelated individuals
ChW	Chest width	**N_a**	Number of alleles
CKM	Creatine kinase, M-type	**N_{Pa}**	Number of private alleles
CW	Croup width	**PCR**	Polymerase Chain Reaction
D-loop	Displacement loop	**PPARGC1A**	PPARG coactivator 1 alpha
ECA	Equus caballus autosome	**S**	Saglawi strain
EDTA	Ethylenediaminetetracetic acid	**SAS**	Statistical Analysis Software
FCL	Fore cannon length	**SD**	Standard deviation
GYS1	Glycogen synthase 1	**SE**	Standard error
H	Hamdani strain	**SNP**	Single nucleotide polymorphism
HCL	Hind cannon length	**TAE**	Tri-acetate and EDTA
HTD	haplotype diversity	**TG**	Throat girth
H_o	Observed heterozygosity	**TAE**	Tri-acetate and EDTA
H_e	Expected heterozygosity	**UTR**	Untranslated region
IGF1	Insulin like growth factor 1	**VDR**	Vitamin D receptor
Kb	Kilobasepair	**WH**	Withers height

List of figures

List of tables

1. General introduction and literature review

Horses have made a remarkable contribution to international human civilization. After domestication, horses were used for hunting, travelling, wars, landscape management, leisure, and served as a nutrition source (milk and meat). Horse industry was an important investment in the 19th century, mainly for the strong economic countries. Today, changes of the public interests and expenses reduce the entries in the horse breeding and industry. However, this sector is still productive, for instance, UK official authorities reported in 2013 approximately 274 million £ was generated in the UK economy from exporting horses worldwide (GOV.UK 2014).

In the last 15 years, equine diversity research has focused on the origin of the horse populations (Royo *et al.* 2005), the maternal genetic variations (Bowling *et al.* 2000; Hill *et al.* 2002), differences explained within and between population (Leroy *et al.* 2009), as well as identification of genes related to the economical traits, like e.g. racing performance (Binns *et al.* 2010).

The release of the first two assemblies of the horse sequences (EquCab1.0 and EquCab2.0) in 2007 and the publication of the whole equine genome sequence in 2009 (Wade *et al.* 2009) enhanced the knowledge about the genome of horse breeds, which encouraged us to perform this diversity study on Arabian horses. Additionally, we investigate the genetic background of the morphological and endurance traits of the Arabian horses.

1.1. Domestication and classification of horses

Domestic horses have descended from the feral type of equines and belong to the *Equus f. caballus* subspecies of the Equidae family (horses, asses, and zebras) (Bokonyi 1987).

Horses passed through multiple domestication events which happened at least four times at different places, namely in Eastern Europe, Southern Russian steppe, Iberia and North Africa (Bennett 2008). It has been suggested that horse domestication started around 3000 BC (Ludwig *et al.* 2009; Cieslak *et al.* 2010). Figure 1.1 shows lineage divergence in 10 taxa of Equidae family including the domesticated horse (*Equus f. caballus)* and the Przewalski horse *(Equus f. przewalskii*). The Przewalski horse will be used as an out-group for comparison in this study.

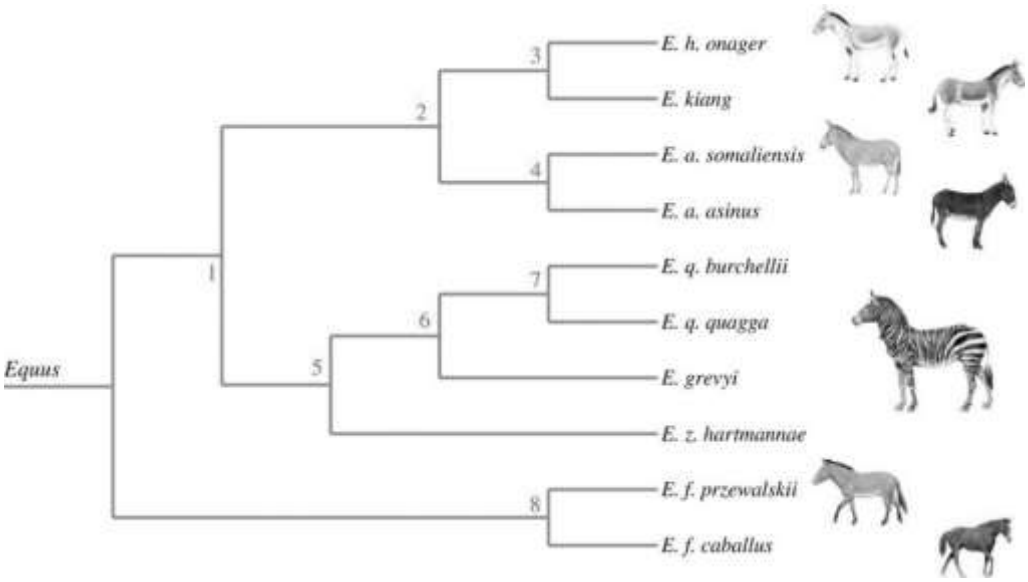


Figure 1.1. Lineage divergence in 10 taxa of Equidae family
Two horses, four zebras, two asses and two hemiones. Numbers are representing the internal nodes based on of 20 374 protein-coding genes phylogeny (Jónsson et al. 2014; Cucchi *et al.* 2017)

The domestic horse has 32 pairs of chromosomes including the sex chromosomes (2n= 64) (Rothfels *et al.* 1959). Domestic Animal Diversity Information System (DAD_IS 2011) reported approximately 800 horse breeds world-wide. Classification of horse breeds is based on different considerations. Generally, horses are classified based on the following major levels:

I. *By body phenotype:* This level discriminates horses into light, heavy horse breeds and ponies. Light horse breeds are small, fine-boned, fast riding and show horses (e.g. Arabian horse breed). Heavy horse breeds are huge, slow and represented mainly by the draft and labor horses (e.g. Dutch Heavy Draft breed or Nederland Trekpaard). Ponies often exhibit thicker manes and necks, as well as shorter legs and wider barrels (e.g. Exmoor pony). Ponies are known for their intelligence and they are currently used for leisure.

II. *By horse characteristics:* This level classifies horses either based on coat color pattern (e.g. spotted Appaloosa breed) or locomotion and gaiting patterns which is happening naturally without prior training or occurring after special exercises in certain breeds. The most common gaits are trot (e.g. Standardbred horse), pace (e.g. Icelandic horse), canter (e.g. Andalusian horse) and gallop (e.g. American Quarter horse).

III. *By horse usage or implementation:* some horse breeds are trained to perform specific sports (e.g. dressage, harness) in addition to land management and labor in constructions.

IV. *By the temperament:* this level includes three terms: hot, cold and warm blooded breeds. Horse breeds which are fall into the hot-blooded group (e.g. Arabian horses) are described as passionate and normally found in warm zones. Cold-blooded group includes horses which are known for their strength and calm nature and they are found in colder climates (e.g. Breton and Percheron horses). While warm-blooded horses have descended from selected cold-blooded horses with influence of hot-blooded ones (e.g. Hanoverian horses).

These classes can overlap each other, for example, hot and warm-blooded horses are mostly light horses (e.g. Arabian horses performing dressage). While the cold-blooded

breeds are mostly the heavy horses (e.g. Belgian draught horses used for land management).

This study will focus on characterization of Arabian horse breed in Syria.

1.2. Arabian horses

1.2.1. History and origin

The Arabian horse is one of the oldest light, hot-blooded horse breeds (Głażewska 2010). Historically, documents described the Arabian horse across the Fertile Crescent and Arab peninsula, where then widely spread through time in this region especially in Egypt, Arabian Gulf states, and Syria. In tracing the origin of the Arabian horse, researchers have not yet solved the dilemma of their foundation. Lately, the horse buried in a wooden coffin at Thebes in Egypt, 1500–1465 BC, which had only five lumbar vertebrae, as it is frequent with the modern Arabian breed, provided one of the oldest evidences of their existence (Schiettecatte & Zouache 2017). It is commonly known that Arabian nomadic breeders (Bedouins) tamed Arabian horses gradually, and refined the recent phenotypes to be adapted to scarcity of pastures, and rough environmental circumstances (WAHO 2015).

Arabian horse is considered as an international trans-boundary horse due to its blood accessibility in all continents through various horse breeds (Khadka 2015). In the 18th century, the horse breeding strategies in Europa and the USA concentrated mainly on having the best stallions to improve their local horse stocks, which required importing high quality stallions, mainly Arabian horses from the Middle East (Wallner *et al.* 2013). Consequently, Arabian horse westward expansion has a great influence on improving modern western horse breeds, for instance, Shagya, Thoroughbred, Lipizzan, Andalusian and Polish Arabian (Bowling *et al.* 2000; Bodó *et al.* 2005; Wallner *et al.* 2013).

1.2.2. Population structure and stud-books

Arabian horse breed is thought to have descended from five founder mares which are known as strains; namely Saglawi, Kahlawi, Hamdani, Obeyan and Hadban (Raswan *et al.* 1981; Mayouf *et al.* 2011). Rarely, stud-books report additional three strains which are Shweemat, Mounagii and Dahmaa. The strains designation system was shaped in the 7th century in the Arabian Peninsula and spread to the neighboring oriental empires and it is still used (Bennett 2008). In particular, Bedouins were careful in mating their horses, keeping the strains distinct and avoiding crossing with foreign horses, which did not belong to a well-known strain.

Stud-books monitor mating events and register the new offspring. Therefore, stud-books are very important for providing strains kinship information. In the western world, official authorities check the Arabian horse stud-books frequently and review them carefully. For verifying individual's relationship and population structure, different techniques based on DNA genotyping have been performed.

1.2.3. Arabian horses in Syria

1.2.3.1. Location and climatic conditions

Syria is located in the region of the Fertile Crescent (Figure 1.2). This region is considered as one of the putative domestication centers for many animal species including donkeys, pigs, sheep, goats and cattle (Larson & Fuller 2014). Sources mentioned that horse bones were found in archeological sites in the north-eastern part of Syria during the Akkad period (2350–2150 BC) (Anthony 2013). This suggests Syria as one of the oldest horse breeding regions. Therefore, it is of utmost interest to study the diversity among modern Syrian

horses. This study will focus on the main Arabian horse strains, which are maintained in Syria.

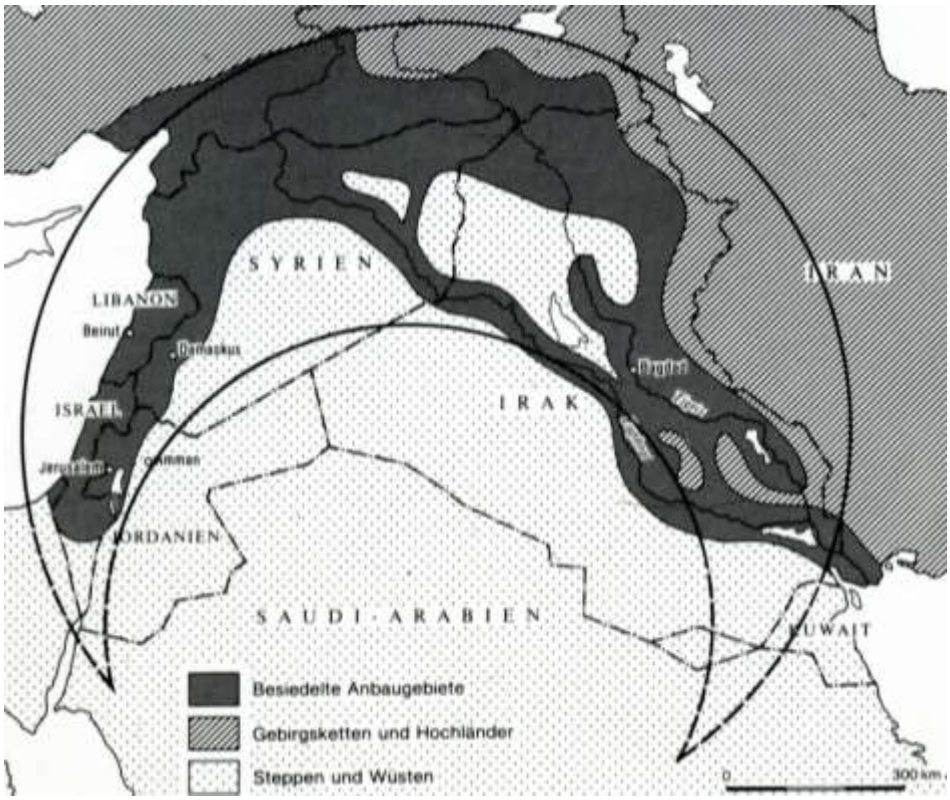


Figure 1.2. Fertile Crescent and its countries including Syria.
From (Hrouda & Bottéro 1991)

Syria's latitude and longitude are 35° 00' N and 38° 00' E, respectively. Generally, Syria (with its 14 governorates) is characterized by arid to semi-arid climate. Most rainfall is seen in January, February and March with an annual average of 884.3 mm in the western coasts, and decreasing till 135.1 mm in the eastern desert. The driest months are June, July and August in the whole country. The annual average temperature ranging in summer between 26.7 °C in the west and 32.6 °C in the east. In winter, the average temperature in the west is 11.6 °C and 6.7 °C in the east. An exception is the mountains where temperatures are much lower in summer and winter.

Livestock species (including Arabian horses) in Syria show strong adaptation and flexibility in living in different Syrian topographic regions (coastal plains, the highlands and the eastern plateau including the Syrian desert), in addition to resistance to diseases.

1.2.3.2. Registries and strains

According to the Ministry of Agriculture and Agrarian Reform database, Syria possesses two horse breeds. The first is namely Local Syrian horse breed which is not described or counted officially yet. The second is the purebred Arabian horse breed, which is registered and certified by the World Arabian Horse Organization (WAHO). Arabian horse registration requires being the offspring of two parents which had already been registered in the Syrian stud-book, given a unique identification number. The Arabian horses are marked with the identifier on the right side of the neck, and every approved horse gets a passport. The first Syrian Arabian horse stud-book has been published by the Arab horse office in 1989 and included the registries of 569 Arabian horses. In 2016, the total number of Arabian horses registered in Syria was 6189, which are distributed a in eight strains (AL Shaieb 2016) as shown in Table 1.1.

Table 1.1. Number of horses in eight Arabian horse strains in Syria, 2016.

Saglawi	Kahlawi	Hamdani	Obeyan	Mouanagii	Shweemat	Hadban	Dahmaa	Total
2907	1488	560	498	416	269	42	9	6189

Resource: (AL Shaieb 2016) January, Syrian Ministry of Agriculture and Agrarian Reform, Arab horse office.

The three major Arabian horse strains with the biggest population size are Saglawi, Khalawi and Hamdani (Figure 1.3). These strains are used in our study. The strains are characterized as follows:

I. *Saglawi strain*: The horses of this strain have been commonly described as lean, fine-boned horses, with relatively flat musculature, upright neck carriage, and a well-developed forehand. The common colors are gray, bay and chestnut. The tendency to white markings is known in the Saglawi strain.

II. *Kahlawi strain:* This type of strains is characterized by a more masculine, balanced and symmetrical appearance. The horses are mostly wider across the chest and the back. The most common colors are gray and chestnut. Kahlawi are heavier in weight and less lean in frame than the Saglawi.

III. *Hamdani strain:* animals are masculine, like the Kahlawi. Hamdani horses are heavier compared to the other strains. The head is straighter in profile with a strong back and more prominent withers. The common colors are gray and bay.

Figure 1.3. Three stallions representing the three Syrian Arabian horse strains
From the left to the right: Saglawi, Kahlawi and Hamdani. Copyright: Mr. Basel Jadaan

1.2.3.3. Breeding system

Syria has been suggested as a hot spot for Arabian horse breeding in the Middle East and the world (Khanshour & Cothran 2013). To date in desert, Bedouins keep their Arabian horses roaming freely, but mating under absolute control. While in urban and suburban

regions, Arabian horses are kept in large paddocks including stables for sleeping, receiving veterinary treatment and protection supported by strand electric or pipe fences.

Usually, traditions of Arabian horse breeding are kept under strict rules, especially with respect to mating, which could require breeders travelling over great distances looking for the appropriate stallions for compatibility with mares (Raswan *et al.* 1981; Mayouf *et al.* 2011). The main criteria for selecting a stallion are positive behavioral patterns, endurance performance and records, correct body conformation, reproductively and healthy progeny and above all it should be provided with recognized pedigree (Samper 2009).

When determining to breed an Arabian horse, the selection index and goals should be set clearly to produce either endurance horses or beauty show horses. Nevertheless, the history of Arabian horse breeding in Syria provides also examples of show horses performing endurance.

Local authorities in Syria realized that the Arabian horse population lost very important stallions without having enough offspring due to the strict traditions. Therefore, artificial insemination and embryo transfer were suggested to support Arabian horse industry, even though most of traditional breeders still prefer natural mating. Syrian Ministry of Agriculture and Agrarian Reform established the first center for artificial insemination and embryo transfer in 2010. The World Arabian Horse Organization (WAHO) permitted officially the center, but, it was destroyed before starting the work during the war in 2012.

Arabian horses in Syria are mainly used for transportation in suburbs, racing, and beauty competition events. The national income (in both, governmental and private sectors) benefits from exporting certified Arabian horses world-wide under very strict policies. Due

to its great cultural value and promising potential in the future Syrian economy, the welfare of Arabian horses in Syria is well protected by the social behavior and policies. Syrian Arabian horses belong to the national heritage.

Despite all that, Syrian Arabian horse breeding is having major obstacles represented by the traditional way of affiliation's documentation, poor characterization, degradation of some strains (e.g. Dahmaa and Hadban), as well as the absence of modern breeding programs. The recent crisis in the country stopped issuing stud-books for three years due to limited access to information of new offspring. Syrian authorities reported in 2017 that Syria lost approximately one third of the total number of Arabian Horses, where the number of missing horses was estimated to be up to 3000 (AL Shaieb 2017).

1.3. Phenotypic characterization of Arabian horses

The characterization of a population in a broad term comprises both the description of the phenotypic traits and genetic information (de Vicente *et al.* 2006). Phenotypic characterization is mainly a process of describing the quantitative and qualitative characteristics that can be observed in a given environment, e.g. production, performance, coat colors and morphological traits. Phenotypic characterization processes routinely carried out by using different methods (e.g descriptor lists of morphological traits) (de Vicente *et al.* 2006). Documentation of phenotypic data is highly important for developing knowledge about patterns distinguishing individuals and populations. This practice is facing challenges. Collecting information demands well planned frameworks and a lot of technical and logistic support (FAO 2012a).

1.3.1. Morphological traits

In horses, morphological traits are beneficial to characterize body conformation, which is used in identifying breeds, describing the growth rates and assessing orthopedic health. The reported heritability values of the equine morphological traits ranged between moderate to high (Table 1.2), which can be used efficiently as selection criteria. Selection for morphological traits led to produce horses suitable for different tasks (Brooks *et al.* 2010a), for example, in Arabian horses, which have been selected for either beauty or endurance (Sadek *et al.* 2006; Gharahveysi *et al.* 2010; Çilek 2012). Both of selection criteria led to a balance between different equine morphological traits (Ricard & Touvais 2007).

Table 1.2. Heritability (h^2) estimates for some equine body conformation traits reported in various horse breeds.

Trait	h^2	Horse breed	Reference
Back line length	0.41	Iranian Thoroughbred	(Bakhtiari & Heshmat 2009)
Body length	0.27	Iranian Arab	(Gharahveysi *et al.* 2008)
	0.72	Andalusian	(Molina *et al.* 1999)
Cannon bone circumference	0.05	Iranian Arab	(Gharahveysi *et al.* 2008)
	0.57	Friesian horse	(Pretorius *et al.* 2004)
Chest depth	0.14	Egyptian purebred	(Sadek 2006)
Chest girth	0.26	Iranian Arab	(Gharahveysi *et al.* 2008)
	0.66	Mangalarga	(Prado & Mota 2008)
Chest width	0.56	Andalusian	(Molina *et al.* 1999)
	0.13	Iranian Arab	(Gharahveysi *et al.* 2008)
Croup height	0.55	Egyptian purebred	(Sadek 2006)
Croup width	0.28	Haflinger	(Samoré *et al.* 1997)
Fore cannon length	0.13	Iranian Arab	(Gharahveysi *et al.* 2008)
Hind cannon length	0.29	Iranian Thoroughbred	(Bakhtiari & Heshmat 2009)
Neck girth	0.35	Egyptian purebred	(Sadek 2006)
Withers height	0.24	Murgese	(Dario *et al.* 2006)
	0.57	Iranian Arab	(Gharahveysi *et al.* 2008)

Some morphological traits describe the ability of adaptation to the surrounding environment. For instance, Arabian horses possess a comparatively small head with large nostrils for better breathing; lean arched neck and clean throat are adapted to blood, air and food passing through; high tail carriage supports better cooling of the body; well-sprung ribs and wide chest offers more lungs capacity; short backs (with only five instead of six lumbar vertebrae) and horizontal croup are advantageous for carrying weight for longer distances; and short cannons with strong hooves are adapted for walking and running on stones and sand (Wood & Jackson 1989; Mayouf *et al.* 2011; AHA 2015; USEF 2016).

There are several classes of morphological traits comprising head, neck and shoulders, chest, back and hip, legs and feet measurements (Sadek et al. 2006; Cervantes et al. 2009; Brooks et al. 2010a; Sánchez et al. 2013). These measurements vary between studies

according to the goal of the study and the nature of horse breeds. Morphological studies of Arabian horses used a panel of measurements ranged between four(Çilek 2012) to eleven (Sadek 2006). In large scale studies, wide variety of horse breeds (65 breeds including Arabian horse breed), were measured for 33 traits (Brooks *et al.* 2010a; Makvandi-Nejad *et al.* 2012). According to the FAO, at least 100 mature individuals (males and females) are required to be analyzed for morphometric body traits to reach a significance level of 0.05 (FAO 2012a). Variations are affected by age and sex (Çilek 2012). Other factors affecting morphological traits include feeding, care, and climate conditions. The genetic determinants of the equine morphological traits are almost unknown, and further intensive work is needed.

1.3.2. Endurance performance

Endurance played a vital role in the evolutionary history of humans and other species, because it enabled them to survive and sustain their lives under different conditions (Maffetone 2010). In horses, endurance is a trait of great economic value. Humans invested this ability to improve the labor capability of horses, as well as for equestrian sport events. According to Bergero et al. (2005), endurance performance of horses is a low-intensity long-term trial.

Endurance racing horses perform hard aerobic work, consume high levels of the stored energy, and involve slow twitching muscle fibers for sustaining speed over great distances (Rivero *et al.* 1993a; Rivero *et al.* 1993b). Endurance performance of horses (as in other species) is affected by technical, physiological, morphological, and psychological factors (Metayer *et al.* 2004). These factors are linked to each other, for instance, the technical

skills are not ideally performed without specific morphological traits (e.g. endurance requires small bodies and slow twitching-long muscles).

The interest of equestrian industry increased dramatically in the early of 19th century, particularly in the horses performing popular types of racing, such as endurance. Today, endurance is one of the most popular types of equestrian athletic performances. Most of the international equestrian endurance distances range between 30 and 160 km that can be run in one day, 200 km over 2 days, or 500 km over 5 days. As in other quantitative traits, evaluation of endurance performance usually requires objective measurements, such as the mean speed (Ricard & Touvais 2007). Depending on the breed and training program, endurance horses can race 160 km at a mean speed of 20 km/h (Votion *et al.* 2012). Horses are eliminated from endurance racing because of lameness and/or metabolic disorders (Adamu *et al.* 2014).

Horses vary in their ability to perform endurance, due to variability of genetic background, morphological differences, health conditions, and training programs. Arabian and Arabian-cross horses are the best competitors considering their stamina and endurance for long distance riding (Metzger *et al.* 2015). It is strongly thought that adaptation to extreme endurance exercise in Arabian horses is influenced by genetic factors (Ricard *et al.* 2017).

In Syria, Arabian horse endurance events became regular; twice a year. The strategies to produce endurance horses in Syria are based on different traditional aspects including housing management, mating, nutrition, and training programs. The endurance trait of Syrian Arabian horses is poorly documented, and not studied yet.

1.3.3. Coat color

Horses have a variability of coat colors and markings which played an important role during horse domestication events and primary selection (Rieder 2009). In Arabian horses, coat color is an eye-catching trait. The most common colors of Arabian horse are bay, grey, chestnut, and black. The markings (mainly white) or spotting patterns do not occur often in Arabian horse, and for a time they were undesirable (Mayouf *et al.* 2011). Colors and markings are documented in the registries and used for describing and distinguishing Arabian horses. Knowledge about coat colors enabled breeders to predict the color phenotypes of their future horses.

Equine genetic research in the early of the 20th century, investigated the inheritance mode of coat colors (Pearson & Lee 1900; Hurst 1906). Pathways of melanin and other pigments production are involved in coat color determination (Marklund *et al.* 1996; Rieder *et al.* 2001; Reissmann *et al.* 2016). Some equine lethal diseases are associated with color phenotypes, for example, lavender foal syndrome (Fanelli 2005) and lethal white foal syndrome (Rieder 2009). Furthermore, studies suggested that some mutations in coloration genes are involved in biochemical pathways that are potentially affecting the physical functions of horses (Marklund et al. 1996; Stachurska et al. 2007; Adrian 2013).

1.4. Molecular genetic characterization of Arabian horses

Genetics and genomics are used to characterize and distinguish populations. Genetic characterization stands for the detection of variation as a result of differences in either DNA sequences or genes or other modifying factors (de Vicente *et al.* 2006). The availability of the whole genome sequence of the horse provides bigger chances for the accurate and fine-scale genetic characterization of equine genome and traits.

1.4.1. Genetic diversity

Genetic diversity is the heritable variations within and between populations of organisms (WCMC 1992). Variation is essential for adapting to environmental changes and have been modulated under long-term and short-term evolutionary impacts of their specific ecosystems elements (Markert *et al.* 2010). Genetic variations occur due to neutral, deleterious or adaptive variants in DNA sequences, individual genes, or chromosomes (Toro & Caballero 2005). Genetic variations are expressed in diverse morphological, physiological and behavioral patterns between individuals and populations (Frankham *et al.* 2002).

Information about genetic diversity are crucial for the development of guidelines for sustainable breeding and setting conservation policies. Knowledge about the genetic diversity helps to obtain a measure for the evolutionary history of individuals and populations and estimate the value of the genetic resources (Toro & Caballero 2005; Tapio *et al.* 2010; Hasler *et al.* 2011; Medugorac *et al.* 2011).

In Arabian horses (like in other livestock species), the maintenance of the genetic diversity is important for sustainability of the breed. The loss of genetic diversity reduces the ability

to improve specific characteristics and the response to unpredictable environmental conditions. Furthermore, a loss of diversity could lead to the loss of desirable traits, which will have negative consequences on the relevant economy (Ehrlich 1988). Therefore, implementation of breeding priorities, conservation options and sustainable management strategies are necessarily should be based on reliable parameters of genetic diversity and population structure (Maudet et al. 2002; Tapio et al. 2010).

1.4.2. Measures of genetic diversity and population structure

The main forces those influence the genetic diversity of populations are genetic drift, mutations, gene flow, isolation, natural and artificial selection. To quantify genetic diversity within populations, geneticists developed various measures, which are mainly the genotypes and alleles frequencies, observed and expected heterozygosity, fixation indices, genetic distances, as well as clustering.

Populations can be described, fundamentally, in terms of its genotypes and alleles frequencies. Genotypes and alleles frequencies are fundamental measures of the genetic variations in a given population. The frequency of each of the three genotypes is the number of individuals in the population with that genotype (genotype numbers) divided by the total number of individuals in the population, noting that that total genotype frequencies sum to 1. Allele frequency describes how common an allele is in a population. The total number of all alleles in a given population is two times the number of individuals in the population (the horse is diploid).

Allele frequency is calculated by determining how many times the allele appears in the population then dividing by the total number of all alleles in the entire population.

$$\text{Frequency of allele i} = \frac{\text{Number of observed copies of allele i in the population}}{\text{total number of all alleles in the population}}$$

1.4.2.1. Genetic variation within population

Generally, the fraction of individuals in a population that are heterozygous for a particular locus (or the fraction of loci within an individual that are heterozygous) is defined as heterozygosity. Heterozygosity values measure the genetic variations within a population. High heterozygosity values indicate high genetic admixture of the studied population, while low heterozygosity can refer to the loss of the genetic potential. A high heterozygosity level has a great value with respect to the population's ability for adaptation to surrounding changes (Hartl *et al.* 1997).

Heterozygosity is often represented by observed heterozygosity (H_o) and expected heterozygosity (H_e). The observed heterozygosity represents the proportion of heterozygous individuals per locus that are observed in a given population. The expected heterozygosity can be calculated from allele frequencies that occur in a population. According to Nei (1973), the expected heterozygosity (H_e) is calculated using the formula:

$$\text{He} = 1 - \sum_{\text{i}=1}^{\text{k}} \text{pi}^2$$

Where Pi is the i^{th} allele frequency for the total number of alleles (K) in a single locus

It is challenging to define populations in terms of their distribution in different ecological patterns, under different evolutionary forces. F-statistics or fixation indices were

introduced by Wright (Wright 1951) which describe the genetic variation in within-subpopulation and between-subpopulations. The fixation indices are applicable to any population (Nagylaki 1998; Lenstra *et al.* 2012).

The key parameter F_{IS} (average subpopulation inbreeding coefficient) refers to the relation of heterozygosity of an individual to its subpopulation, and reveals the degree of inbreeding under random mating conditions within a subpopulation (Lenstra *et al.* 2012). Inbreeding decreases the heterozygosity within populations, while outbreeding increases genetic diversity. In homogenous populations where the individuals are related, F_{IS} values are expected to be positive. In contrast, in heterogeneous population where individuals are less related, F_{IS} values will be negative.

$$F_{IS} = (H_S - H_I) / H_S$$

Where H_S represents the expected heterozygosity within the studied subset, and H_I is the observed heterozygosity in each individual in the examined subpopulations.

Introducing unrelated individuals into a breeding line is more beneficial and saves the population from negative effects of inbreeding, which would reduce the fitness of the population (Leutenegger *et al.* 2003; Charlesworth & Willis 2009).

1.4.2.2. Genetic variation between populations

For assessing genetic differentiation between populations, Wright's fixation index F_{ST} which describes the between-subpopulations component is often used. While F_{IT} is the overall inbreeding coefficient of an individual relative to the total population, but it is not often used.

The F_{ST} measures the divergence value between subpopulations, relative to the total genetic variation of the entire population. F_{ST} is calculated using different ways (Lenstra *et al.* 2012). The original formula of Wright (1969) was for two-alleles locus, then it has been reformulated for multiple alleles (Nei 1973, 1977; Weir & Cockerham 1984):

$$F_{ST} = (H_T - H_S) / H_T$$

Where H_T represents the expected heterozygosity in the total set of populations, and H_S refers to the expected heterozygosity partitioned within subpopulations.

F_{ST} values range between 0 and 1. Generally, levels are evaluated using the standard quantitative scale: 0 to 0.05 means none or low genetic differentiation, 0.05 to 0.15 means moderate genetic differentiation, 0.15 to 0.25 for high genetic differentiation, above 0.25 for a very high genetic differentiation between populations (Wright 1978). Although studies showed that significant F_{ST} can be indicated even with the smaller values (Frankham *et al.* 2002; Lenstra *et al.* 2012).

Genetic distances have been developed in terms of genetic diversity and evolutionary forces (e.g. complete or partial isolation). Genetic distance is a term describes mathematical method that calculate the distance between the studied taxa (individuals, populations, or species) based on their genetic profile (Nei 1987; Ruane 1999; Laval *et al.* 2002). Several types of genetic distance methods have been invoked to assess genetic diversity between populations. Genetic distances can be visualized with heatmaps, dendrograms, or minimum spanning networks.

Nei *et al.* (1983) formulated his genetic distance (DA) to be performed in order to generate the topology of the examined genetic data. Nei's genetic distance method has been widely

adopted due to its ability to be applied to any taxa under the condition of sufficient data (Nei 1972, 1987).

$$DA = 1 - 1/r \sum_{j}^{r} \sum_{i}^{mj} \sqrt{Xij\ Yij}$$

Considering that X and Y are the studied populations. Where *Xij* and *Yij* are the frequencies of the *i*th allele at the *j*th locus in populations X and Y, respectively, and *mj* is the number of alleles at the *j*th locus. While *r* is the total number of loci.

Reynolds genetic distance is another method assumed that genetic differences in a population occur due to genetic drift without mutations. This method measures the probability that any two alleles, sampled at random (one from each individual), are identical copies of an ancestral allele. This method is developed to include the pedigree information (Reynolds *et al.* 1983).

The genetic distance is supposed to be linear with time, therefore, it has been used for estimating the genetic relationships between populations by construction of phylogenetic trees, for example, the neighbor joining (NJ) tree (Nei & Kumar 2000).

Additionally, clustering is used as a tool to classify the examined samples, clarify relationships between individuals, and determine the population structure. Clustering produces informative dendrograms that help to assign individuals to one or more populations. Clustering relies on arranging genetic data of individuals into groups based on their similar genetic patterns without prior information on the structure of population (Toro & Caballero 2005; Ojango *et al.* 2011). Clustering uses different algorithms based on

the cluster models which vary according to the nature of the data set and the intended use of the outcome.

Prior to clustering, a distance matrix should be constructed using one of the distance measures, then select the appropriate clustering linkage. For example, hierarchical clustering can be done using Manhattan distance (d_{man}). For a given pair of individuals, Manhattan distance measures the absolute differences between their SNPs using the formula:

$$d_{man}(X,Y) = \sum_{i=1}^{n} |Xi - Yi|$$

Where d_{man} is the distance between the individuals X (SNP_1,..., SNPn) and Y (SNP_1,..., SNPn). Noting that when using the three genotypes for each SNP, we code A, H, B as 0, 1, and 2, respectively (according to Manhattan distance, if Individual X has genotype A and individual Y has genotype H, then the distance is 1, and if Individual X has genotype A and individual Y has genotype B, then the distance is 2).

Genetic analysis softwares are currently integrating multiple clustering methods to infer the population structure, e.g. the STRUCTURE program (Pritchard *et al.* 2010).

1.4.3. Molecular markers for genetic diversity studies in horses

Molecular markers have been developed in light of many considerations in the last decades, including the throughput and costs of producing and applying them (Bernardo 2008).

Different molecular markers were used efficiently to characterize horse populations (e.g. Thoroughbred, Quarter horses), and to investigate their origins (Bowling *et al.* 2000;

Heaton *et al.* 2002; Cieslak *et al.* 2010; Williams *et al.* 2010; Hasler *et al.* 2011; Achilli *et al.* 2012; Petersen *et al.* 2013a).

The more frequently markers used in horse diversity studies are:

1.4.3.1. Mitochondrial markers

Hypotheses assume that mitochondria developed either from an anaerobic nucleus-bearing cell which requires an aerobic lifestyle developing its own O_2 metabolic system, or they have been descended (approximately 1.5 million years ago) from archaebacteria which can live with or without O_2, (Martin & Mentel 2010). Functionally, mitochondria are power generators of the living cells. They have an essential role in the metabolism providing the energy for all types of cells (Hanna & Nelson 1999; Ning *et al.* 2010).

The equine mitochondrium (like most of mammals) possesses a small (16.6 Kb), closed circular double-stranded DNA molecule. Typical mammalian inner mitochondrial membrane encodes 22 genes for tRNA, 13 subunits of multimeric proteins, 2 ribosomal RNA genes in addition to the displacement loop (D-loop) which is noncoding region. The D-loop with an approximately length of 1,100 base pairs (Xiufeng & Árnason 1994; Ishida *et al.* 1995; Jansen *et al.* 2002) has two hyper variable regions (HVR1 and HVR2), which are regions with highest mutation rates in the mitochondria. Due to lack of repair mechanisms and proofreading capabilities which makes mitochondria susceptible to base substitutions, leading to high mutation rates. Therefore, mitochondrial DNA accumulates mutations about 10-20 times faster than nuclear DNA (Wallace *et al.* 1987).

Figure 1.4 represents the structure of the mammalian mitochondrial molecule (mtDNA) and the two D-loop hyper variable regions; HVR1 and HVR2.

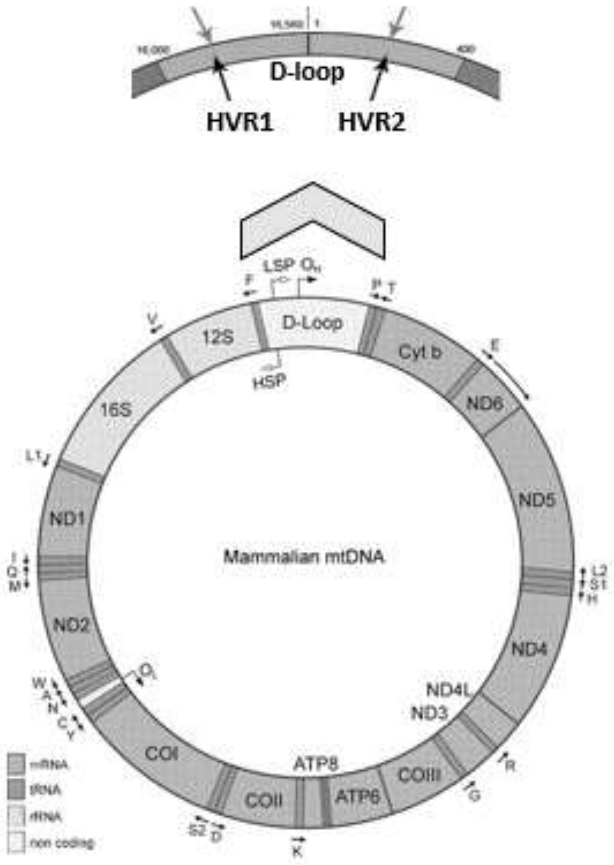


Figure 1.4. Structure of mammalian mtDNA and D-loop.
Modified from (Park & Larsson 2011)

Mitochondrial DNA (mtDNA) is inherited from the mother with the maternal oocyte and does not undergo recombination with paternal DNA. Therefore, offspring from a single mother share identical mtDNA sequences forming maternal haplotype (the arrangement of genetic polymorphisms within the mitochondrial genome which is inherited in the same pattern) (Moritz *et al.* 1987).

Mitochondrial DNA can easily be extracted and sequenced, which is a big advantage in maternal lineages discrimination studies (Harrison 1989).

Mitochondrial DNA has been used to study the current genetic diversity of the maternal lineages to give a clearer view of the female's historical contribution to the population architecture. Mitochondrial DNA (sequencing D-loop and the entire mitochondrial

molecule) has also been used for the validation of pedigrees (Forstén 1991; Bowling *et al.* 2000; Hill *et al.* 2002; Jansen *et al.* 2002; Kavar *et al.* 2002; Keyser-Tracqui *et al.* 2005; Royo *et al.* 2005; Cieslak *et al.* 2010; Achilli *et al.* 2012).

Arabian horses have been bred within strains (maternal lineages) and breeders kept them strictly separated (Raswan et al. 1981; WAHO 2015). Although Arabian horse pedigrees have been recorded (orally or in written form), many of the original pedigree registries of the beginning of breeding got lost. Now genetic information can be used to reconstruct the lost pedigree information. Mitochondrial haplotypes can be scrutinized across generations. Therefore, mitochondrial DNA is suited to show female ancestors in the Arabian horse population and detect current maternal lineages. The first study using mitochondrial DNA variants was performed on American Arabian horses (Bowling et al. 2000). The analysis revealed 27 mitochondrial haplotypes belonging to 18 dam lines which are descended from the Middle East. Recently, the whole mitochondrial genome of Arabian horse has been investigated, samples were taken from different Arabian horse populations around the world. Although many additional haplotypes were identified, the phylogenetic analysis did not clearly reveal Arabian horse strains (Khanshour & Cothran 2013).

Nevertheless, I like to mention that in few cases mitochondrial DNA within an individual is not uniform, but consists of different origin. In such cases heteroplasmy occurs (Harrison 1989; Hanna & Nelson 1999; Jansen *et al.* 2002). Therefore, it is recommended to include complementary markers from the chromosomes into the genetic diversity studies (FAO 2011).

1.4.3.2. Y chromosome markers

Assessment of the genetic diversity using the Y chromosome can support the bias of using only the mitochondrial DNA markers. The Y chromosome variation reveals the male contribution to the gene pool in a population.

Published studies on diversity in horses suffer limitations in identifying significant polymorphic variants in the Y chromosome (Lindgren *et al.* 2004). The horse Y chromosome was the third mapped after human and mouse. Raudsepp *et al.* (2004) presented the detailed map of the horse Y chromosome including the euchromatic region which comprises approximately 15 megabases (Mb) and lies in the long arm, where the pseudoautosomal region (PAR) is located terminally. The rest of the chromosome is predominantly heterochromatic. Figure 1.5 shows comparative illustrations representing the horse Y chromosome map and six mammalian species (including human). Additionally, Wallner *et al.* (2004) reported the first six Y chromosome-specific markers in the domestic horse which infer a single haplotype.

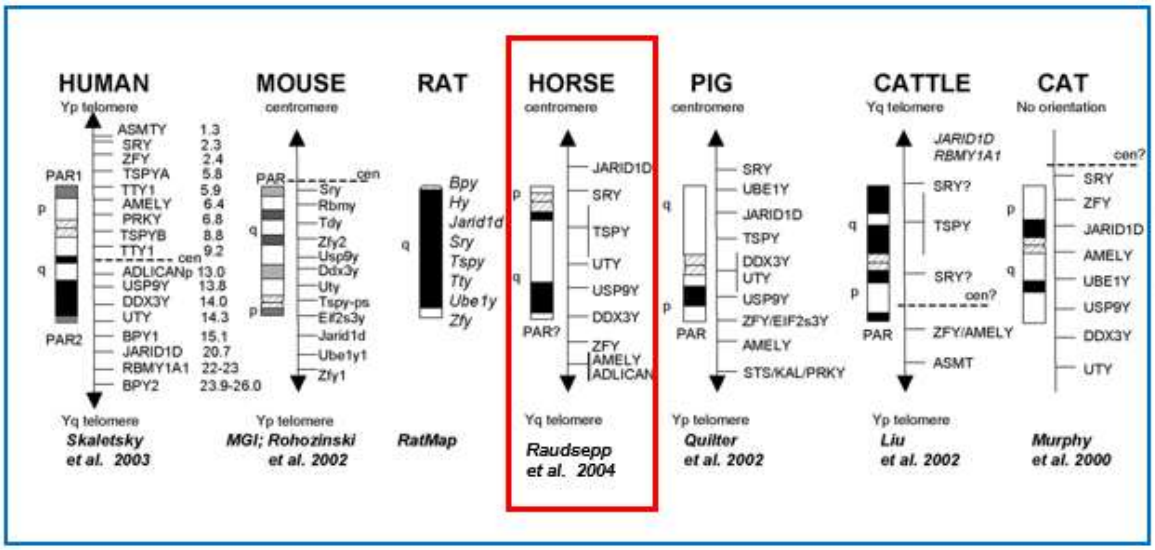


Figure 1.5. Comparative illustrations representing the horse Y chromosome map and six mammalian species (including human).
Modified from Raudsepp *et al.* (2004)

Findings over years displayed low variability and, therefore, suggest strongly that only few stallions participated in establishing the modern horse breeds (Groeneveld *et al.* 2010). Despite all difficulties faced by genetic marker on the Y chromosome, a study ofdiversity and tracing back the paternal horse lines, amplified long-range PCR (LRP) Y-chromosomal products from 18 stallions including 8 Lipizzan, 9 stallions representing distinct horse breeds (Icelandic horse, Thouroghbred, Warmblood (Trakehner), Quarter horse, Shetland pony, Shire horse, Shagya Araber, Norwegian Fjord horse and Arabian), as well as one Przewalski horse and one Shetland pony (Wallner *et al.* 2013). The samples were sequenced and the marker variants resulted in six haplotypes. Arabian horses were presented in the highest frequent haplotypes. Interestingly, the topology of the Y chromosome haplotypes in different horse breeds suggested that 48.3% of the studied horse breeds are descendent from Arabian founders.

1.4.3.3. Whole genome chromosomal markers

Molecular markers at the whole-genomic level include microsatellites and simple sequence repeats (SSRs, *alias* microsatellites), single nucleotide polymorphisms (SNPs), insertions and deletions (InDels) (Hu *et al.* 2015). Studies display that microsatellites and SNP markers have become the favorable markers for genetic diversity studies. In particular high density SNP arrays are attractive for genotyping animals for diversity studies.

Microsatellite markers

Microsatellites are short repeats which can be di-nucleotides (e.g. CA), tri-nucleotides (e.g. TCT), tetra-nucleotides (e.g. GATA), penta or hexa nucleotides (Ellegren 2004). Microsatellites replaced more complex genetic marker technologies such as RFLP and AFLP in the nineties of last century. Microsatellites have the advantage of being highly

polymorphic detecting often more than two alleles at one locus. Most microsatellites are neutral in their function; they are evenly distribution throughout the genome, co-dominantly inherited and had low mutation rates. First equine microsatellites were found in 1992 (Ellegren *et al.* 1992). According to recommendations of FAO, the standard microsatellite panel for diversity studies in horses should consist of 30 microsatellites (FAO 2011).

Using microsatellites, origins, genetic structure, and relationships of several horse breeds have been assessed (Cho 2006; Seyedabadi *et al.* 2006; Kakoi *et al.* 2007; Giacomoni *et al.* 2008; Leroy *et al.* 2009; Ling *et al.* 2011; Xu *et al.* 2012; Mackowski *et al.* 2015). Few studies have been carried out using microsatellites to investigate the genetic diversity and the population structure within (Georgescu *et al.* 2005; Mahrous *et al.* 2011; Mostafa *et al.* 2011; Sargious *et al.* 2014) and between (Khanshour *et al.* 2013) Arabian horse populations world-wide. In terms of microsatellites properties, goal of the study and nature of each horse breed, published equine studies used a panel of microsatellites ranges between seven (Seyedabadi *et al.* 2006) to thirty two microsatellites (Senju *et al.* 2017).

Even though the wide implementation of SSRs in genetic diversity studies, but SSRs have limitations with respect to complexity in designing and applying, time consuming, variations in findings from lab to lab (Vignal *et al.* 2002).

Single nucleotide polymorphism markers

SNP is a variation of a single nucleotide (A, C, G, or T) at a certain position in the DNA. SNPs have only two alleles. They are either coding or noncoding. During the last years, SNPs became the most attractive form of molecular markers. They are highly reliable, easy and cheap to genotype in particular in batches of SNP arrays.

The first whole genome sequence was generated for a female of the breed Thoroughbred (Twilight) designated EquCab2.0. Whole genome sequencing of the domestic horse enabled detecting approximately 1.5 million SNPs at once (Wade *et al.* 2009). Sequencing was done by the National Human Genome Research Institute (NHGRI) by shotgun sequencing. The project identified immediately ∼750,000 SNPs from Twilight and additional ∼ 400,000 SNPs from seven horses of different breeds (including Arabian horse breed). The latest genome assembly EquCab3 accession corresponds to GenBank Assembly ID GCA_002863925.1 (2018).

Using the comprehensive SNP database, the equine SNP array was produced for the effective genotyping of horses. For instance, the Equine SNP50 and SNP70 BeadChips (Illumina Inc., San Diego, USA) were developed with approximately 54K and 74K SNPs, respectively.

Table 1.3 represents briefly a comparison between the two equine arrays. For the development of the 70K chip more breeds were included in the chip design. This reduces ascertainment between breeds due to a bias of breeds represented on the chip versus breeds that are not directly (breeds that did not contribute to the selection of SNPs on the chip) presented there. SNP, could produce bias for example if some SNPs segregate in one population only, but not in the other. Such bias could affect the reliability of findings (Albrechtsen *et al.* 2010; Gärke *et al.* 2012; Fernández *et al.* 2013). Therefore, the genotyped SNP set plays a crucial role in minimizing false negative results.

Table 1.3. Characterizations of Equine SNP50 and SNP70 BeadChips (Illumina)

	Equine SNP50 Beadchip (2008)	**Equine SNP70 Beadchip (2011)**
Discovery breeds	Akhal-Teke, Andalusian, Arabian, Icelandic, Quarter Horse, Standardbred, Thoroughbred	Akhal-Teke, Andalusian, Arabian, Icelandic, Quarter Horse, Standardbred, Thoroughbred Plus: Twilight (Thoroughbred model horse) and RNA sequencing (RNA-seq) data
Nr. of SNPs	54,602	74,500 (including 53,500 from the SNP50)
Genome coverage	Entire equine genome, with the exception of the Y chromosome	Entire equine genome, with enhancement of the coverage on ECA20 and X chromosome and 2 SNPs on the Y chromosome
Gap size (density)	Average 1 SNP per 43.1 kb (few gaps larger than 500 kb)	Average of 1.5 SNPs per 50 kb

In 2017, a 2 million (MNEc2M) and a 670,000 (670K) SNP array (MNEc670k) were developed from over 23 million variants discovered from whole genome sequence of 24 breeds. The SNP chips were designed for genotyping at very high density and genotype imputation compared with the lower density chips (Schaefer et al. 2017).

The SNP arrays were used to evaluate genetic diversity and population structure of different breeds including Arabian, cross-Arabian, and several other horse populations world-wide (McCue *et al.* 2012; Petersen *et al.* 2013a; Kader *et al.* 2016). Furthermore, SNP arrays used to detect signatures of selection and to perform association studies for finding causal variants for equine-specific traits (Chen & Abecasis 2007; Wollstein *et al.* 2007; Petersen *et al.* 2013b; François *et al.* 2016).

Thorough quality control of genotypes generated with SNP arrays on genotype platforms is required prior performing the downstream statistical analyses. Criteria for for quality control include:

I. *GenTrain scores:* Genotype calling is based on a clustering algorithm for occuring genotypes which is known as GenTrain. Good clustering can be identified when the genotypes AA and AB do not overlapp. Every locus in the analyses is costumized and provided with a score (GenTrain score) which ranges between 0 and 1, where 1 indicates best clustering. Low GenTrain can happened due to incorrect placement of the cluster which will result in errors (Guo *et al.* 2014).

II. *Minor allele frequency (MAF):* This criterion provides information on rare alleles. The statistical power drops extremely for rare SNPs (Turner *et al.* 2011). Typically, SNPs with minor allele frequency (MAF) less than 1 or 5 % are exculded from the data set, depending on the size of the examined population. GWAS studies showed that common SNPs (MAF > 25%) give less false positive associations at different thresholds (Tabangin *et al.* 2009).

III. *Call rate: Genotyping* call rate measure is used for both SNPs and studied individiuals. The low call rate is happening due to poor DNA quality which leads to failure in genotyping (Anderson *et al.* 2010). SNPs with call rate less than 90% are usually excluded from further analysis (Turner *et al.* 2011). In addition, indidviduals with more than 10 % missing SNPs are excluded from analyses. The thresholds for individual and genotyping call rate are applied for autosomal SNPs. They differ for SNPs on the X and Y chromosomes (Wiggans *et al.* 2011; Cooper *et al.* 2013).

IV. *Unassigned SNPs:* the chromosomal position of the genotyped SNPs must be identified correctly. In the recently provided SNP BeadChips, SNPs are mostly

positionally verified. Luckily, methods have been developed which enable re-mapping SNPs to their precise chromosomal positions (Schmitt et al. 2010). SNPs that cannot be assigned to a unique positions are removed from the SNP set.

V. *Hardy-Weinberg Equilibrium (HWE):* checking for HWE is required for subsequent analyses. For genome wide association studies, for example, HWE is expected across most populations. Deviation from HWE can be due to genotyping errors or poulation stratification (Turner *et al.* 2011). Nonetheless, deviation from HWE has to be considered if lethal alleles are searched for.

1.4.4. Candidate genes for endurance performance

In classical animal breeding, endurance horse breeders would base their mating decision on pedigree information, attempting to have the progeny of the best performing ancestors (Harrison & Turrion-Gomez 2006). Using genomic information, the goal is to provide genetic predictors for the potential high endurance performance.

Endurance, as a complex quantitative trait, is regulated by multiple genes. Heritability for mean speed estimated in eight endurance horse breeds (including Arabian horses) vary considerably between 0.16 and 0.40 at distances 90 km and ≥ 120 km, respectively (Ricard & Touvais 2007). This makes the mean speed as one selection parameter for endurance performance.

Horse breeds that differ in their endurance performing also differ in many physiological and metabolic parameters (Bergero *et al.* 2005; Castejon *et al.* 2006; Joyner & Coyle 2008), morphology and gaiting (Metayer *et al.* 2004; Cottin *et al.* 2010), skeletal muscle fiber types (fast/low twitch fibers) and muscle fiber composition (Rivero *et al.* 1993b; Rivero & Barrey

2001). Pathways contributing to endurance performance provide a list of candidate genes that can be tested for association with this trait. Candidate genes are genes that likely contribute to a trait due to the known or predicted function of the gene product. The candidate gene approach is quick to perform and a comparatively cheap method (Patnala *et al.* 2013). Even though, it remains difficult to shorten the list of candidate genes that could be associated with the endurance trait, which is difficult to measure and likely many traits contribute to it.

Candidate genes counted for athletic and physical performance in human are approximately 230 genes (Bray *et al.* 2009; Schröder *et al.* 2011). Recent meta study tried to identify a common genetic profile specific to the endurance performance in human,, but it is still under review (Rankinen *et al.* 2016). In horses, although the development of equine genetics, only a small set of genes related to physical performance were genotyped in some horse breeds (Hill *et al.* 2010a; Silva *et al.* 2015).

In this study, we will focus on Arabian horses, which provide a valuable model to investigate genes (mitochondrial and nuclear) that have potential effect on the endurance trait.

1.4.4.1. Mitochondrial candidate genes

Mitochondrial DNA encodes for the structural subunits of the respiratory chain, which produces energy based on consumption of oxygen and liberation of carbon dioxide and water. The released energy (in the form of hydrogen ions H^+) is utilized in regulation of the adenosine triphosphate (ATP) formation process which is known as the oxidative phosphorylation (OXPHOS), which occurs entirely in mitochondria (Hatefi 1985).

Mitochondrial respiratory chain complex includes 13 essential proteins of the OXPHOS located in the inner mitochondrial membrane. These proteins belong to five complexes: complex I which includes NADH dehydrogenase subunits (encoded in seven mitochondrial genes), complex II with succinate dehydrogenase subunits, complex III which comprises ubiquinol-cytochrome c reductase complex subunits (one mitochondrial gene), complex IV of cytochrome c oxidase subunits (three mitochondrial genes), and complex V which include ATP synthase subunits (two mitochondrial genes) (Hanna & Nelson 1999; Kühlbrandt 2015). Mitochondrial genes lack introns and intergenic spaces are short or absent (Shokolenko & Alexeyev 2015).

For better view, Figure 1.6 displays the ATP synthesis pathway in the mitochondrial respiratory chain. All genes of the five complexes are mitochondrially encoded except proteins in the complex II, which are nuclear encoded (Eynon *et al.* 2011).

Only little information is available on polymorphisms in these genes in species that are selected for high athletic performance, like horses (Votion *et al.* 2012). The energy for endurance is mainly produced via ATP synthesis in the aerobic energy metabolism. Therefore, non-neutral polymorphisms in mtDNA could eventually be advantageous and improve endurance (Niemi & Majamaa 2005; Harrison & Turrion-Gomez 2006; Barrey 2010; Davie *et al.* 2011).

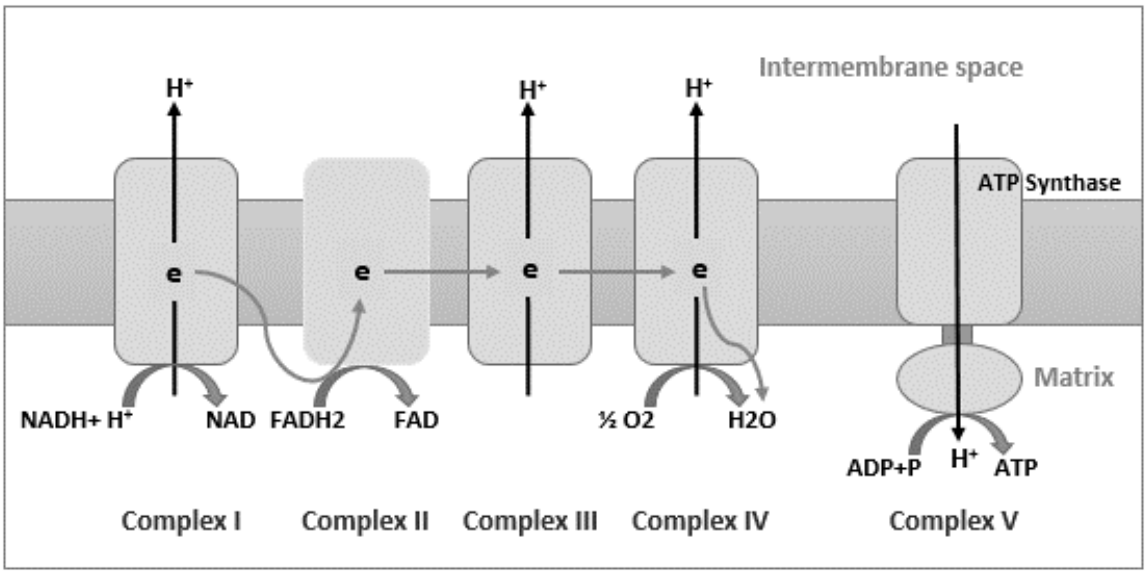


Figure 1.6. The five mitochondrial protein complexes of respiratory chain.
The complex II subunits are encoded in the nuclear genome (without frame).

Mitochondrial genes are strongly involved in the respiratory functions. Studies in human suggested that the utilizing of oxygen has a major effect on ATP synthesis in endurance performance and it is associated to mitochondrial DNA polymorphisms (Dionne *et al.* 1991; Niemi 2005).

Figure 1.7 represents the equine mitochondrially encoded genes (*MT-ATP6, MT-COX3, MT-ND4* and *MT-CYB*) which are potentially involved in the endurance performance.

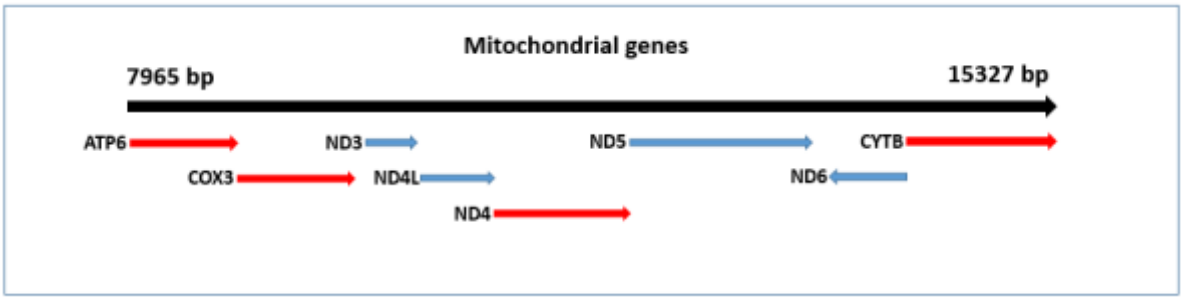


Figure 1.7. The mitochondrial genes *MT-ATP6, MT-COX3, MT-ND4* and *MT-CYB* (in red) expansion in the equine mitochondrial genome.

1.4.4.1.1. *MT-ND4*

The equine *MT-ND4* gene encodes *NADH-ubiquinone oxidoreductase chain 4* (ENSECAG00000027675) that spans positions 10,205 to 11,582 bp on the mitochondrial genome. *MT-ND4* belongs to the respiratory chain complex I, which encodes the enzyme NADH dehydrogenase. The transfer of electrons from NADH to ubiquinone that occurs during oxidative phosphorylation (OXPHOS) is mediated by the complex I enzyme (GeneCards). Studies on humans showed that exercise intolerance represented by metabolic disorders and heart failure is associated with a nonsense mutation (11832 G>A) which changed Tryptophan to a stop codon (Andreu *et al.* 1999; Gorman *et al.* 2015). To our knowledge, this gene has not examined yet for its potential contribution in endurance performance of horses.

1.4.4.1.2. *MT-CYB*

The equine *MT-CYB gene* encodes mitochondrial *Cytochrome b* (ENSECAG00000027669) that spans positions 14,188 to 15,327 bp on the mitochondrial genome. *MT-CYB* belongs to the respiratory chain complex III. *MT-CYB* contributes to the generation of energy by mediating the transfer of negative charged electrons from ubiquinol to cytochrome c (UniProt). Although *MT-CYB* mutations in humans are associated with wide spectrum of physical phenotypes (De Coo *et al.* 1999; Keightley *et al.* 2000), an interesting study showed

that a mutation in *MT-CYB* (Arg318Pro) is combined with sever disorders in the skeletal muscles (Blakely *et al.* 2005).

Till now, equine studies presented only the variants of *MT-CYB* in order to characterize the mitochondrial diversity of different horse breeds (Achilli *et al.* 2012; Yue *et al.* 2012; Sziszkosz *et al.* 2016).

1.4.4.1.3. *MT-COX3*

The equine *MT-COX3* gene encodes the *Cytochrome C Oxidase 3* (ENSECAG00000027672), which is located between positions 8,645 and 9,428 bp on the mitochondrial genome. It encodes the functional core of the complex IV (cytochrome c oxidase subunits) contributing to the respiratory electron transport, ATP synthesis, and heat production (GeneCards). Studies in humans suggest that mutations in this gene are associated with myopathy resulting mainly in muscular weakness (Horvath et al. 2005). In Norman horses, several cases were diagnosed with polysaccharide storage myopathy which could be a result of *MT-COX3* down regulation in muscles (Barrey et al. 2009).

1.4.4.1.4. *MT-ATP6*

The equine *MT-ATP6* gene encodes the *Adenosine triphosphate synthase subunit 6* (ENSECAG00000027671), spanning less than 1 Kb (7,965 - 8,645 bp) on the mitochondrial genome. *MT-ATP6* belongs to the complex V (ATP synthase subunits), and its biological role is represented by transporting of protons across the mitochondrial membrane to generate an electrochemical disturbance that stimulates ATP synthesis (GeneCards).

The transition variant 8794C>T in the humans *MT-ATP6* gene has been suggested to be associated with good performance in endurance running (Tanaka *et al.* 2004). In

Thoroughbred horses performing short distance racing, *MT-ATP6* variants linked in the same haplotype with additional variants of the genes *MT-CYB* and *MT-ND4* suggesting that the presence of any or a combination of these variants support enhanced racing performance (Harrison & Turrion-Gomez 2006). A study in Arabian horses provided evidence for an association between *MT-ATP6* variants and enhancement of work capacity and exercise tolerance (Ahmed *et al.* 2011).

1.4.4.2. Autosomal candidate genes

Studies in humans provided a list of autosomal genes, which are thought to be involved in the endurance performance. Some of the putative autosomal genes for endurance trait are listed in Table 1.4.

Table 1.4. Endurance-related candidate genes with their positions and biological functions

Gene	Ensembl ID	ECA	location	Biological Function
ACE	ENSECAG00000012910	11	15,829,612-15,849,932	Cell proliferation in bone marrow, heart muscle contraction, positive regulation of blood pressure
ACTN3	ENSECAG00000018961	12	26,511,750-26,524,992	Muscle contraction, skeletal muscle atrophy, positive regulation of skeletal muscle fiber growth, regulation of bone mineralization.
ADRB2	ENSECAG00000004810	14	28,966,506-28,967,756	Negative regulation of smooth muscle contraction, positive regulation of bone mineralization, heat generation.
AMPD1	ENSECAG00000014907	5	54,062,569-54,083,314	Instructions for producing an enzyme AMP deaminase. This enzyme is found in the skeletal muscles and contributing in producing energy.
CKM	ENSECAG00000022433	10	15,881,172-15,890,008	Response to heat stimulus above the optimal temperature for the organism, have an impact on cardiorespiratory endurance.
GYS1	ENSECAG00000021428	10	18,933,089-18,946,403	Heart development, glycogen biosynthetic process.
IGF1	ENSECAG00000010109	28	26,183,639-26,252,893	Bone mineralization, response to heat, skeletal muscle regeneration, regulation of T cells and smooth muscle cell proliferation, regulation of insulin-like growth factor receptor signaling pathway.
MSTN	ENSECAG00000021373	18	66,490,208 – 66,495,180	Growth, negative regulation of skeletal muscle satellite cell proliferation and differentiation, skeletal muscle fiber development.
PPARGC1A	ENSECAG00000009164	3	100,784,624-100,876,530	Cellular respiration, cellular response to oxidative stress, regulation of mitochondrial DNA metabolic activities, response to muscle activity stimulus.
VDR	ENSECAG00000015822	6	65,504,403-65,533,482	Skeletal system development, which regulates the bony framework formation to the mature structure.

ECA (*Equus caballus* autosome), Genes: *ACE (angiotensin I converting enzyme), ACTN3 (Actinin alpha 3), ADRB2 (β2-adrenergic receptor), AMPD1 (adenosine monophosphate deaminase 1), CKM (Creatine kinase, M-type), GYS1 (Glycogen synthase 1), IGF1 (Insulin like growth factor 1), MSTN (Myostatin), PPARGC1A (PPARG coactivator 1 alpha), and VDR (Vitamin D receptor)*.

1.4.4.2.1. *ACTN3*

The *Alpha-actinin skeletal muscle isoform 3* (*ACTN3*) (ENSECAG00000018961) encodes the equine α-actinin 3 protein. The gene is located between positions 26,511,750 and 26,524,992 bp on horse chromosome 12. *ACTN3* gene is composed of 21 exons (Figure 1.8).

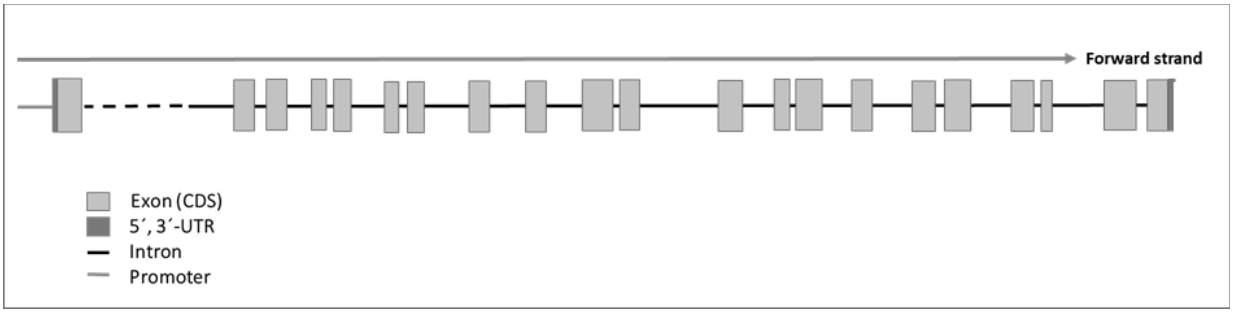


Figure 1.8. Structure of the *alpha-actinin-3 gene (ACTN3)* including the promoter, exons, introns and the untranslated regions (5' and 3' UTRs).

The encoded protein is primarily expressed in skeletal muscle, mainly in the fast twitch muscle fibers (type 2 muscle fibers) which are responsible for high speed and have a substantial role in the maintenance of muscle contraction (Yang et al. 2003; MacArthur & North 2004; Sjöblom et al. 2008). The *ACTN3* gene is highly conserved and its mutation rate is lower than average, which reflects the importance of its function.

In humans, the homozygosity for the nonsense polymorphism (R577X), which converts the Arginine at the position 577 of the protein into a stop codon, causing complete deficiency of the fast skeletal muscle fiber protein α-actinin-3 (Mata *et al.* 2012; Orysiak *et al.* 2014). In a study of endurance athletes, the XX genotype was over-represented (Yang *et al.* 2003). This suggested that ACTN3 variants may contribute to enhancing the endurance performance (Yang *et al.* 2003; Zanoteli *et al.* 2003). MacArthur *et al.* (2008) supported the human studies by mouse studies. Their analysis of knockout mouse muscle showed a shift in the properties from fast fibers towards slow fibers, increased activity of the metabolic

enzymes and better resistance to fatigue. In horses, *ACTN3* is suggested to affect muscle strength and insulin sensitivity which are related to endurance performance in different horse breeds (Gu et al. 2009; Thomas et al. 2014). Moreover, *ACTN3* polymorphisms showed positive association with muscle strength during training (Clarkson *et al.* 2005).

In Thoroughbreds horses, *ACTN3* is suggested to be involved in functions such as muscle strength and insulin sensitivity, which are related to endurance trait (Gu *et al.* 2009). Recently, Thomas *et al.* (2014) sequenced *ACTN3* of five horse breeds in Australia including Arabian, Thoroughbred, Standardbred, Shire, and Clydesdale, which are representing different patterns of physical phenotypes. This study identified 34 variants, twelve of them had been reported before in various horse breeds (Wade *et al.* 2009; Gu *et al.* 2010; Mata *et al.* 2012). Therefore, these variants were suggested as a biomarker for muscle performance in horses.

1.4.4.2.2. *ADRB2*

The equine *β2-adrenergic receptor* (ENSECAG00000004810) is located on the horse chromosome 14 spanning approximately 1.2 Kb (28,966,506-28,967,756 pb). The transcript *ADRB2-201* (ENSECAT00000004802) consists only of one exon (1,251 bp) (Figure 1.9).

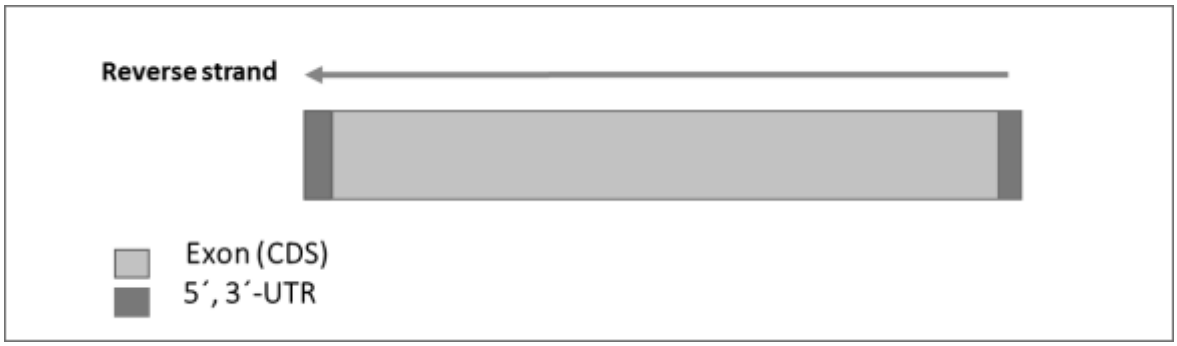


Figure 1.9. Structure of the *β2-adrenergic receptor* gene (*ADRB2*) including the promoter, exons, introns and the untranslated regions (5' and 3' UTRs).

ADRB2 is suggested to have biological effects represented by negative regulation of smooth muscle contraction, positive regulation of bone mineralization, as well as heat generation.

Researches in humans suggested that *ADRB2* polymorphisms are related to endurance and maintenance of the exercise (Wolfarth *et al.* 2007; Sarpeshkar & Bentley 2010).

Recently in Thoroughbreds, analysis of the expression pattern of *ADRB2*, before and after exercise, showed that ADRB2 responded to stress caused by exercise, and it was expressed in both; muscles and blood cells (Cho *et al.* 2015).

1.4.4.2.3. *MSTN*

The *Myostatin* (*MSTN*) gene (ENSECAG00000021373) is located between 66,490,208 and 66,495,180 bp on horse chromosome 18. *MSTN* comprises three exons. (Figure 1.10).

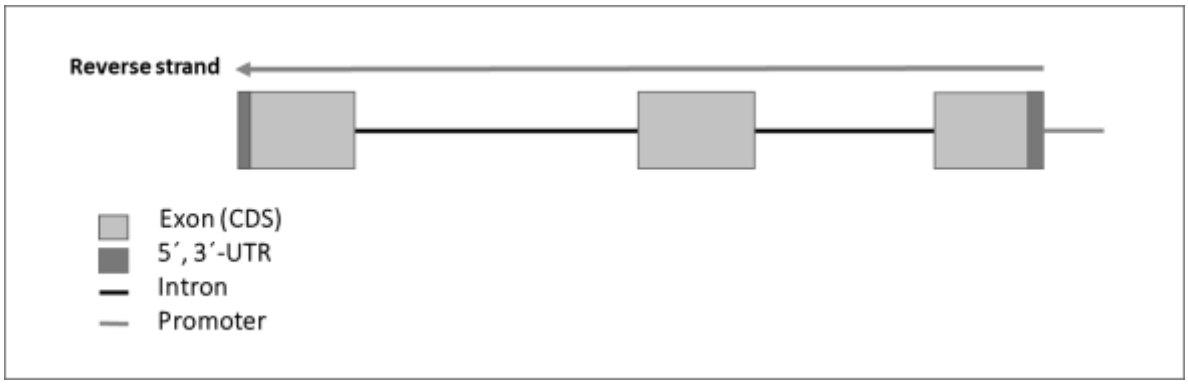


Figure 1.10. Structure of the *myostatin* gene (*MSTN*) including the promoter, exons, introns and the untranslated regions (5' and 3' UTRs).

MSTN encodes the growth differentiation factor 8 (GDF-8) which belongs to the TGF-β protein family affecting growth, differentiation and regulation of muscle proliferation as well as controlling the muscle fiber's growth (Carnac et al. 2006). Additionally, MSTN is involved in performance relevant functions such as regeneration of skeletal muscles, bone

formation, glucose metabolisms and adipocyte proliferation. In different species, mutations which result in an inhibition of MSTN cause increased muscle mass, for instance in Bully Whippet dogs and Belgian blue cattle (Mosher et al. 2007).

In Thoroughbred horses, association studies on racing stamina and optimum racing distance, identified a region on chromosome 18 and suggested significant association between the *MSTN* intronic variant 66493737C>T and optimum racing distance (Binns *et al.* 2010; Hill *et al.* 2010a; Tozaki *et al.* 2012). A 227 bp insertion of a SINE element into the *MSTN* promoter showed association with racing performance and muscle fiber proportions in Thoroughbred and Quarter Horse, respectively (Dall'Olio *et al.* 2014). The promoter insertion suggested to have an effect on the *MSTN* expression (Santagostino *et al.* 2015a). Further studies showed that the detected promoter insertion was in high linkage disequilibrium with the intronic variant 66493737C>T, mentioned above (Hill *et al.* 2010a; Hill *et al.* 2012a). Tozaki *et al.* (2011) identified downstream variants which have a significant association with racing distance and performance in Thoroughbred horses.

Recent studies in Arabian horses revealed variants in the exon 2, which defined a haplotype described as an Arabian horse-specific haplotype (Baron *et al.* 2012). Furthermore, Arabian horses showed more polymorphisms in the promoter region compared to other breeds (Stefaniuk *et al.* 2014). These findings implied that *MSTN* variants can be potential predictors of racing performance and morphological traits (Hill *et al.* 2010b; Tozaki *et al.* 2011; François *et al.* 2016).

1.4.4.2.4. *PPARGC1A*

The *Peroxisome proliferator-activated receptor-γ coactivator 1α* (ENSECAG00000009164) gene (PPARGC1A) is located on horse chromosome 3 spanning 91 Kb (100,784,624 - 100,876,530 bp); it consists of 13 exons (Figure 1.11).

The *PPARGC1A* gene is considered as a master regulator for cellular respiration, mitochondrial DNA metabolic processes, cellular response to oxidative stress, and response to muscle activity stimulus.

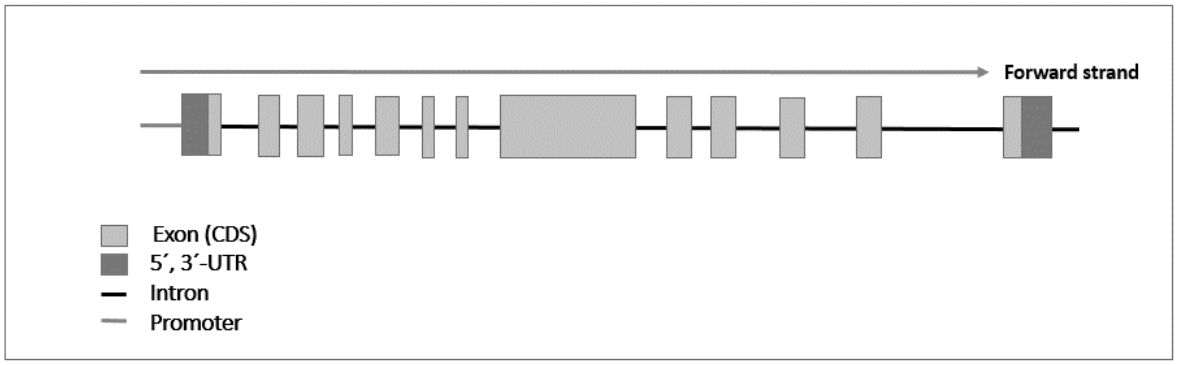


Figure 1.11. Structure of the *Peroxisome proliferator-activated receptor-γ coactivator 1α* gene (*PPARGC1A*) including the promoter, exons, introns and the untranslated regions (5' and 3' UTRs).

Human studies showed that *PPARGC1A* polymorphisms were more frequently found in the endurance elites. Therefore, the *PPARGC1A* gene considered to be a strong candidate for endurance performance (Lucia *et al.* 2005; Eynon *et al.* 2009; Ostrander *et al.* 2009; Tsianos *et al.* 2010).

In horses, a study confirmed that *PPARGC1A* regulates oxidative energy metabolism in equine skeletal muscle during exercise, showing a significant increase in mRNA expression in skeletal muscle following endurance exercise in Thoroughbreds (Eivers *et al.* 2012). A recent study analyzed the regulatory region of *PPARGC1A* and detected the polymorphisms

in 9 horse breeds (including Arabian horses). The SNP 100784525 C>G could be associated with strength and endurance (Polasik *et al.* 2017) .

1.4.4.2.5. *CALCA*

In horses, as in all mammals, bones, joints, and muscles are functionally connected. To perform endurance in the best way, it is important to protect bones, joints and muscles from potential disorders. A large and diverse set of factors affect musculoskeletal system, including calcitonin and vitamin D. Therefore, two genes are important to be investigated, CALCA and CYP2R1.

The *Calcitonin* gene (ENSECAG00000007825) is located on horse chromosome 7, spanning nearly 3 Kb (83,443,346 – 83,446,573 bp). Its transcript *CALCA-201* (ENSECAT00000008055) comprises 4 exons, three of them are coding for the protein (Figure 1.12).

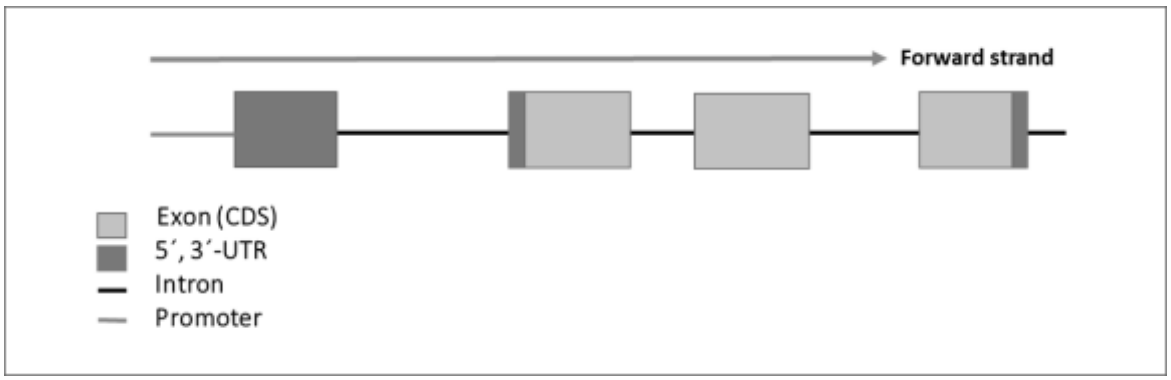


Figure 1.12. Structure of the *Calcitonin* gene (*CALCA*) including the promoter, exons, introns and the untranslated regions (5' and 3' UTRs).

Calcitonin inhibits the osteoclast activity in bones, and it promotes calcium loss from skeleton during periods of calcium mobilization, such as pregnancy and lactation (Goodman 2010).

In human, association studies suggested that bone resorption by osteoclasts is down regulated by Calcitonin (Meleleo & Picciarelli 2016). In mice, CALCA-deficient mice *CALCA* (-/) shows an increase of bones mass (Schinke et al. 2004). The deficiency also associated with an increase of bone resorption with age, where bones become more fragile and fractures occur most often in the deficient mice (Huebner *et al.* 2006).

1.4.4.2.6. *CYP2R1*

The gene encoding *Cytochrome P450 family 2 subfamily R member 1* (ENSECAG00000007597), is located on horse chromosome 7, spanning 148 Kb (83,383,774 - 83,532,095 bp). Its transcript *CYP2R1-201* (ENSECAT00000007831) has 5 exons (Figure 1.13).

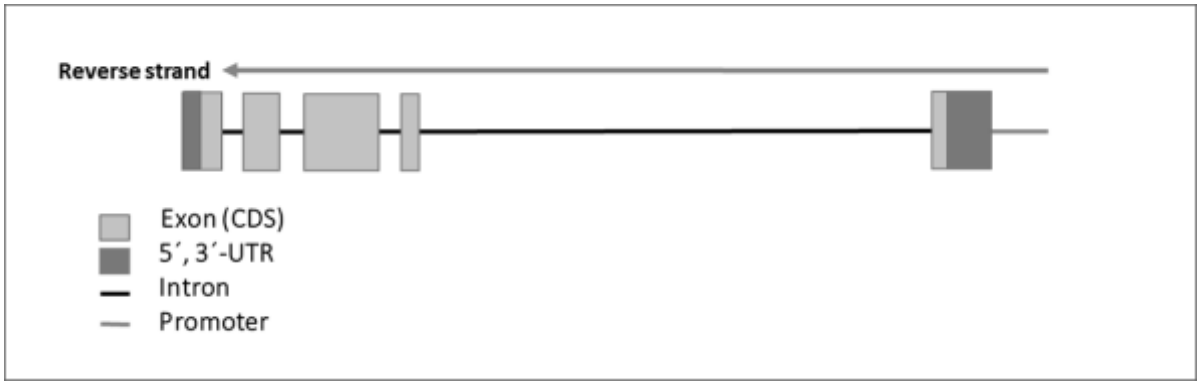


Figure 1.13. Structure of the *Cytochrome P450 family 2 subfamily R member 1* gene (*CYP2R1*) including the promoter, exons, introns and the untranslated regions (5' and 3' UTRs).

The CYP2R1 is an enzyme, which catalyzes many reactions involved in the synthesis of cholesterol, steroids and other lipids. This enzyme is a microsomal vitamin D hydroxylase, which converts vitamin D into the active ligand for the vitamin D receptor. Vitamin D is essential in calcium homeostasis and in the development and maintenance of the skeleton.

In human studies, a mutation in this gene has been associated with vitamin D deficiency. The deficiency in vitamin D may lead to bone loss, growth problems, and muscle weakness, causing falls and fractures (Lips & van Schoor 2011).

Both genes, *CALCA* and *CYP2R1,* are not tested yet in Arabian horses, or any other horse breed, for their effect on endurance. It will be of great interest to identify variants of these genes in Arabian horses, particularly those excluded from endurance races due to skeletal disturbances (e.g. lameness).

1.4.5. Genome-wide association studies (GWAS)

Investigating the causal relationship between phenotypes and genetic variants remains one of the most challenging barriers in human and livestock genetics. In light of the genome-wide SNPs identification and the complexity of the quantitative traits (Collins *et al.* 1999; Zhang *et al.* 2012), genome-wide association study (GWAS) with dense markers are considered as a powerful tool to discover the underlying genes for the examined traits (Korte & Farlow 2013).

GWAS is the examination of a genome-wide set of genetic variants in a group of Individuals to see if any genomic region is associated with a relevant trait (Hirschhorn & Daly 2005; Zhang *et al.* 2012). GWAS is relying on the linkage disequilibrium (LD) of causal and non-causal alleles which occurs more or less frequently in a population than would be expected under random conditions (Pearson & Manolio 2008; Ojango *et al.* 2011; Visscher *et al.* 2012).

Therefore, contrary to the candidate gene approach (where genes are selected based on their known effects on a trait), GWAS does not require prior-knowledge about the genomic regions contributing to the phenotype or pedigrees of the tested individuals (Visscher *et al.* 2012).

GWAS is performed using several models (e.g. linear model) to test the significance of association between every SNPs and the relevant trait(s) by comparing with a null hypothesis (no association is existing). GWAS might fail due to false positive or negative associations. Therefore, an appropriate sample size, refined phenotypes, and well-distributed - high quality genotypes (criteria for quality control are mentioned previously), are the primary prerequisite for successful GWAS (Laurie *et al.* 2010; Turner *et al.* 2011). Figure 1.14 represents schematic work illustrating GWAS.

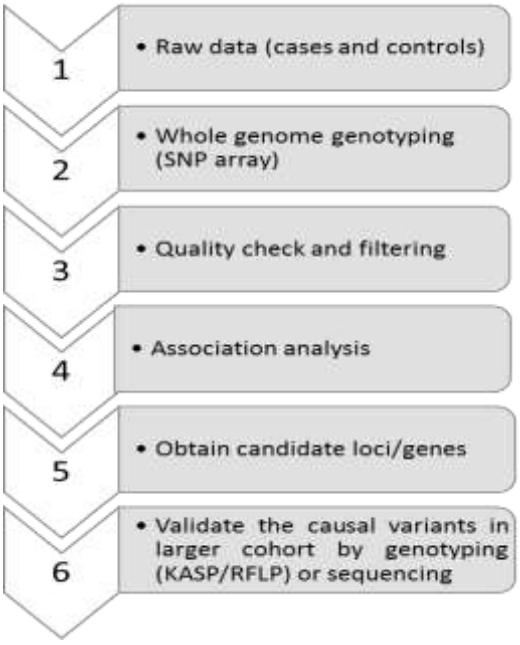


Figure 1.14. Schematic representation of the GWAS and the expected output.

By far, several GWAS studies in horses have been reported, particularly with using equine SNP arrays. One of the most remarkable results of a GWAS performed in Thoroughbreds

was the detection of a chromosomal region close to the *MSTN* gene on chromosome 18 that was associated with racing trait (Binns *et al.* 2010; Hill *et al.* 2010c; Petersen *et al.* 2013b). Additional GWAS studies have been performed for the variability of performances patterns in several horse breeds such as jumping in Hanoverian and Anglo-Arab horses (Schröder *et al.* 2012; Ricard *et al.* 2013; Brard & Ricard 2015). Other GWAS studies have identified SNPs associated with equine morphological traits (e.g. withers height, body length and back line length) in various horse breeds such as Franches-Montagnes, German Warmblood, and Quarter horses (Signer-Hasler *et al.* 2012; Tetens *et al.* 2013; Meira *et al.* 2014).

These studies and others have suggested relevant candidate genes like for example *LCORL (ligand dependent nuclear receptor corepressor like*) and *NCAPG (non-SMC condensin I complex, subunit G*) on chromosome 3 to be associated with withers height, as well as *ZFAT (zinc finger and AT-hook domain containing)* on chromosome 9 to be associated with back line length (Makvandi-Nejad *et al.* 2012; Tetens *et al.* 2013; Staiger *et al.* 2016).

Furthermore, in the equine disease field GWAS was performed to identify chromosomal regions associated with lavender foal syndrome in Arabian horse (Brooks *et al.* 2010b), and with the fatal foal immunodeficiency syndrome in Fell and Dales ponies (Fox-Clipsham *et al.* 2011). On the long term, GWAS results will contribute to find biomarkers for the important equine traits. The use of genetic markers allows for predicting offspring genotypes and phenotypes in earlier life (Gorbach *et al.* 2010). An overview on notably GWAS results obtained for various horse traits is given in Table 1.5.

Table 1.5. Some genes identified by GWAS for equine traits/diseases in various horse breeds

Horse breed	N	Trait / disease	Gene	ECA	Reference
Thoroughbred	118	Racing distance	*MSTN*	18	(Hill *et al.* 2010c)
Friesian dwarf horses	20	Dwarfism	*PROP1*	14	(Orr *et al.* 2010)
Arabian horse	36	Lavender foal syndrome	*MYO5A*	1	(Brooks *et al.* 2010b)
Fell and Dales ponies	51	Fatal foal immunodeficiency syndrome	*SLC5A3*	26	(Fox-Clipsham *et al.* 2011)
Hanoverian	115	Jumping	*PAPSS2* *NRAP* *MYL2* *TRHR* *TBX4* *GABPA*	1 1 8 9 11 26	(Schröder *et al.* 2012)
Franches-Montagnes	1,077	Withers height Croup length Back line length	*LCORL* *NCAPG* *ZFAT*	3 3 9	(Signer-Hasler *et al.* 2012)
Arabian and German Warmblood Horses		Equine guttural pouch tympany (GPT)	*SLC39A8* *SLC30A6*	3 15	(Metzger *et al.* 2012)
Holsteiner Oldenburger Hanoverian Trakehner	782	Withers height	*LCORL* *NCAPG*	3 3	(Tetens *et al.* 2013)
33 international breeds	744	Draft Gaiting	*MSTN* *DMRT3*	18 23	(Petersen *et al.* 2013b)
German Warmblood Horses	144	Equine recurrent uveitis	*CRYGA-CRYG* *IL-17A* *IL-17F*	18 20 20	(Kulbrock *et al.* 2013)
Quarter Horse	184	Weight Rump length	*WWOX* *AVPR1A*	3 6	(Meira et al. 2014)
Selle Francais Anglo-Arab	999	Jumping	*RYR2*	1	(Brard & Ricard 2015)

N: number of animals in the study, ECA: equine chromosome autosome

2. Objectives of the study

Syria is one of the least-studied major domestication spots of the world. Despite the high importance of Arabian horse, which is considered as a national treasure, studies about the evolutionary history, population structure, genetic diversity, and phenotypic traits are still few. Although the Arabian horse breed is not on the list of endangered breeds, the unstable conditions in the Middle East, particularly in Syria, raised the need for documentation, research and preservation actions. According to FAO (2012b), the Syrian livestock breeds including horses suffer poor phenotypic and genotypic characterization.

In light of these conditions, in the present study a total of 200 Arabian horses representing the three major strains in Syria, were sampled from five distinct governorates in Syria, The data used include pedigrees back to three generations, twelve morphological traits, and speed records of national endurance races. The study had following goals:

- Determining, genetically, the maternal lineages of the Syrian Arabian horses and pedigrees validation in the three major Arabian horse strains (Saglawi, Kahlawi and Hamdani) using the mitochondrial D-loop region as a genetic marker,
- Providing better knowledge concerning the level of genetic diversity within and between the major strains and their relatedness using genome-wide genetic markers including equine microsatellites and the equine SNP array,
- Presenting a Genome-wide association study (GWAS) of twelve morphological traits in Arabian horses,
- Identification of variants in mitochondrial (*MT-ND4, MT-COX3* and *MT-CYB*) and autosomal (*ACTN3* and *MSTN*) candidate genes with potential contribution in

endurance trait to examine their effects on low and high endurance performance of Syrian Arabian horses.

Jointly, the findings of this study can be considered as key facts for further research about Arabian horses in Syria. This knowledge can be applied in drawing better breeding strategies that can support the future Arabian horse industry in Syria.

3. Materials and methods

3.1. Materials

3.1.1. Chemicals

The chemicals and solutions which were used for DNA extraction, and molecular genotyping are shown in Table 3.1.

Table 3.1. Chemicals, chemical solutions, buffers and kits used for the genotyping

Chemicals	Manufacture
Agarose	SERVA GmbH (Heidelberg, Germany)
Buffer B without $MgCl_2$	PROMEGA (Mannheim, Germany)
DNA-Marker	Sigma-Genosys (Steinheim, Germany)
dNTP	Carl Roth GmbH (Karlsruhe, Germany)
EDTA	Carl Roth GmbH (Karlsruhe, Germany)
Ethanol (70% and 99%)	Carl Roth GmbH (Karlsruhe, Germany)
EthidiumBromide	Carl Roth GmbH (Karlsruhe, Germany)
FirePol 5 U/µl	Solis BioDyne OÜ (Estonia)
Gel	AppliChem (Darmstadt, Germany)
Gel extraction kit (Gene JET)	Thermoscientific (Germany)
Go Taq 5X without Mg^{+2} (5X)	Solis BioDyne OÜ (Estonia)
GoTaq flexi 5U/µl	PROMEGA (Mannheim, Germany)
Isopropanol	Carl Roth GmbH (Karlsruhe, Germany)
KASP assay mix	LGC Genomics Ltd. (UK)
KASP reaction mix	LGC Genomics Ltd. (UK)
M13 -700/800 (Microsatellites genotyping)	OPERON (Spain)
Magnesium acetate-tetrahydrate	Carl Roth GmbH (Karlsruhe, Germany)
Magnesium chloride ($MgCl_2$)	Carl Roth GmbH (Karlsruhe, Germany)
NaCl	Carl Roth GmbH (Karlsruhe, Germany)
PCR reaction buffer	Genaxxon BioScience GmbH (Germany)
Polyacrylamide gel Matrix 6.5%	LI-COR Biosciences (Nebraska, USA)
Premix for Sequencing PCR (B.D. Terminator)	Applied Biosystem (USA)
Primer, F, 5 µM (Microsatellites genotyping)	Invitrogen AG (USA)
Primers for genotyping polymorphisms	Carl Roth GmbH (Karlsruhe, Germany)
Protenase K	AppliChem GmbH (Darmstadt, Germany)
Puregene Core Kit	QIAGENE (Hilden, Germany)
SDS	Carl Roth GmbH (Karlsruhe, Germany)
Sequencing buffer 5X (Big Dye Terminator)	KingsLand Grange (UK)
Sodium acetate	Carl Roth GmbH (Karlsruhe, Germany)
Sodium hydroxide	Carl Roth GmbH (Karlsruhe, Germany)

Chemicals	Manufacture
T1 buffer	Macherey-Nagel (Germany)
Taq DNA polymerase	Genaxxon BioScience GmbH (Germany)
TEMED(N,N,N,N-Tetra methylamine)	Sigma-Genosys (Steinheim, Germany)
Tris	Carl Roth GmbH (Karlsruhe, Germany)

3.1.2. Instruments

Descriptive instruments and devices which were used in the variable genotyping techniques in this study are shown in Table 3.2.

Table 3.2. Equipment used for the genotyping

Instrument	Model	Manufacture
Bio Imaging System	GENE GENIUS	VWR international (Germany)
Centrifuge apparatus	Universal 5804R 5415C Universal 320 Mikro Centrifuger	Eppendorf (Germany) Eppendorf (Germany) HETTICH (Germany) ROTH (Taiwan)
Digital graphic printer	Up-D895	VWR International (Germany)
Digital timer		Carl Roth GmbH (Germany)
DNA Analyzer	LI-COR 4300	Licor Biosciences (USA)
Electrophoresis power supply	GPS (200/400)	Pharmacia GmbH (Germany)
Fine weigh scale	Horst Schirmer	Sartorius Fachhändler (Germany)
Freezer (- 20°C)	-	LIEBHERR (Germany)
Fridge (4°C)	-	Siemens (Germany) LIEBHERR (Germany) AEG (Germany)
Gel Electrophoresis	Horizontal	SERVA Electrophoresis (Germany)
Gene JET purification columns and collection tubes	-	Thermoscientific (Germany)
Ice machine	-	Ziegra (Germany)
Microwave	R-3V10	SHARP EU (Germany)
Mixer	VortexGenie2	Scientific Industries, Inc. (USA)
PCR plates	96, semi-skirted	Brinkmann Instruments, Inc. (USA)
PCR reaction tube	1.5 ml	Carl Roth GmbH (Germany)
Low volume PCR plates	4 Titude / Frame Star Fast Plates 96	Eppendorf (Germany)

Instrument	Model	Manufacture
Real Time PCR system thermal cycling block	StepOnePlus	Applied Biosystem (USA)
Sample box	12x12	Carl Roth GmbH (Germany)
Sequencer + monitor	ABI PRISM 310	Applied Biosystem (USA)
Spectrophotometer	ND-1000	Nanodrop technologies (Germany)
T Gradient Thermo Cycler	-	Biometra Ltd (Goettingen, Germany)
Thermo mixer	-	Eppendorf (Germany)
Thermoblock	TBI	Biometra Ltd (Goettingen, Germany)
Thermocycler	-	Biometra Ltd (Goettingen, Germany)
Ultra Violet Spectro	Gene Genius	VWR International (Germany)
Vortex	Vortex Gene2	Scientific industries (USA)

3.1.3. Arabian horse data

3.1.3.1. Animals

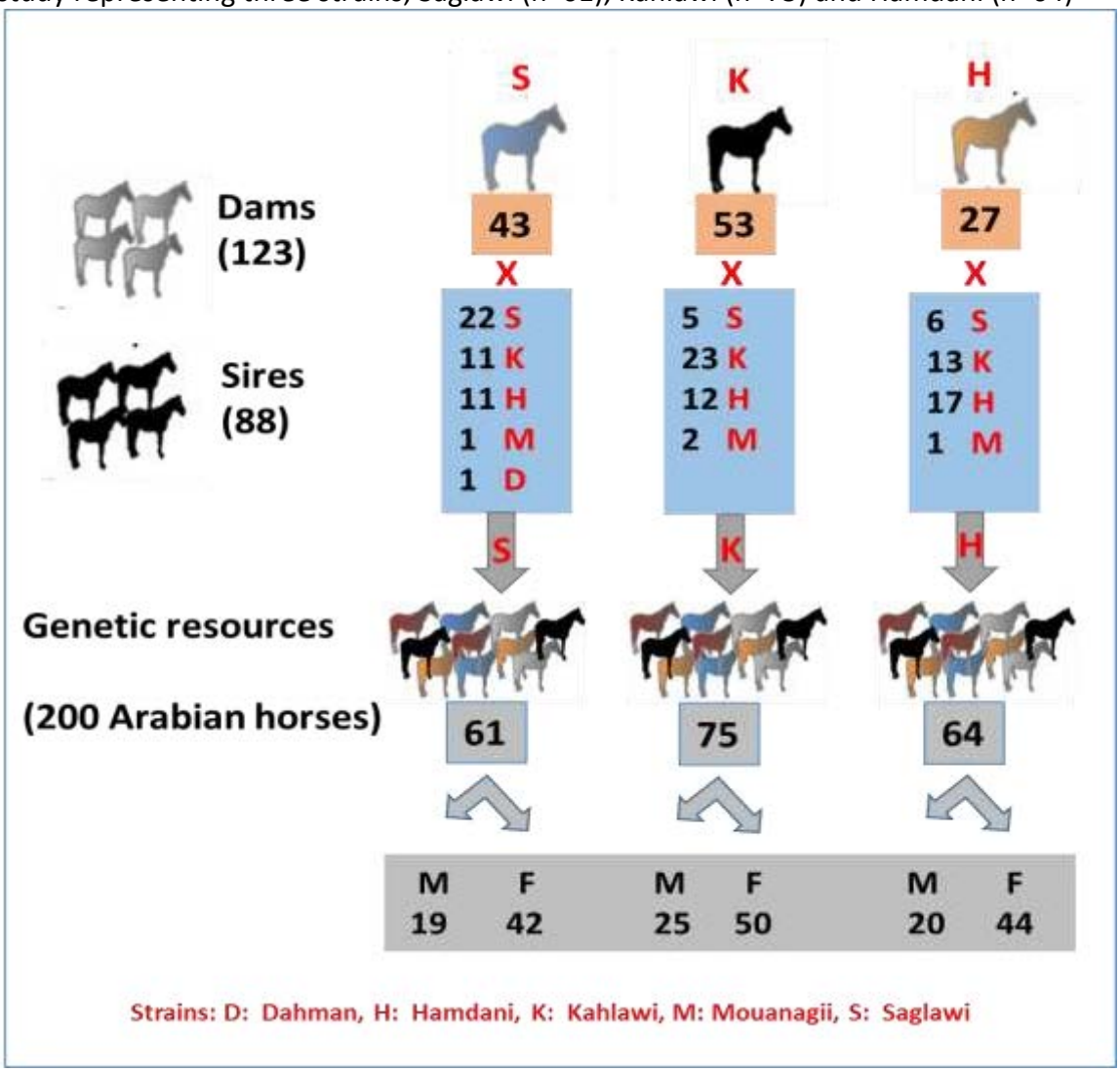
A total number of 200 Syrian Arabian horses (136 females and 64 males) representing three strains, namely Saglawi (n= 61), Kahlawi (n= 75) and Hamdani (n=64) descendant from 123 dams and 88 sires, born between 1986 and 2013 were sampled. Horses originated from five governance regions of Syria (Damascus; n=179 and Dara; n=3 in south, Hama; n=4 and Hims; n=2 in the middle-west and Al Hasakah in the north-east; n=12). Sampling overviews as well as the origin of horses are given in Table 3.3, and Figures 3.1 and 3.2.

All information and pedigree data were verified by the Syrian national stud-book.

Table 3.3. Origin, strains, number and sexes of 200 Syrian Arabian horses examined in this study

Region	Total			**Strain**				
	number		Saglawi		Kahlawi		Hamdani	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**
Al Hasakah	1	11	1	6	-	3	-	2
Damascus	61	118	17	33	24	45	20	40
Dara	-	3	-	1	-	-	-	2
Hama	1	3	1	2	-	1	-	-
Hims	1	1	-	-	1	1	-	-
Total	64	136	19	42	25	50	20	44

Figure 3.1. Schematic representation of the Syrian Arabian horses (n= 200) sampled for this study representing three strains; Saglawi (n=61), Kahlawi (n=75) and Hamdani (n=64)



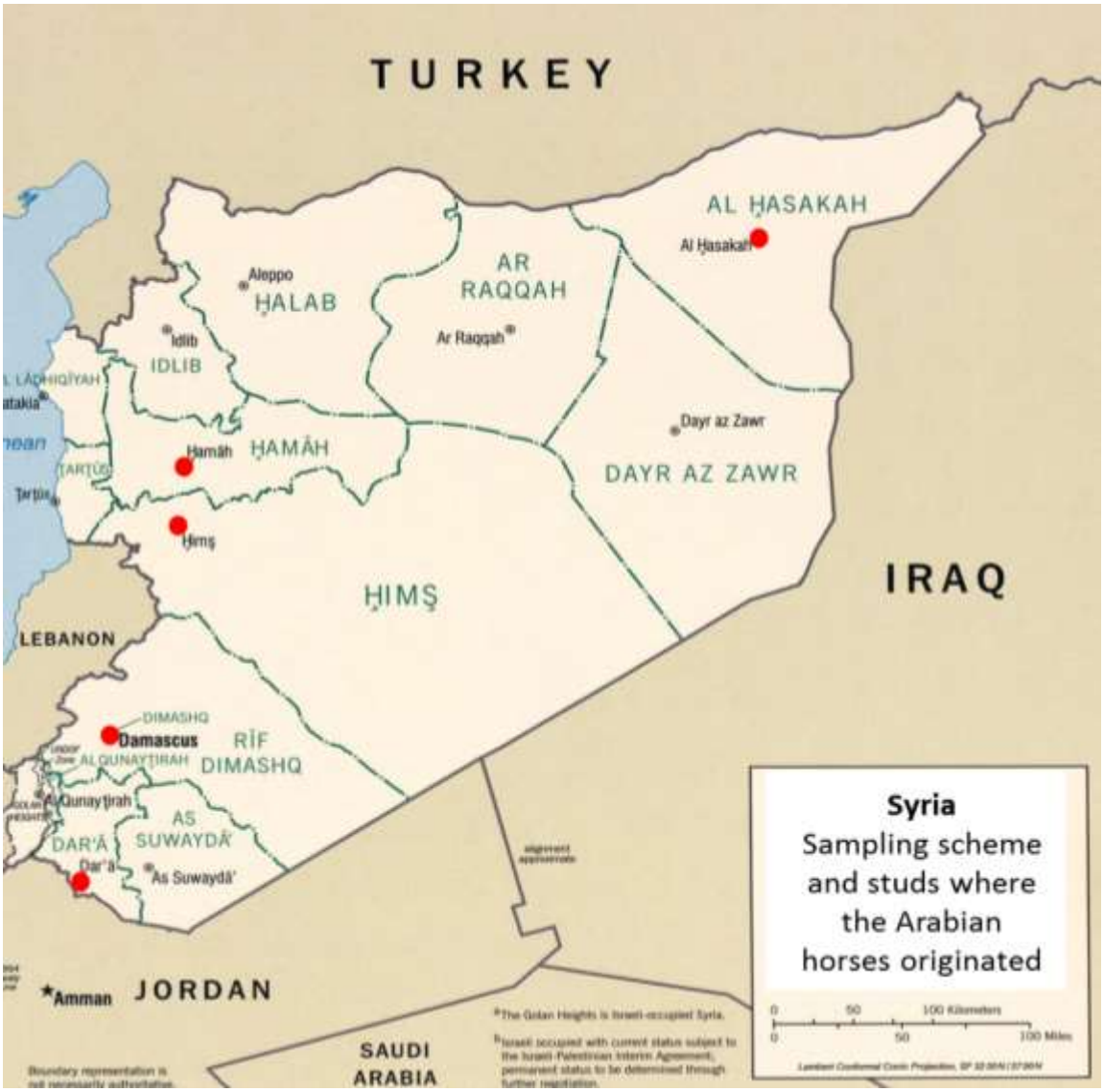


Figure 3.2. Sampling scheme of Syrian Arabian horses
Red dots show sampling regions in Syria. Syria: Administrative Divisions, 2007. The map is used with courtesy of the University of Texas Libraries, The University of Texas.

For the mitochondrial D-loop genotyping study, hair or blood samples were collected from 192 Syrian Arabian horses representing 68 maternal lineages from the three major strains Saglawi (27 maternal lineages with 57 horses), Kahlawi (31 maternal lineages with 75 horses) and Hamdani (10 maternal lineages with 60 horses).

For the microsatellites genotyping study, DNA was collected from 84 Arabian horse samples including hair roots and whole blood South (Daraa, n = 2), Middle-West (Hims, n = 2 and Hama, n = 2) and North-East (Al-Hasakeh, n = 10), as well as from the National Centre

of the Arab Horse in Damascus (n = 68). The sampled Arabian horses are registered in the national stud-books and representing the Syrian strains: Hamdani (n = 26), Kahlawi (n = 30) and Saglawi (n = 28).

For the whole genome Equine SNP array study, DNA was collected from 48 Syrian Arabian horses (36 females and 12 males) representing the strains Saglawi (n=18), Kahlawi (n=16), and Hamdani (n=14) collected from Syria. Genotypes of additional 24 Arabian horses from the USA and three Przewalski`s horses were provided by the Equine Genetic Diversity Consortium (Petersen *et al.* 2013a) and the Molekularbiologisches Zentrum, Humboldt University of Berlin, respectively.

For the candidate genes approach, we sampled (blood and hair) 42 endurance Arabian horses born between 1994 and 2007, performed endurance in official national events carried out in Syria between 2001 and 2010. More details are in the endurance records section.

3.1.3.2. Morphological traits

In this study, we used 12 morphological traits that characterize a wide range of morphological features in horses (Brooks *et al.* 2010a; Sánchez *et al.* 2013). Measurements consisted of back line length (BLL), body length (BL), chest depth (ChD), chest girth (ChG), chest width (ChW), croup height (CH), croup width (CW), fore cannon length (FCL), hind cannon length (HCL), neck girth (NG), throat girth (TG), and withers height (WH).

A detailed description of each trait is provided in Table 3.4 and Figure 3.3. Horses were measured from the left side, starting with the head and continuing with the rest of the body, while standing on a solid ground using measure stake and tape.

Table 3.4. A List of morphological traits, abbreviations and morphometric description

Traits	Abbreviation	Description
Body length	**BL**	Distance from the shoulders till the point of tuber ischium (slopping line)
Back line length	**BLL**	Distance between the highest point of the withers till the pin bone
Croup height	**CH**	Distance from the highest point in the rump (tuber sacral) till the ground
Chest depth	**ChD**	Distance from the highest point of the dorsal in the chest to the sternum (parallel point at the chest floor)
Chest girth	**ChG**	Circumference around the chest behind the front arms
Chest width	**ChW**	The outside distance between the humeral bones (distance from the middle point of the chest to the paralleled point)
Croup width	**CW**	Distance from tuber sacral of the hip from one point to the paralleled point
Fore cannon length	**FCL**	Distance from the lateral tuberculum of the os metacarpale IV to the middle of the fetlock joint
Hind cannon length	**HCL**	Distance from the point of the hock to the ergot (of the hind leg)
Neck girth	**NG**	Circumference of the neck at the attachment area between the breast and the neck
Throat girth	**TG**	Neck circumference at throat latch
Withers height	**WH**	Distance from the highest point of the withers till the ground

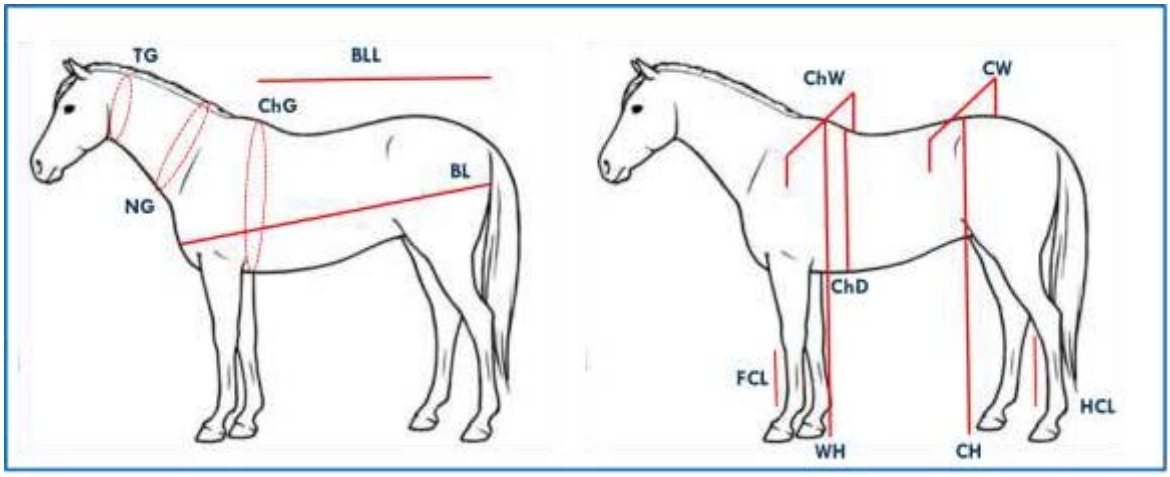


Figure 3.3. Morphological measurements of Syrian Arabian horses
Dashed and solid red lines represent BLL: Back line length, BL: body length, ChD: chest depth, ChG: chest girth, ChW: chest width, CH: croup height, CW: croup width, FCL: fore cannon length, HCL: hind cannon length, NG: neck girth, TG: throat girth and WH: withers height.

Due to difficulties of sampling and measuring traits in Syria, the morphological traits were measured for the GWAS is relatively small sample of 37 mature Syrian Arabian horses, (8 males and 29 females) belonging to three strains: Hamdani (n=9), Kahlawi (n=15) and Saglawi (n=13) born between 1995 and 2010. Horses originated from five governance regions in Syria (Damascus; n=20 and Dara; n=2 in south, Hama; n=4 and Hims; n=1 in the middle-west as well as Al hasakah in the north-east; n=10) (Table 3.5).

Table 3.5. Origin, strains, numbers and sexes of 37 Syrian Arabian horses measured for twelve morphological traits in the current study.

Region	Total				**Strain**			
	number		Saglawi		Kahlawi		Hamdani	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**
Al Hasakah	2	8	1	4	1	3	-	1
Damascus	4	16	1	4	2	6	1	6
Dara	-	2	-	1	-	-	-	1
Hama	1	3	1	1	-	2	-	-
Hims	1	-	-	-	1	-	-	-
Total	8	29	3	10	4	11	1	8

3.1.3.3. Endurance records

For this study, we sampled (blood and hair) 42 endurance Arabian horses born between 1994 and 2007, and have endurance records in official national events carried out in Syria between 2001 and 2010. Average speed records for short and medium endurance distances based on international rules (Ricard & Touvais 2007) have been recorded as following: 40 km (n=12), 80 km (n=18) and 120 km (n=12). Twenty four individuals were rated for high endurance performance (18-25 km/h) and 18 low endurance performance (6-16 km/h), while failed individuals due to lameness were excluded from race but included in our study as low performance horses (Table 3.6). Weights (jockeys and their kits including lead weights) were optimized to 75 kg. Among the horses, 78.5% were born in the South (Damascus and Dara), meaning high homogeneity in geographical affiliation of the studied group (Table 3.7).

Table 3.6. Arabian horses reported for endurance performance (high and low) for three distances in the current study

	40 km		**80 km**		**120 km**	
	High performance	Low performance	High performance	Low performance	High performance	Low performance
	7	5	7	11	10	?
SUM	**12**		**18**		**12**	

Table 3.7. Origin, strains and sexes of 42 Syrian Arabian horses reported for endurance performance in the current study

Region in Syria	**Number of individuals**	
	Male	**Female**
Al Hasakah	1	6
Damascus	8	23
Dara	-	2
Hama	1	-
Hims	1	-
Total	11	31

3.1.3.4. Ethics

Blood and hair samples were collected by state veterinarian according to the animal welfare regulations set by the Syrian Ministry of Agriculture and Agrarian Reform and the Syrian Arabian horse official authorities. All horse owners provided written informed consent after full explanation of the purpose and procedures of the study. All sampling measurements were carried out under supervision of Arab Horses Office, and this research was conducted in full consideration of Syrian Ministry of Agriculture and Agrarian Reform Syria legislations.

3.2. Molecular methods

3.2.1. DNA extraction

Total DNA was extracted from either hair roots (25 pulled mane hairs), or whole blood (5 ml from the jugular vein).

Extraction DNA from hair roots was done by incubating the roots in 180 µl T1 buffer followed by the salting out procedure with some modifications (Miller *et al.* 1988). This method involves salting out of the cellular proteins by dehydration and precipitation with a saturated NaCl solution. While for extraction DNA from whole blood, we used the Puregene Core Kit (Qiagen, Hilden, Germany). We used two tubes for each sample filled with 500 µl fresh EDTA as an anticoagulant. The aim was to have a reserve DNA tube for each horse sampled. Prior DNA extraction, we applied 1000 µl of TE buffer in each tube and centrifuging for 10 minutes at 4.000 rpm. We took the supernatant out and kept the pellet in the tubes. We added 1000 µl TE-buffer again and mixed it before spinning it 10 minutes at 4.000 rpm. We then removed the supernatant and kept washing the pellet until

it became clearly white. Genomic DNA was isolated using Puregene Core kit (QIAGENE) according the manufacturer instructions. DNA was solved in 70 µl of solution mix provided in the kit (Hydration Solution: pure water = 1: 1). The DNA concentration of each sample was adjusted to 50 ng/µl (diluted later to be 10 ng/µl).

3.2.2. DNA fragment amplification

To generate larger quantities of DNA fragments, we used the polymerase chain reaction (PCR) technique. This technique is involving enzymatic amplification of nucleic acid sequences by repeated cycles of denaturation, oligonucleotide annealing, and DNA polymerase extension (Gibbs 1990). The protocol for polymerase chain reaction conditions for the amplification of microsatellite markers are provided by the ISAG-FAO advisory group (FAO 2011). The PCR mixture consisted of the following ingredients (Table 3.8).

Table 3.8. Components, concentrations and quantity per sample, which are required for preparing the PCR mixture of 40 µl.

Chemical components	Concentration (40 µl)	Quantity per sample (µl)
H_2O	-	17.1 µl
GoTaq-buffer, 5X without Mg^{+2}	1 X	8.0 µl
$Mg^{+2,}$ 25 mM	2.5 mM	4.0 µl
dNTP, 5mM	0.2 mM	1.6 µl
Primer upper, 5µM	0.2 mM	1.6 µl
Primer lower, 5µM	0.2 mM	1.6 µl
GoTaq flexi, 5U/µl Promega	Total of 1 U	0.1 µl
DNA (10 ng/µl)	Total of 60 ng	6.0 µl

3.2.3. DNA quantification

Conventional gel electrophoresis of PCR products was used to assess the reaction quality and quantifying the yield for the mitochondrial D-loop, mitochondrial and autosomal genes PCR amplicons. DNA concentration was also determined using the Nanodrop ND-1000. The DNA concentration of the samples was adjusted to 10 ng/µl for mitochondrial and

candidate genes genotyping and to 50 ng/µl for SNP array genotyping. Each PCR amplicons was mixed with 8 µl of Gel-Red (staining pigment) and loaded onto a 2% agarose gel. The components, concentration, and amount of the chemicals, which were used for the preparation of the agarose gel, are presented in Table 3.9.

Table 3.9. Composition of 2% agarose gel

Chemical components	Concentration	Quantity
Agarose	2%	2.3 g
Tris acetic acid EDTA (TAE) Buffer	1%	115 ml

Electrophoresis of PCR-fragments is expected to yield a single strong band of correct size, as determined by comparison with 3 µl size marker (ladder) run on the same gel. After visualizing the bands using the UV- imaging system, we cut the bands and obtained the DNA using the extraction gel kit according to the manufacturer's protocol.

3.2.4. Sequencing of mitochondrial PCR amplicons

For the 192 Syrian Arabian horses, mitochondrial D-loop region was amplified using primers tested by Głażewska et al. (2007). The reference sequence was: NC001640; MT: 15382-15863 bp (Table 3.10).

Table 3.10. The mitochondrial D-loop primers

D-loop primer	Sequence	Length (bp)	Product (bp)	Annealing T˚
mtDNA up	AAC GTT TCC TCC CAA GGA CT	20	486	56
mtDNA low	ATG GCC CTG AAG AAA GAA CC	20		

For amplified PCR products, the framework for sequencing was performed using ABI PRISM 310 sequencer (Applied Biosystem) according to manufacturer protocol. For sequencing in a 10 µl reaction volume, 2 µl of the amplified DNA fragment, 1.0 µl premix, 1.5 µl of Big Dye Sequencing Buffer 5X, 1.0 µl of the upper primer (same of the PCR) with calculated amount of distilled water to fill 10 µl were used.

After pre-sequencing PCR, the reaction mix was cleaned up using the following mixture: 1 µl of 125 mM EDTA, 1 µl 3M Na-acetate, 25 µl of 100% Ethanol, respectively. After mixing gently and incubating for 15 minutes at room temperature, spinning with 6.500 rpm for 30 minutes was performed. Then the supernatant was gently discarded with a 100 µl pipette, followed by adding 35 µl of 70% Ethanol and mixing carefully. Spinning and discarding for the rest of the supernatant were repeated. Finally, the sample was incubated for 10 minutes at room temperature with an open cover, and the pellet was dissolved in 12 µl formamid solution, which was kept in the fridge at 4°C for at least 30 minutes.

3.2.5. Microsatellites genotyping

Eighty four Syrian Arabian horses were genotyped for 12 equine microsatellite markers: AHT4, ASB17, ASB23, HMS1, HMS2, HMS3, HMS6, HMS7, HTG4, HTG7, LEX3 and VHL20 chosen from the recommended standard panel of the International Society of Animal Genetics (ISAG 2014). Polymerase chain reaction (PCR) conditions were provided by the ISAG-FAO advisory group protocol (FAO 2011). The detection of the microsatellite allele length was performed with LI-COR gel electrophoresis standard system based upon thermo-cycling process was used to visualize the PCR products labeled with a fluorescence labelled M13-tailed primer. Reference samples were used for correct allele length

detection. The 12 equine microsatellites were chosen for this study, primers, chromosomal affiliation, sequences, lengths (bp), and repeat motif are in the Table 3.11.

Table 3.11. List of 12 microsatellites with their primers, affiliation to chromosomes, sequences, lengths (bp), repeat motifs and references

Loci	ECA	Primer Sequence (5'-3')	Expected allelic range (bp)	Repeat motif	Reference
AHT4	24	AAC CGC CTG AGC AAG GAA GT GCT CCC AGA GAG TTT ACC CT	142-164	(AC)nAT(AC)n	(Binns et al. 1995)
ASB17	2	ACC AGT CAG GAT CTC CAC CG GAG GGC GGT ACC TTT GTA CC	81-125	(AC)n	(Breen et al. 1997)
ASB23	3	ACA TCC TGG TCA AAT CAC AGT CC GAG GGC AGC AGG TTG GGA AGG	181-209	(TG)n and (TG)nTT(TG)4	(Breen et al. 1997)
HMS1	15	CAT CAC TCT TCA TGT CTG CTT GG TTG ACA TAA ATG CTT ATC CTA TGG C	170-186	(TG)n	(Guérin et al. 1994)
HMS2	10	CTT GCA GTC GAA TGT GTA TTA AAT G ACG GTG GCA ACT GCC AAG GAA G	216-244	(CA)n(TC)2	(Guérin et al. 1994)
HMS3	9	CCA ACT CTT TGT CAC ATA ACA AGA GCC ATC CTC ACT TTT TCA CTT TGT T	152-180	(TG)2(CA)2TC(CA)n and (TG)2(CA)2TC (CA)n GA(CA)5	(Guérin et al. 1994)
HMS6	4	CTC CAT CTT GTG AAG TGT AAC TCA GAA GCT GCC AGT ATT CAA CCA TTG	155-169	(GT)n	(Guérin et al. 1994)
HMS7	1	CAG GAA ACT CTC ATG TTG ATA CCA TC GTG TTG TTG AAA CAT ACC TTG ACT GT	171-189	(AC)2(CA)n	(Guérin et al. 1994)
HTG4	9	CTA TCT CAG TCT TGA TTG CAG GAC GCT CCC TCC CTC CCT CTG TTC TC	126-142	(TG)nAT(AG)5AAG (GA)5, ACAG(AGGG)3	(Ellegren et al. 1992)
HTG7	4	CCT GAA GCA GAA CAT CCC TCC TTG ATA AAG TGT CTG GGC AGA GCT GCT	118–128	(GT)n	(Marklund et al. 1994)
LEX3	X	ACA CTC TAA CCA GTG CTG AGA CT GAA GGA AAA AAA GGA GGA AGA C	142-164	(TG)n	(Coogle et al. 1996)
VHL20	30	CAA GTC CTC TTA CTT GAA GAC TAG AAC TCA GGG AGA ATC TTC CTC A	85-109	(TG)n	(Haeringen et al. 1994)

3.2.6. Genotyping with the equine SNP array

Blood samples were collected from 48 Syrian Arabian horses and prepared using the Purgene Core Kit (QIAGENE). The 48 high quality DNA samples were genotyped using Equine SNP70 BeadChip using the standard procedures of the manufacturer (Illumina Inc., San Diego, USA). The Equine SNP70 BeadChip contains 74,500 SNPs and has an average density of 1 SNP per 33 Kb. A total number of 65.156 SNP calls were generated after hybridization. SNP calling was done from the raw intensity data with the minimum score cutoff of 0.05 by Delta Genomics, CA (Figure 3.4).

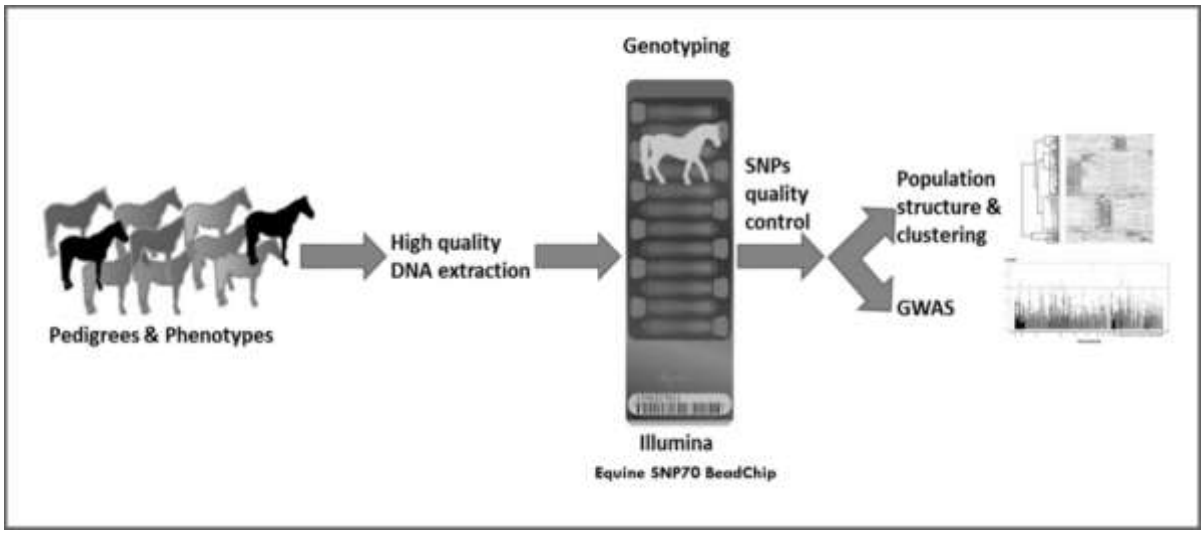


Figure 3.4. Schematic representation of the genotyping using the Equine SNP70 BeadChip

3.2.7. Sequencing of candidate genes

A total of 2.578 bp of three mitochondrial genes (*MT-COX3*: 8651-9403 bp, *MT-CYB*: 14307-15150 bp, and *MT-ND4*: 10345-11325 bp), as well as, promoters, 5' and 3' UTRs and exons of the autosomal genes (*ACTN3* and *MSTN*) were amplified and sequenced using self-designed primers done by means of Primer3 online tool (Untergasser *et al.* 2012). If the intron was smaller than 350 bp, two exons were amplified using one primer based on the

reference equine genome assembly: EquCab2 GCA_000002305.1. Primers information with PCR product sizes are listed in the Table 3.12.

Table 3.12. Information of gene specific sequences, PCR product sizes and annealing temperature for the analyzed fragments

ıers ID	Sequence	Product (bp)	Annealing T˚
Mitochondrial genes			
MT-COX3 up	CAC CAA ACC CAC GCT TAC	752	57
MT-COX3 low	TCC TCA TCA ATA AAT AGA GACG		
MT-ND4 up	CAA TAG CCT AAA CTT CTC AC	981	57
MT-ND4 low	GAA TAG CTC TCC AAT TAG G		
MT-CYB up	CCT AAT CCT CCA AAT CTT AAC	845	50
MT-CYB low	CTA AGA GTC AGA ATA CGC ATT G		
Autosomal genes			
ACTN3 prom up	AGG TTG AGC AGC TGG AAG G	633	59
ACTN3 prom low	CTG TTC CAT ATA CTC GCC GC		
ACTN3 Exon1 up	CTT TCC CAA GGT CAC ACA GC	633	59
ACTN3 Exon1 low	TCC CCT TGT CAC CCT AAA CC		
ACTN3 Exon2-3 up	ACT AGA GCT CAG GGA GGG AA	635	59
ACTN3 Exon2-3 low	TGT GAG GCA TGG GTG GTT AT		
ACTN3 Exon4-5 up	GAT CTG AAC CCG TGA AGC TG	802	59
ACTN3 Exon4-5 low	CAT TAC CAG ACT TGC GCC AT		
ACTN3 Exon6-7 up	TGG TAA TGA AGG GCC TCA CA	642	59
ACTN3 Exon6-7 low	GGG ACC AAT ATG CTC CCA GA		
ACTN3 Exon8 up	CAG GGA AGA AGA CAC TGG GT	489	59
ACTN3 Exon8 low	CTC CCT GTG TGA TGC CCT TA		
ACTN3 Exon9 up	CTT TGC ATG GGT CCA GGT TT	363	59
ACTN3 Exon9 low	GAG CTT GGA TGG GCA GAA AG		
ACTN3 Exon10 up	GAG ATG GGT GGA TGA GGT GA	400	59
ACTN3 Exon10 low	CCA TCA CGG TTC ACC CAT TG		
ACTN3 Exon11 up	ATC AAC TTC AAC ACG CTG CA	645	59
ACTN3 Exon11 low	CCT TTG GAC ACC TGC TAT GC		
ACTN3 Exon12 up	TAT CAC ACT AGC GCC TCA GG	482	59
ACTN3 Exon12 low	GGG ACA AGT GAT GAT GGG GA		
ACTN3 Exon13-14 up	GCA GGC AAG GAG GAA ATC TG	462	59
ACTN3 Exon13-14low	AGC TTC CCT GTC ATC CCA TC		
ACTN3 Exon15 up	AAA GCG CCA GTT CTT GAG TG	417	59
ACTN3 Exon15 low	TGA GGT TTC AGG GTG GCT AG		
ACTN3 Exon16-17 up	GTA AAT GGT GCA CTG ACC CC	774	59
ACTN3 Exon16-17low	TTA GAC TGC TCT GTG ACC GG		
ACTN3 Exon18-19 up	AAC CTC CAG ATG CGG ACA G		

Primers ID	Sequence	Product (bp)	Annealing T˚
ACTN3 Exon18-19low	GCG TGA TGA GGA GGA AGT GA	547	59
ACTN3 Exon20-21 up	TCT GTG TGA CTC CAA AGC CT		
ACTN3 Exon20-21low	TGT TCC CTT CCA CGG TGT AA	1052	59
MSTN prom up	TGC CCT GGT AAT AAC AAT GAA GA	1200	58
MSTN prom low	TGC CTG TAC AGT CTG AGA GA		
MSTN Exon1 up	CTG GTG TGG CAA GTT GTC TC	682	58
MSTN Exon1 low	TGC AGC AGA TTT CAG TCT CA		
MSTN Exon2 up	GTT CCT CCA CGG TGT CTC TT	878	59
MSTN Exon2 low	TTA TTG GGT ACA GGG CTG CC		
MSTN Exon3 up	AAC AAG CGT GAA GAG AGG GA	801	59
MSTN Exon3 low	AAT TGT GAG GGG AAG GCC TT		

Genes were initially sequenced in 10 endurance Arabian horses; including sub cohort of high performance horses (n=5) and low performance horses (n=5). The identified variants were verified in a total of 42 horses sequences; 24 with high performance, and 18 horses with low performance over three distances 40 km (n=12), 80 km (n=18), 120 km (n=12).

For both, mitochondrial and autosomal genes amplified PCR products, the sequencing was performed according to information that are provided above in the chapter 3.2.4. The genomic positions of the identified sequence variants were determined according to the *Equus caballus* genome assembly EquCab2 GCA_000002305.1, and the protein sequence that are available in Ensembl, Release 90, 2017. *In silico* analysis were performed using the Variant Effect Predictor toolset (Ensembl) to determine functional consequences and novelty of the identified variants. Furthermore, we checked both of the transcription factor binding sites (TFBSs) within promoter regions, which are essential for enhancing the routine cellular functions, using ConSite online toolset (http://consite.genereg.net/), as well as miRNA target sites in the 5' and 3' UTRs using miRBase Database, Release 21, with filtering for *Equus caballus* (Griffiths-Jones *et al.* 2006; Griffiths-Jones *et al.* 2007).

For determining alleles and genotypes frequencies after finding sequence variants, we additionally genotyped 32 Arabian horses for the important identified *MT-COX3, MT-CYB*, and *MT-ND4* variants by sequencing and restriction fragment length polymorphism (RFLP) technique detection using specific restriction enzymes (Table 3.13).

For genotyping the 32 Arabian horses for the autosomal genes variants, we used customized allele-specific PCR assays. Reagents were obtained from KBioscience (UK), PCR was performed on a StepOnePlus set, (Applied Biosystems, USA) based on a protocol from Kreuzer *et al.* (2013) (Table 3.14).

Table 3.13. The sequencing and RFLP primers for the important identified *MT-COX3*, *MT-CYB*, and *MT-ND4* in 32 Arabian horses

Primer ID	Gene	Upper primer	Lower primer	Length (bp)	Temp
H-mtCO3	*MT-COX3*	CAC CAA ACC CAC GCT TAC	TCC TCA TCA ATA AAT AGA GAC G	773	57
H-mtCyB	*MT-CYB*	CCT AAT CCT CCA AAT CTT AAC	CTA AGA GTC AGA ATA CGC ATT G	866	50
RFLP P-6065	*MT-ND4*	CAA TAG CCT AAA CTT CTC AC	GAA TAG CTC TCC AAT TAG G	1000	57

Table 3.14. The customized allele-specific PCR assays and primers for the important identified *ACTN3* and *MSTN* in 32 Arabian horses

Assay ID	Gene	Primer A1	Primer A2	Primer C	Temp
KP-049	*MSTN*	TAT TAA GTA ATC AGG TTA TAA TGC ACC AAA	ATT AAG TAA TCA GGT TAT AAT GCA CCA AG	CCA GGA CTA TTT GAT AGC AGA GTC ATA AA	57
KP-051	*MSTN*	ATT CTT TCT ATT TCA AAT GTT TGC CTA AAT AAT	CTT TCT ATT TCA AAT GTT TGC CTA AAT AAC	GAA ATG TTA CTT CCT CAG AAA TTA AGA TTT	57
KP-053	*ACTN3*	GGG GCC TCG TTA AGT AGC GT	GGG GCC TCG TTA AGT AGC GC	CCC CAT ATT TAG CGC GAA TCC GAT	57
KP-054	*ACTN3*	GAC CCC TTG ACC TCT CCT CTT A	GAC CCC TTG ACC TCT CCT CTT T	GAT TTT GTG GAA GCG CAT CTT GCC TT	57
KP-055	*ACTN3*	GTT CTC CAC GCA AGT AGG AGC	GGT TCT CCA CGC AAG TAG GAG T	TGG GAT CAG CCA GAG GGA GCA A	57

3.3. Analyses of diversity and Statistics

3.3.1. Genetic diversity indices and phylogeny

For the mitochondrial D-loop region sequences (15495-15847 bp), Sequence editing and multiple alignments were done by using Clustal Omega package (Sievers & Higgins 2014), sequence scanner software v2.0 (Applied Biosystems, USA) and the BioEdit (Hall 1999). For estimation of the mitochondrial haplotype diversity (HTD) of each horse strain, we measured the degree of haplotype uniqueness (Nei & Tajima 1981).

The pair-wise population differentiation index (F_{ST}) was used for quantification of the genetic differentiation between the three strains. For the evaluation of the results, we followed the quantitative scale of Wright (Wright 1978). To assess the genetic variance among horses within and between strains, we performed the analysis of molecular variance (AMOVA) using Arlequin v5.3 (Excoffier & Lischer 2010).

For the neighbor-joining tree reconstruction and the median-joining network, we used MEGA 5.2 (Tamura *et al.* 2011) and NETWORK 4.6.1 software (Bandelt *et al.* 1995), respectively. The GenBank horse sequence NC001640 was used as a reference and Przewalski's horse GenBank sequence JN398402 as an out-group.

For the Arabian horses microsatellites, standard genetic diversity analyses were carried out using GenAlEx 6.5 (Peakall & Smouse 2012) including number of alleles for individual markers (N_a), number of private alleles (N_{Pa}), observed (H_o) and expected (H_e) heterozygosity, and allele frequencies.

The genetic variation existing within and between the three strains was analyzed by Analysis of Molecular Variance (AMOVA) and Wright`s F statistics (F_{IS}, F_{IT} and F_{ST}) were used to quantify the genetic variances within and between the three strains. The polymorphism

information content (PIC) was calculated to investigate the polymorphic status of each locus per each strain. Furthermore, we produced the first three principal coordinates (PCoA) for 12 microsatellites of the three Arabian horse strains based on genetic distance matrix.

We used POPTREE2 (Takezaki *et al.* 2010) to construct phylogenetic tree from genetic distance matrix of allele frequencies data by using the neighbor-joining (NJ) method (Saitou & Nei 1987), performing 5000 bootstrap tests.

3.3.2. SNP array data analyses

For genotyping 48 Syrian Arabian horses we used the equine SNP array (SNP70 BeadChip). Additionally, the study is provided with genotypes of 24 Arabian horses from the USA and three Przewalski`s horses.

We used R for statistical computing to perform data quality control (QC). The following steps were taken to reduce the input set of 65.156 SNPs to a set of high confidence SNPs: (1) SNP markers with a GenTrain Score $\leq$ 0.6 were removed, (2) markers were harmonized between our data and genotyped horses from the Equine Genetic Diversity Consortium, (3) non-informative and low informative markers were removed, (4) duplicated markers were removed, and (5) allele frequencies of markers were checked to be ≥ 5%. These procedures ended with 38.671 informative SNPs. The number of markers per chromosome before and after QC are found in supplementary Table S3.1.

The diveRsity package (R 3.3.1) was used to calculate heterozygosity values (H_e, H_o), inbreeding coefficients (F_{IS}) and pairwise population genetic differentiation estimates (F_{ST}). Hierarchical clustering hclust package, which is available in R 3.3.1, was used to generate a dendrogram showing the genetic distance (using Manhattan distance) among all samples

(Arabian horses from Syria and the USA as well as Przewalski's horses). STRUCTURE was used for a more fine grained analysis of population structure among Syrian Arabian horses (n=48) together with the Przewalski's horses (n=3). We tested K=2 to K=6 clusters using the following settings for 51 individuals: N=2 for diploidy, N=38.671 for SNP loci. Parameter settings for the analysis were burnin=5000 and MCMC reps=1000. Each analysis was repeated 10 times, to make sure convergence on similar F_{ST} values was observed between repeated runs (for the same K clusters). Results from STRUCTURE were analyzed and visualized using R.

3.3.3. Genome-wide association study

The GWAS was performed with 37 mature Syrian Arabian horses which are genotyped using the equine SNP70 BeadChip and provided with data of 12 morphological traits.

When performing GWAS on low sample sizes, quality control of genotype and phenotype data is particularly important, since false associations are very likely to occur. That is why we performed extensive quality control, using seven different metrics to remove SNPs which could lead to false associations.

We employed a rigorous quality control (QC) of our genotype data using the following steps: We removed (1) SNPs with GenTrain scores < 0.6, (2) non-segregating S, (3) SNPs with 5% or more missing data, (4) markers with a minor allele frequency (MAF) < 5%, (5) markers with less than 4 individuals in a genotype class, (6) Markers not in Hardy Weinberg equilibrium ($P_{(Bon)}$ < 0.05) and (7) we pruned remaining SNPs for LD in a 2 Mb window surrounding the SNP, grouping markers which show an $r^2 > 0.5$, and then keep the marker with the highest MAF. The remaining 14.920 SNPs that passed the quality control were used for the GWAS.

The twelve morphological traits (back line length, body length, chest depth, chest girth, chest width, croup height, croup width, fore cannon length, hind cannon length, neck girth, throat girth, and withers height) were corrected for sex differences. Afterwards, outliers (mean +/- 2 SD) were removed from the phenotype data for individuals that show extreme phenotypes to prevent false positive associations. In addition, morphological traits were corrected for strain effects, if a suggestive strain effect was observed ($P < 0.1$). The traits that were corrected for strain are body length, chest depth, chest width, and throat girth. Trait correction was performed using the mean of the original data with addition of the residuals after fitting a linear model using the lm function in R 3.3.1. For the association analysis, we applied a standard linear model with Y as corrected morphological trait:

$$\boldsymbol{Y}_{ij} = \mu + \boldsymbol{SNP}_{i} + \boldsymbol{e}_{j}$$

P values were corrected for multiple testing using Bonferroni correction to avoid false positive associations. Results were visualized using Manhattan plots and investigated using quantile-quantile (QQ) plots and genomic inflation factor (ƛ). Associations were visualized using boxplots.

4. Results

4.1. Diversity of mitochondrial DNA in three Arabian horse strains

4.1.1. Genetic variation and haplotypes of the mitochondrial D-loop

In comparison with the reference sequence (NC001640), the alignment of 192 sequences revealed 38 DNA variations, among them 37 transitions and one deletion (at the position 15532). Using all variations, 28 haplotypes were derived representing 68 maternal lineages (Supplementary Table S4.1). Haplotypes differed from each other by one to 11 variations.

Five haplotypes belong to the strain Hamdani (H001-H005), 12 to Kahlawi (K002-K013) and nine to Saglawi (S002-S010). Only two haplotypes were shared among different maternal lineages: KS01 (represents 3 maternal lineages from Saglawi and Kahlawi) and KSH1 (represents 6 maternal lineages from all three strains). Compared to other studies, we identified three novel haplotypes that received following GenBank accession numbers: KT724968 (K012), KT724969 (K006) and KT724970 (S003) (Supplementary Table S4.2). The highest haplotype diversity value (HTD) was found in Kahlawi (0.95) followed by nearly equal values in Hamdani (0.91) and Saglawi (0.90) (Table 4.1).

Table 4.1. Mitochondrial diversity in three Arabian horse strains

Strain	N_T	N_U	N_P	N_{HT}	HTD
Hamdani (H)	60	10	15	6	0.91
Kahlawi (K)	75	31	33	14	0.95
Saglawi (S)	57	27	26	11	0.90

Total number of animals (N_T), number of unrelated individuals (N_U), number of DNA sequence polymorphisms (N_P), number of haplotypes (N_{HT}), and haplotype diversity (HTD)

Table 4.2. Population differentiation (F_{ST}) estimates of three Arabian horse strains

Strain	Hamdani	Kahlawi	Saglawi
Hamdani	0.000		
Kahlawi	0.138	0.000	
Saglawi	0.205	0.098	0.000

The pair-wise population differentiation index (F_{ST}) between the three strains ranged between 0.098 and 0.205. The highest value was found between Hamdani and Saglawi (0.205). It is lower between Hamdani and Kahlawi (0.138) and lowest between Saglawi and

Kahlawi (0.098) (Table 4.2). The analysis of molecular variance (AMOVA) revealed genetic variation of 14.18 % and 85.82 % between and within the three strains, respectively.

4.1.2. Maternal phylogenetic relationship

The neighbor-joining tree displayed a mixture of the mitochondrial haplotypes from the three strains. All haplotypes from all maternal lineages were scattered across the neighbor-joining tree. The sub-division between the three strains that was expected according to the pedigree could not be confirmed (Figure 4.1).

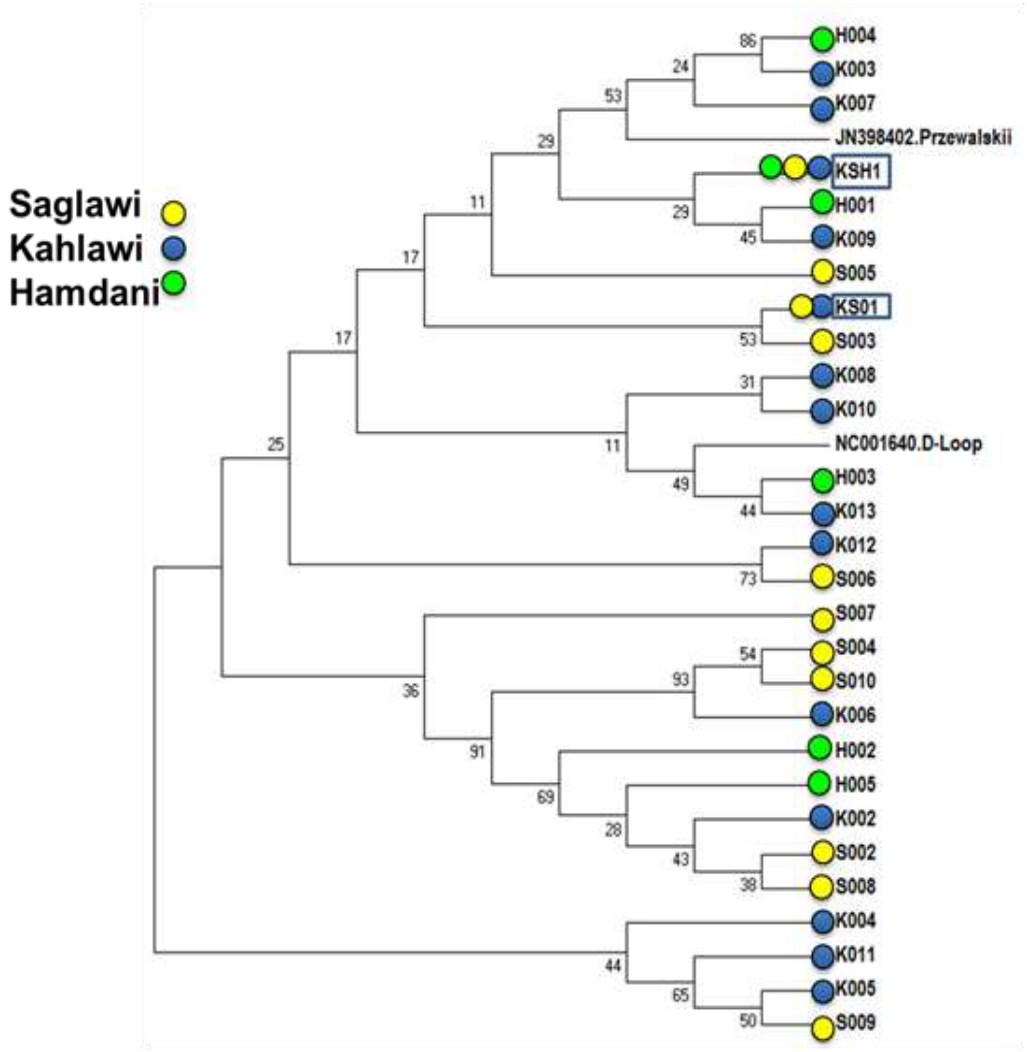


Figure 4.1. Neighbor-joining tree of mitochondrial haplotypes found in three Arabian horse strains.
Mitochondrial variation motifs of each haplotype are provided in the supplementary Table S4.1.

The median-joining network clustered the haplotypes into seven haplogroups (A, B, C, G, M, P and Q) which were named according to Achilli *et al*. (Achilli *et al.* 2012) (Figure 4.2).

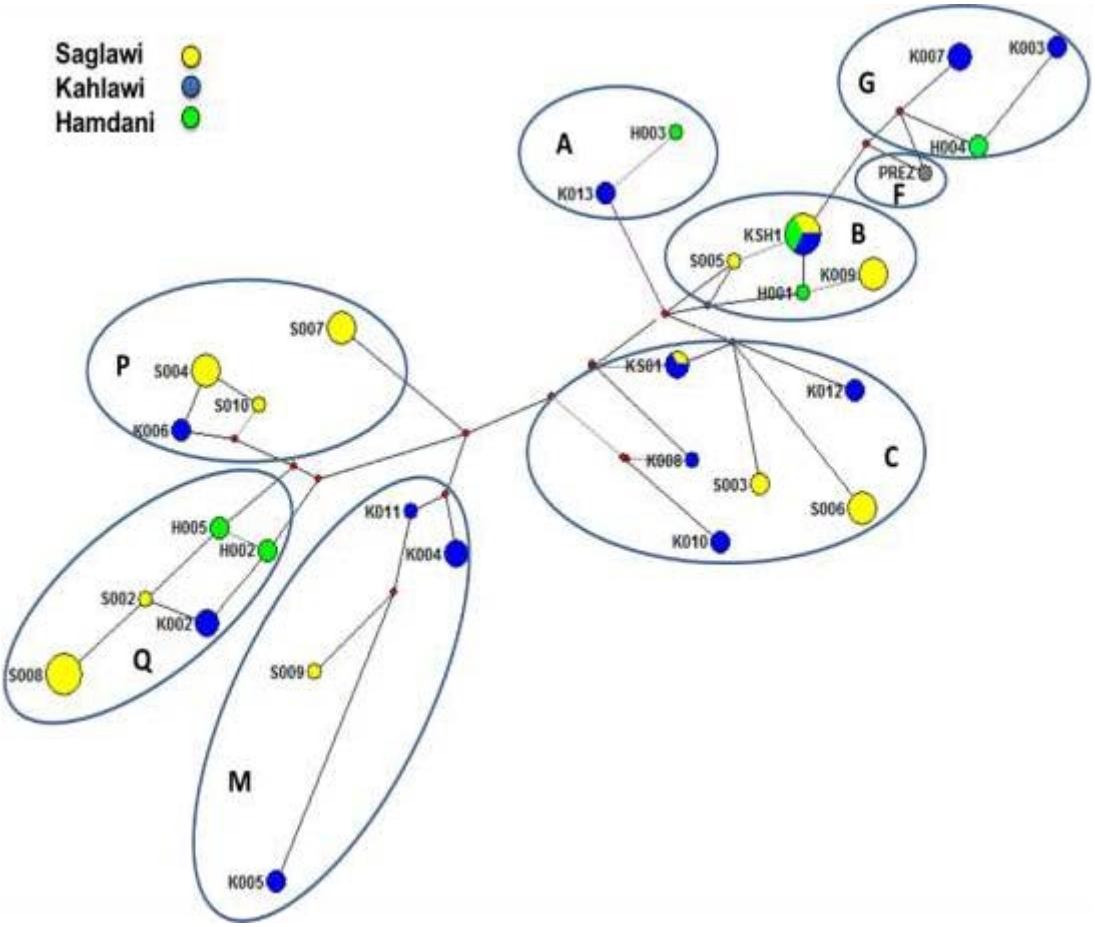


Figure 4.2. Median-joining network based on 353 bp of mitochondrial D-loop sequences of 192 Arabian horses represented by 28 haplotypes
The individuals of each haplotype are represented by colored circles depending on strains (Saglawi (S) - yellow, Kahlawi (K) - blue, Hamdani (H) - green). Sizes of the nodes are proportional to haplotype frequencies of maternal lineages represented in each strain. Przewalski sequence (GenBank JN398402) is labeled with a grey circle in haplogroup F.

The highest number of maternal lineages was found in haplogroups C and Q (14 in each with over 40 percent of animals). Haplogroup A contained only three maternal lineages, haplogroup B had 12 lineages presenting only 14 horses, and the haplogroups G and M contained seven lineages in each group presenting 15 and 26 horses, respectively. Every haplogroup comprised a mixture of haplotypes and thus a mixture of maternal lineages

from more than one strain. For example, Kahlawi haplotypes were represented in all haplogroups, while Saglawi haplotypes were present in five (B, C, M, P, Q) and Hamdani haplotypes only in four haplogroups (A, B, G and Q).

With respect to the pedigree information, we found minor errors. For example, we expected with view on their mother that two Kahlawi individuals would represent the haplotypes K003 and K004, but they were genetically profiled as the haplotypes K004 and K005, respectively.

4.2. Genetic diversity based on 12 microsatellite loci

Using 84 Syrian Arabian horses and 12 equine microsatellites, we identified genetic diversity within and between the three strains Kahlawi, Hamdani and Saglawi.

4.2.1. Within the three strains

A total of 251 alleles were detected, number of alleles per strain was 84, 76 and 91 in Kahlawi, Hamdani and Saglawi, respectively (Table 4.3 and supplementary Table S4.3). All loci were polymorphic and number of alleles (N_a) per locus ranged from 3 (HTG4 and HTG7) to 11 (ASB17 and LEX3) (Table S4.2). The average number of alleles varied only slightly between the strains with an average of 6.972 ±0.364 overall strains. The effective number of alleles (N_e) was the lowest in Hamdani (3.797 ±0.357), and the highest in Saglawi (7.583 ±0.621).

Microsatellites overall the three Syrian strains showed high heterozygosity values. The mean observed heterozygosity (H_o) and the mean expected heterozygosity (H_e) were similar in the three strains but slightly higher in the Saglawi strain (Table 4.3). Detailed heterozygosity values per each locus per strain are shown in the supplementary Table S4.4.

Inbreeding coefficients (F_{IS}) per strain for the 12 loci are also shown in Table 4.3. Low F_{IS} values, on average, were present in Hamdani and Kahlawi (0.010 and 0.027, respectively), while Saglawi showed a higher value (0.123). Polymorphic information content (PIC) within the strains fluctuated from 0.363 (HTG7 in Kahlawi) to 0.890 (ASB17 in Saglawi). All PIC values and their means (±SE) per each strain/each locus are shown in the supplementary Table S4.4. For the three strains, no significant variations were detected.

Table 4.3. Microsatellites diversity indices in Syrian Arabian horses
The indices include total number of alleles (TN_a), number of private alleles or rare alleles with frequency less than 0.1 (N_{Pa}), average number of alleles per locus per strain (N_a), average number of effective alleles (N_e), observed heterozygosity (H_o), expected heterozygosity (H_e) and the inbreeding coefficients (F_{IS}) per strain for the 12 microsatellite markers

Strain	TN_a	N_{pa}	N_a (±SE*)	N_e (±SE)	H_o (±SE)	H_e (±SE)	F_{IS} (±SE)
Kahlawi	84	9	7.000 ±0.663	4.044 ±0.414	0.697 ±0.050	0.717 ±0.037	0.027 ±0.047
Hamdani	76	5	6.333 ±0.607	3.797 ±0.357	0.696 ±0.031	0.707 ±0.031	0.010 ±0.032
Saglawi	91	12	7.583 ±0.621	4.582 ±0.537	0.646 ±0.046	0.743 ±0.037	0.123 ±0.054
Overall mean	251	26	6.972 ±0.364	4.141 ±0.254	0.680 ±0.025	0.722 ±0.020	0.053 ±0.027

*All values are provided with their standard error (SE)

Wright`s F-statistics indices F_{IS}, F_{ST}, F_{IT} estimated overall individuals per loci are listed in Table 4.4. Average F_{IS} overall horses was 0.054, and average F_{ST} of all strains per each locus was 0.019 ranging from 0.007 (HTG07) to 0.037 (LEX3). Additionally, the mean of global deficit of heterozygotes across strains (F_{IT}) amounted to the value of 0.072.

Table 4.4. F-statistics of 12 microsatellite loci analyzed in 84 Syrian Arabian horses

Locus	F_{IS}	F_{IT}	F_{ST}
AHT04	-0.195	-0.180	0.013
ASB17	0.025	0.046	0.021
ASB23	0.221	0.242	0.027
HMS01	-0.083	-0.053	0.028
HMS02	0.064	0.073	0.010
HMS03	0.107	0.125	0.020
HMS06	0.008	0.028	0.020
HMS07	0.115	0.126	0.012
HTG04	0.031	0.043	0.013
HTG07	-0.037	-0.030	0.007
LEX3	0.248	0.275	0.037
VHL20	0.150	0.167	0.021
Overall mean	0.053	0.072	0.019
±SE	±0.036	±0.037	±0.003

F_{IS}: inbreeding coefficients of individuals relative to each strain. F_{IT}: inbreeding coefficients of individuals relative to the total population. F_{ST}: fixation index (measure of population differentiation due to genetic structure).

4.2.2. Between the three strains

The analysis of molecular variance (AMOVA) of the twelve microsatellites revealed estimated variation among the strains of 0.045 (1%) and among individuals within the three strains of 0.343 (8%) (Table 4.5).

Table 4.5. Analysis of molecular variance (AMOVA) of 12 microsatellite loci in the three strains of Syrian Arabian horses

Source of variation	df	Sum of squares	Mean squares	Estimated variation	Percentage variation (%)
Among strains	2	14.512	7.256	0.045	1%
Among individuals within strains	81	385.833	4.763	0.343	8%

Genetic differentiation estimates (F_{ST}) between pairs of strains are shown in Table 4.6. The F_{ST} estimations ranged between 0.013 (Saglawi vs. Kahlawi) and 0.015 (Hamdani vs. Saglawi and Kahlawi). In general, values showed low magnitude of differentiation among the three tested strains.

Table 4.6. Pairwise population differentiations (F_{ST}) among the three Arabian horse strains

	Kahlawi	**Hamdani**	**Saglawi**
Kahlawi	0.000		
Hamdani	**0.015**	0.000	
Saglawi	**0.013**	**0.015**	0.000

In the three based on genetic distance principal coordination figures (Figure 4.3; A, B and C), no partition among Arabian horses was detected.

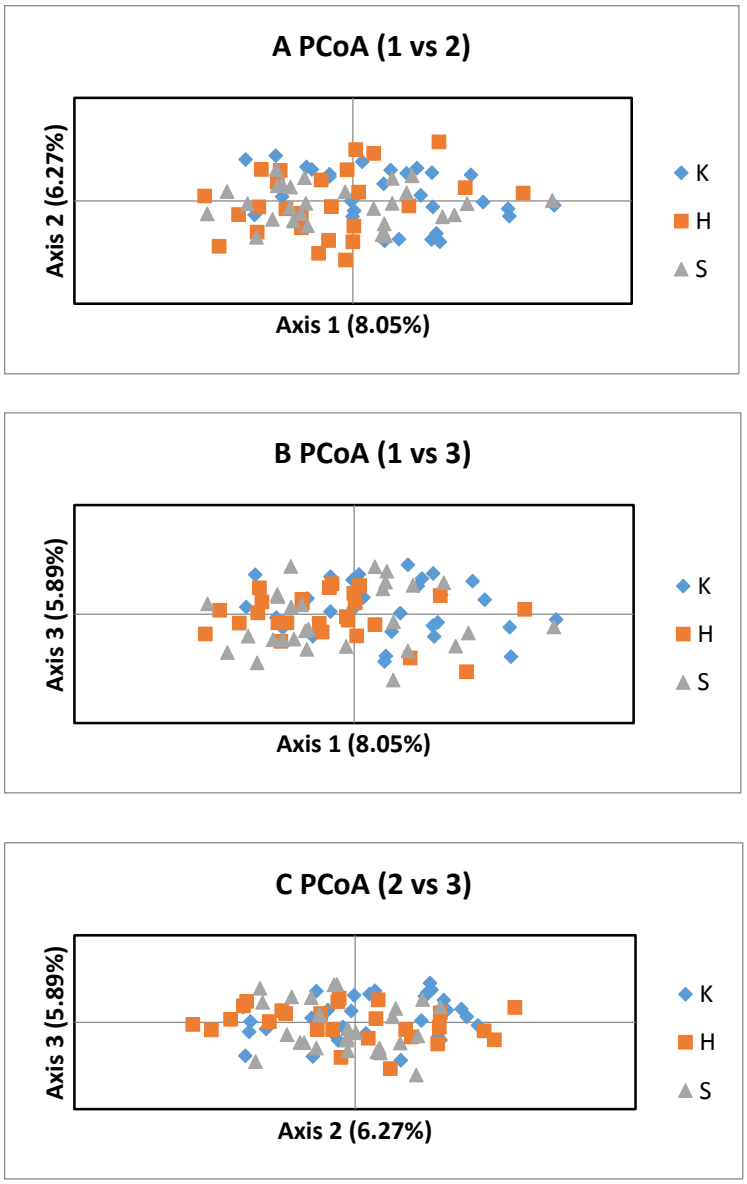


Figure 4.3. First three PCoAs of the Syrian Arabian horses descended from three strains; K: Kahlawi, H: Hamdani and S: Saglawi.

Moreover, in Nei's phylogenetic tree Hamdani (POP2) showed a separation from Kahlawi and Saglawi (POP1 and POP 3) (Figure 4.4).

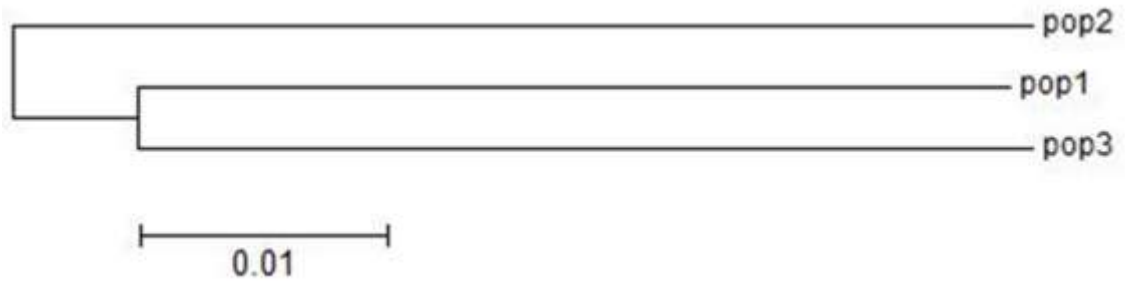


Figure 4.4. Neighbor-joining dendrogram based on Nei's genetic distance of the three Syrian strains (POP1: Kahlawi, POP2: Hamdani, POP3: Saglawi) using POPTREE2

4.3. Genetic diversity using the equine SNP70K BeadChip

For this study, 48 Syrian Arabian horses were genotyped using SNP70 BeadChip, genotypes of 24 Arabian horses from the USA, and three Przewalski`s horses. Out of the total 65.156 SNPs, 38.671 informative SNPs passed the quality check as it has been described under Material and methods (3.3.2).

4.3.1. Diversity indices

The observed heterozygosity (H_o) of the three strains ranged between 0.30 (Hamdani) and 0.32 (Saglawi, Kahlawi) and was similar to the expected heterozygosity (H_e) (Hamdani 0.30, Saglawi 0.31, Kahlawi 0.31).

The inbreeding coefficients (F_{IS}) within the three strains had a tendency to negative values (Saglawi -0.05, Kahlawi -0.02, Hamdani -0.04). According to the quantitative scale for F_{ST} values, as proposed by Wright (1978), we observed very low population differentiation between the three Syrian strains with F_{ST} values ranging between 0.005 and 0.009 (Table 4.7).

Table 4.7. Pairwise F_{ST} values between the Syrian Arabian horse strains, Arabian horses from the USA and Przewalski's horses (calculated by diveRsity R package, F_{ST} estimates after 1,000 bootstraps)

	Saglawi	**Kahlawi**	**Hamdani**	**Arabian (USA)**
Kahlawi	**0.005**			
Hamdani	**0.009**	**0.009**		
Arabian (USA)	0.032	0.025	0.040	
Przewalski's horse	0.196	0.194	0.206	0.200

4.3.2. Population structure

Hierarchical cluster analysis showed a clear distance between Arabian and Przewalski's horses and two clusters among Arabian horses (Figure 4.5.A). While one cluster of Arabian horses contained 12 Arabian horses exclusively from the USA (group I), the other cluster contained horses from Syria and the USA. The mixed cluster was further split into two sub-clusters which comprise either only Syrian strains (37 horses; group II) or Syrian and the remaining horses from the USA (11 horses of each; group III).

The population structure model K=2 (Figure 4.5.B) made a clear separation between Przewalski's and Arabian horses. When analyzing Syrian Arabian horses and the Przewalski's horses, STRUCTURE provided the best K value when K=4 (Figure 4.5.B) where the variance of Ln likelihood between runs increases continuously and the plateau of the mean value of Ln likelihood is stable (Evanno *et al.* 2005). Analysis when K=4 showed three clusters among the examined Syrian Arabian horses.

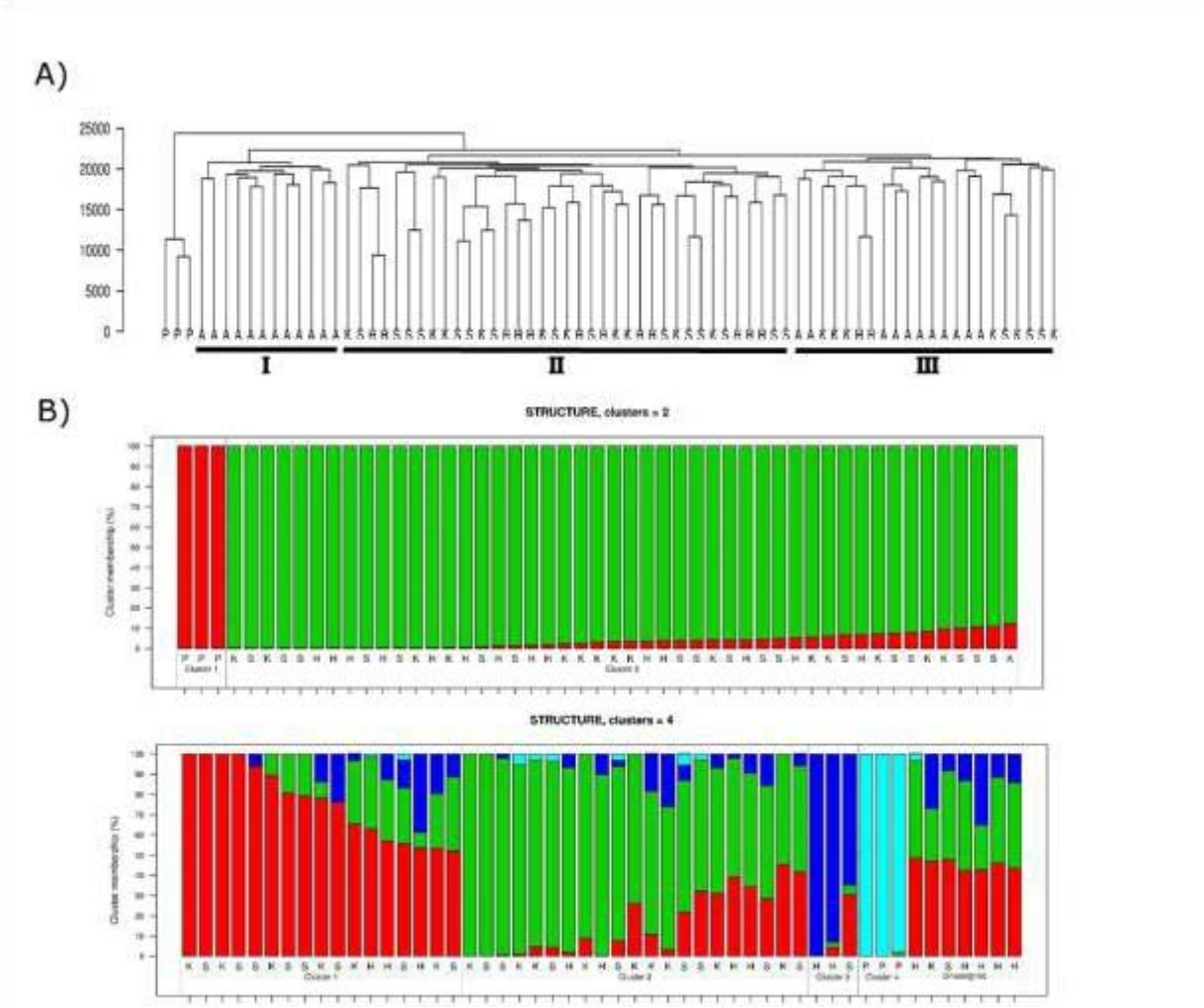


Figure 4.5. Hierarchical clustering and population structure of Syrian Arabian horses
A) Dendrogram showing the hierarchical cluster configuration of 71 Arabian horses including 48 from Syria, 23 from the USA (a single horse from the USA was excluded due to an error in the pedigree) and 3 Przewalski's horses. Arabian horses cluster in three groups: group I includes only American Arabian horses, group II comprises only Syrian Arabian horses and group III is an admix of Syrian and American Arabian horses. Horses are presented in the horizontal axis and genetic distances are represented in the vertical axis.
B) Clustering assignment of the 48 Syrian Arabian horses (three strains clustered in groups II and III (Figure 1A)) and three Przewalski's horses obtained by STRUCTURE analysis. Each individual is represented by a vertical bar; size and color correspond to the proportion of the horse genome corresponding to a particular cluster when K=2 (above panel) and K=4 (bottom panel). K=2 shows clear separation between Przewalski's and Syrian Arabian horses. K=4 (the best value of K) distinguishes three clusters among Syrian Arabian horses; 7 horses could not be assigned to any cluster. The cluster membership (%) is shown on the y-axis.
Breed abbreviations are S - Saglawi, K - Kahlawi, H - Hamdani, A - Arabian horses from the USA, P - Przewalski.

4.4. Genome-wide association study of morphological traits in Arabian horses

In this part, we studied association between the twelve corrected morphological traits and 14.920 informative SNPs passed the quality check on the SNP70 BeadChip in 37 Arabian horses representing the three major strains in Syria.

Analysis of Morphological traits showed a strong influence of strain on body length (BL), chest depth (ChD), chest width (ChW), neck girth (NG) and throat girth (TG) morphological traits (Table 4.8).

Table 4.8. Median values of morphological traits per strain, and significance of strain effects

Morphological trait	**Saglawi**		**Kahlawi**		**Hamdani**		**Strain P≤0.05**
	Median	**SD**	**Median**	**SD**	**Median**	**SD**	
Back line length (BLL)	97.4	1.12	99.0	1.88	98.0	2.18	0.32
Body length (BL)	160.0	7.57	149.0	6.46	160.0	8.64	**0.012**
Chest depth (ChD)	72.0	2.70	75.0	1.61	70.0	1.48	**<0.001**
Chest girth (ChG)	180.0	5.38	181.0	4.95	176.0	0.80	0.269
Chest width (ChW)	52.0	2.31	53.0	2.68	49.0	2.77	**0.011**
Croup height (CH)	151.5	1.62	152.0	2.29	151.0	3.00	0.589
Croup width (CW)	56.0	0.97	56.0	2.81	55.0	3.34	0.186
Fore cannon length (FCL)	20.3	0.89	20.2	1.02	20.0	0.85	0.386
Hind cannon length (HCL)	23.0	0.98	23.0	1.14	22.0	1.11	0.193
Neck girth (NG)	112.0	3.91	110.0	4.62	108.0	2.63	**0.033**
Throat girth (TG)	77.0	1.47	72.0	2.99	72.0	1.98	**<0.001**
Withers height (WH)	150.0	1.76	150.7	2.18	151.5	1.41	0.504

SD: standard deviation

Association analysis showed no significant association (LOD > 5.5), which can be attributed to the low sample size used in this study.

However, two morphological traits showed suggestive associations (LOD > 5). Neck girth showed an suggestive association (LOD = 5.18, P = 6.58x 10^{-6}) with a SNP BIEC2_772752 ($MAF_{(T)}$ of 0.49) on ECA3 at 13.7 Mb (13.763.517 bp) (Figure 4.6).

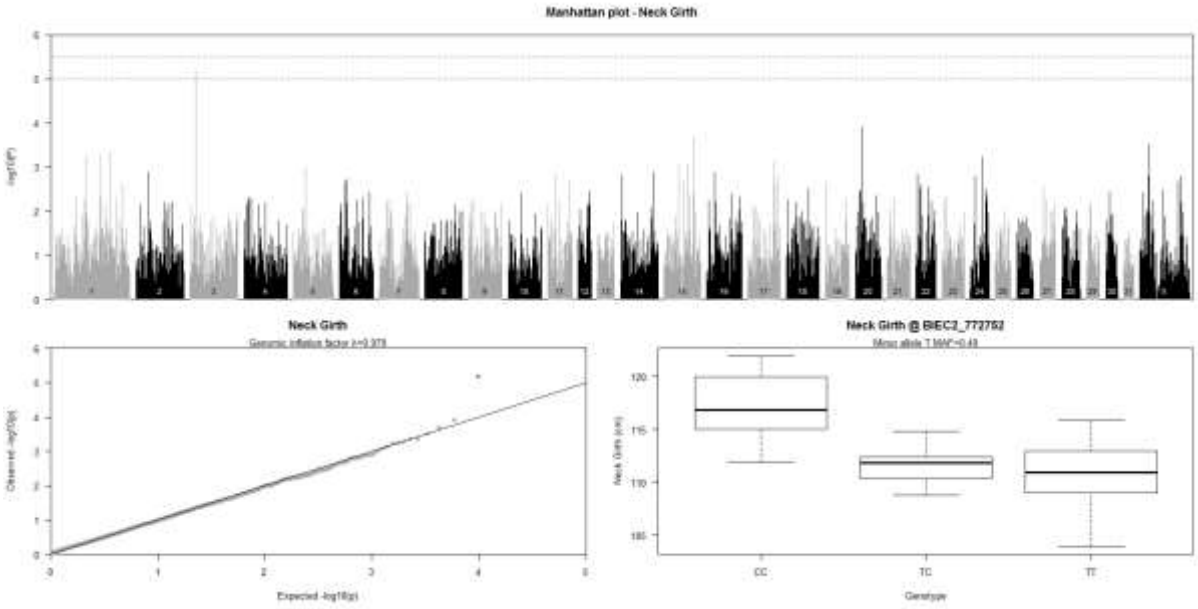


Figure 4.6. Manhattan plot, QQ plot, and boxplot of the suggestive associated SNP on ECA 3
Top: Manhattan plot for neck girth showing the suggestive associated SNP on ECA 3 in Syrian Arabian horses. Bottom Left: QQ plot of the expected versus observed LOD scores for neck girth. Bottom Right: Boxplot showing the dominant effect of The CC genotype on neck girth in our sample. (median $_{(CC)}$: 116.84, median $_{(TC)}$: 111.86, median $_{(TT)}$: 110.92; $n_{(CC)}$ = 10, $n_{(TC)}$ = 18, $n_{(TT)}$ = 9 individuals).

The second trait showing a suggestive association was chest width (LOD = 5.10, P = 6.95x10^{-5}) with a SNP BIEC2_444806 ($MAF_{(A)}$: 0.34) on ECA19 at 50.9 Mb (50.879.577 bp).

4.5. Test of candidate genes polymorphisms in endurance Arabian horses

The comparative sequence analysis of the candidate genes in 10 endurance Arabian horses led to identify 26 allelic variants, 13 in the mitochondrial genes and another 13 in the autosomal genes. The most promising SNPs were genotyped across the 42 horses.

4.5.1. Mitochondrial genes

The identified 13 mitochondrial variants in the exploratory 10 horses were genotyped further in 32 horses using RFLP for the *MT-ND4* variants and full sequencing for the *MT-COX3* and *MT_CYB* variants. Haplotypes of the mitochondrial genes *MT-COX3, MT-CYB* and *MT-ND4* were identified by combining variants and verified by pedigree information.

4.5.1.1. *MT-COX3*

We found in the 752 bp fragment of *MT-COX3* five synonymous variants, which were all transitions. Analysis of allele frequency showed that the G allele (mt:9241A>G) occurred most frequent across all horses (0.72). It was slightly higher in the group of high performing horses (0.75) compared to the low performing horses (0.66). The synonymous variant mt:9280T>C occurred only in the high performing group of Syrian Arabian horses (0.13) (Table 4.9).

Six Arabian horse *MT-COX3* haplotypes were defined with the five variants. Six individuals were identical to the reference sequence (EquCab2:ENSECAG00000027672) representing the haplotype HTCOX3_Ar0 (3 individuals with high and 3 with low performance). The most frequent haplotypes were HTCOX3_Ar3, which was present in 14 individuals (33 % of the sample set), nine with high endurance performance, and five with low performance, followed by the haplotype HTCOX3_Ar5 that was present in 12 individuals, six in each low and high performance groups. The *MT-COX3* haplotypes HTCOX3_Ar1, HTCOX3_Ar2, and HTCOX3_Ar4 were present in two, four and three individuals, respectively. HTCOX3_Ar1, HTCOX3_Ar2 haplotypes showed no specific affiliation to endurance performance, while HTCOX3_Ar4 was only present in three individuals with high performance. The haplotype

HTCOX3_Ar6 was unique in one individual with low performance. Haplotypes with their frequencies are listed in Table 4.10.

Novel haplotype sequences were submitted to GenBank. At the time point of printing this thesis, the accession numbers have not yet been provided.

4.5.1.2. *MT-CYB*

By sequencing 845 bp of the *MT-CYB* gene, two synonymous and one missense variants were found (Table 4.9). We have shown that the frequency of the variant mt:14628G>A in the Arabian horses with high endurance performance (0.38) was slightly higher than its frequency in the low performance group (0.33). The missense variant mt:14827G>A was present only in a single high performing individual. Twenty horses had the same haplotype as the reference, while 22 horses were grouped in three different *MT-CYB* haplotypes (Table 4.10). HTCYB_Ar1 was found in 6 individuals (3 with high and 3 with low performance), HTCYB_Ar2 was higher frequent with 15 individuals (9 with high and 6 with low performance). Both haplotypes show no affiliation to a specific endurance performance group. The HTCYB_Ar3 was unique to one individual with high performance.

Table 4.9. The detected variants in the mitochondrial genes (*MT-COX3, MT-CYB* and *MT-ND4*) of 42 Syrian Arabian horses, their locations and mutated alleles frequencies

Gene	Variant position	Reference> mutated allele	Amino acid change and positions	Variant effect	SNP ID	Allele frequency of mutated variant in Arabian horses			
						Sequenced individuals N=10	Total N=42	High performers N=24	Low performers N=18
MT-COX3	mt:8857	C>T	His71His	synonymous	-	0.50	0.5	0.50	0.50
	mt:9088	C>T	His148His	synonymous	-	0.50	0.50	0.50	0.50
	mt:9205	T>C	Thr187Thr	synonymous	-	0.30	0.38	0.50	0.33
	mt:9241	A>G	Val199Val	synonymous	rs394787910	0.50	0.72	0.75	0.66
	mt:9280	T>C	Ser212Ser	synonymous	-	0.20	0.07	0.13	-
MT-CYB	mt:14628	G>A	Thr147Thr	synonymous	(Sziszkosz et al. 2016)	0.20	0.36	0.38	0.33
	mt:14805	T>C	Asn206Asn	synonymous	-	0.20	0.50	0.50	0.50
	mt:14827	G>A	Asp214Asn	missense	(Sziszkosz et al. 2016)	0.10	0.03	0.04	-
MT-ND4	mt:10450	C>T	His82His	synonymous	-	0.50	0.50	0.46	0.55
	mt:10828	G>C	Pro208Pro	synonymous	rs869311971	1	1	1	1
	mt:10829	G>C	Leu209Val	missense	-	1	1	1	1
	mt:10861	C>T	Ala219Ala	synonymous	-	0.50	0.50	0.50	0.50
	mt:11242	C>T	Arg346Arg	synonymous	-	0.90	0.93	0.96	0.89

Table 4.10. Frequencies of haplotypes in the mitochondrial genes (MT-COX3, MT-CYB and MT-ND4) in 42 endurance Arabian horses

Mitochondrial genes	Name of HT	Position 1	Position 2	Position 3	Position 4	Position 5	Individuals per haplotype	Haplotype frequency	Gene Bank Accession ID
MT-COX3	HTCOX3_Ar1	8857_{T}	9088_{T}	9205_{C}	9241_{A}	9280_{T}	2	0.05	-
	HTCOX3_Ar2	8857_{T}	9088_{T}	9205_{T}	9241_{A}	9280_{T}	4	0.09	-
	HTCOX3_Ar3	8857_{T}	9088_{T}	9205_{C}	9241_{G}	9280_{T}	14	0.33	-
	HTCOX3_Ar4	8857_{C}	9088_{C}	9205_{T}	9241_{G}	9280_{C}	3	0.07	-
	HTCOX3_Ar5	8857_{C}	9088_{C}	9205_{T}	9241_{G}	9280_{T}	12	0.28	-
	HTCOX3_Ar6	8857_{T}	9088_{T}	9205_{T}	9241_{G}	9280_{T}	1	0.02	-
MT-CYB	HTCYB_Ar1	14628_{G}	14805_{C}	14827_{G}	-	-	6	0.14	-
	HTCYB_Ar2	14628_{A}	14805_{C}	14827_{G}	-	-	15	0.36	-
	HTCYB_Ar3	14628_{G}	14805_{T}	14827_{A}	-	-	1	0.02	-
MT-ND4	HTND4_Ar1	10450_{C}	10828_{C}	10829_{C}	10861_{C}	11242_{C}	3	0.07	KT757739.1 KT368740.1 KT368757.1 JN398384.1
	HTND4_Ar2	10450_{C}	10828_{C}	10829_{C}	10861_{C}	11242_{T}	16	0.38	JQ340142.1
	HTND4_Ar3	10450_{C}	10828_{C}	10829_{C}	10861_{T}	11242_{T}	2	0.05	-
	HTND4_Ar4	10450_{T}	10828_{C}	10829_{C}	10861_{T}	11242_{T}	19	0.45	-
	HTND4_Ar5	10450_{T}	10828_{C}	10829_{C}	10861_{C}	11242_{T}	2	0.05	-

4.5.1.3. *MT-ND4*

The analysis of 981 bp of the *MT-ND4* gene revealed four synonymous and one missense variants in the positions mt:10450C>T, mt:10828G>C, mt:10829G>C, mt:10861 C>T, mt:11242C>T (Table 4.9). Frequency of the variants showed that the transversions mt:10828G>C and mt:10829G>C were present in all individuals.

The *MT-ND4* variants defined five mitochondrial haplotypes. The major haplotypes were HTND4_Ar4 including 45% of individuals (n=19), and HTND4_Ar2 with 38% (n=16), while the three haplotypes HTND4_Ar1, HTND4_Ar3 and HTND4_Ar5 were present in less than 3 endurance horses (Table 4.10).

4.5.2. Autosomal genes

The identified 13 autosomal variants in the exploratory 10 horses were genotyped further for ACTN3 and MSTN genes in 32 Arabian horses.

4.5.2.1. *ACTN3*

By sequencing the promoter and 21 exons with flanking intron regions of *ACTN3* (approximately 8.979 bp), we found a total of 12 variants: one 5´ UTR variant, five intronic variants, four exonic synonymous variants, and two 3´ UTR variants (Table 4.11). No change in the amino acid was detected. Three *ACTN3* polymorphisms were genotyped further in 32 horses which are 12:26511704G>A, 12:26515885A>T, 12:26524894T>C using customized allele-specific PCR assays (Table 3.14).

In general, the more frequent variants were the intronic variant 12:26515885A>T (splice region variant), and the exonic variant 12:26524717A>G (synonymous variant) but no frequency differences have been detected between the high and low performance groups.

One variant in the 3´ UTR (12:26524930T>C) was analyzed for its potential effect on miRNA binding (eca-miR-1296 and eca-miR-326), additionally the TFBSs analysis of the *ACTN3* 5' UTR 12:26511704G>A was predicted to cause gaining of an E4BP4 binding site (Table 4.12).

Table 4.11. **The detected variants in the autosomal genes (*ACTN3* and *MSTN*) of Arabian horses, their locations and the mutated alleles frequencies**

Gene	SNP position	Reference> mutated allele	Amino acid change and positions	Variant effect	SNP ID	Frequency of the mutated allele in Arabian horses			
						Sequenced individuals N=10	Total N=42	High performers N=24	Low performers N=18
ACTN3	12:26511704	G>A	-	5´ UTR	I1 (Thomas et al. 2014)	0.50	0.42	0.33	0.53
	12:26515793	C>G	-	Intronic (I12)	rs68947239	0.45	0.43	0.46	0.39
	12:26515795	C>T	-	Intronic (I12)	rs68947240	0.45	0.43	0.46	0.39
	12:26515807	T>C	-	Intronic (I12)	I4 (Thomas et al. 2014)	0.45	0.43	0.46	0.39
	12:26515885	A>T	-	splice region (I12)	rs68947241	0.45	0.43	0.46	0.39
	12:26515942	C>T	Ile105Ile	Synonymous (E3)	rs68947242	0.45	0.43	0.46	0.39
	12:26516020	G>C	-	splice region(I3)	rs68947243	0.45	0.43	0.46	0.39
	12:26519406	A>G	Pro366Pro	synonymous (E10)	E3 (Thomas et al. 2014)	0.60	-	-	-
	12:26524504	T>C	Ala814Ala	synonymous (E20)	rs394353570	0.95	-	-	-
	12:26524717	A>G	Leu858Leu	synonymous (E21)	E6 (Thomas et al. 2014)	0.30	0.61	0.58	0.64
	12:26524894	T>C	917	3´ UTR	rs396350893	0.70	0.39	0.42	0.36
	12:26524930	T>C	929	3´ UTR	rs396948497	0.30	0.61	0.58	0.64
MSTN	18:66495696	A>G	-	Upstream	(Stefaniuk et al. 2016)	0.15	0.09	0.10	0.08

E: Exon, I: Intron

Table 4.12. Potential effects of variants within the untranslated and promoter regions of *ACTN3* and *MSTN* genes on miRNA and transcription factor binding sites (TFBSs)

Gene	Variant position	Location of variant	Effect of the allelic change	
	miRNA, gain/loss of binding	**TFBSs,** gain/loss of binding		
ACTN3	12:26511704G>A	5´ UTR	-	E4BP4 [+]
ACTN3	12:26524930T>C	3´ UTR	eca-miR-1296 [-] eca-miR-326 [-]	-
MSTN	18:66495696A>G	Promoter	-	HFH-1 [-] Sox-5 [-] HFH-3[+] E4BP4[+]

TFBSs: transcription factor binding sites, [+]=gain, [-]=loss

4.5.2.2. *MSTN*

By sequencing the promoter and three exons with their flanking intron of *MSTN* in 10 horses, one transition was detected at 18:66495696A>G within the promoter region. The *MSTN* promoter polymorphism 18:66495696A>G was genotyped in further 32 horses using customized allele-specific PCR assays (Table 3.14). The alternative allele G frequencies of the variant 18:66495696A>G are listed in Table 4.11.

The analysis of TFBSs of the *MSTN* promoter substitution of A to G at the position 18:66495696 is predicted to cause substitution of two binding sites for HFH-1 and Sox-5 by two binding sites for HFH-3 and E4BP4 (Table 4.12).

Four autosomal polymorphisms were genotyped further in 32 horses which are 12:26511704G>A, 12:26515885A>T, 12:26524894T>C in the *ACTN3* gene, as well as 18:66495696A>G in the *MSTN* gene using customized allele-specific PCR assays (Primers are listed in chapter 3.2.7).

5. Discussion

5.1. Maternal diversity and phylogeny of three Arabian horse strains

This study was designed to test the maternal genetic diversity of 192 Syrian Arabian horses sampled from three strains Saglawi (n=57), Kahlawi (n=75) and Hamdani (n=60) based on the partial mitochondrial D-loop sequencing approach. We found 38 variations in the mitochondrial D-loop region that had been reported previously in different equine breeds. Out of 28 derived mitochondrial haplotypes, 25 had been described before in other horse breeds including Arabian horses (Bowling *et al.* 2000; Kavar *et al.* 2002; Lopes *et al.* 2005; Głażewska *et al.* 2007; Achilli *et al.* 2012; Khanshour & Cothran 2013). Three novel haplotypes (K012, K006 and S003) that were detected in the present study occur in four maternal lineages in Kahlawi and in two maternal lineages in Saglawi according to the recorded pedigree information. The two haplotypes KS01 and SKH1 that segregate in the three strains suggest common maternal ancestors.

The results show that Kahlawi is the most diverse strain in our study. The high haplotype diversity found in Kahlawi (0.95) is consistent with previous reports for the same strain (0.94) (Khanshour & Cothran 2013). Hamdani and Saglawi show a little lower diversity (0.91 and 0.90, respectively). A former study reported that the haplotype diversity of Saglawi was equal to Kahlawi, while the lowest value was found in Hamdani (0.87) (Khanshour & Cothran 2013). These former studies also show the importance of the breeding region. Arabian horses bred in Syria (all strains) present a haplotype diversity of 0.96 while Arabian horses from the USA have a diversity of 0.79 to 0.83 and Polish Arabian horses 0.82 (Khanshour & Cothran 2013). Nevertheless, the haplotype diversity in our study is close to

the average value reported for other horse breeds like for North China breeds (0.95), South China breeds (0.92), and Italian breeds (0.900 till 0.991) (Bigi *et al.* 2014; Yang *et al.* 2016).

The pair-wise population differentiation index (F_{ST}) provides evidence for little genetic differentiation between strains. The biggest genetic differentiation was found between Saglawi and Hamdani (0.205) compared to Kahlawi *versus* Hamdani (0.138) and Kahlawi *versus* Saglawi (0.098). These findings suggest that individuals of Kahlawi and Saglawi strains are more closely related than the other strains and they share more frequently their maternal haplotypes. Similar results were found for F_{ST} values between different Arabian horse populations world-wide (Khanshour & Cothran 2013).

The analysis of molecular variance showed low genetic variation between the three Syrian Arabian horse strains (14.18%), while substantial differentiation was found within strains (85.82%). Other data for Arabian horses explain 8.25% of genetic variation among and 91.75 % within populations (Khanshour & Cothran 2013). In comparison to Arabian horses, 21.06 % of genetic variation was found among three Indian horse breed populations, for example (Chauhan *et al.* 2011). Therefore, our results provide further evidence for high variability of maternal mitochondrial DNA within each Syrian Arabian horse strain while the genetic differentiation between the three strains is low.

These results are additionally confirmed by the phylogenetic analyses of the neighbor-joining tree. The studied individuals have low distance values and show no clear clustering. No sub-structure of the proposed three strains could be found. Thus, the Arabian horses from the three strains are strongly related to each other. The Kahlawi individuals were scattered all over the tree clusters which could be interpreted as Kahlawi being common maternal ancestors of all three strains. This substantiates our findings which show no

evidence that the Arabian horse strains are genetically separated. The clustering of our haplotypes into seven haplogroups known for other horse breeds (Achilli *et al.* 2012) in the median-joining network shows that Syrian Arabian horses share their mitochondrial haplotypes in cluster A with breeds such as Caspian pony, Maremmano, Akhal-Teke and Iranian; in cluster B with Westphalian, Maremmano and Italian; in cluster C with Akhal-Teke and Suffolk pony; in cluster G with Akhal-Teke, Iranian, Asian Naqu, Giara and Italian; in cluster M with Akhal-Teke, Caspian pony, Maremmano and Friesian; in cluster P with Iranian and Caspian pony; and in cluster Q with Akhal-Teke, Maremmano, Iranian and Asian Deqin. Only the Przewalski's horse split from the Arabian and other horse breeds in the separate cluster F. Neither breed nor geographical clusters could be found among the many known equine mitochondrial DNA variants (Yang *et al.* 2016). Clusters exist only for animals in very isolated populations with limited ancestors (Devi & Ghosh 2013; Winton *et al.* 2015; Cardinali *et al.* 2016). This high maternal diversity was reflected in a study of seven Arabian horse strains with 191 animals where 12 haplogroups containing 55 haplotypes were found (Khanshour & Cothran 2013) as well as in our study with Syrian Arabian horses.

This shows that the genetic scenario observed for the three Syrian strains is more likely the result of contributions from multiple female ancestors, share common maternal ancestors with other breeds. One hypothesis assumes that the Arabian horse breed has a heterogeneous origin of a wide variety of horses from Central and East Asia is possibly correct (Cieslak *et al.* 2010; Głażewska 2010). The opposite probability that the common haplotypes originate from Arabian mares that were introduced into all the other breeds is rather small. Since archeological material of Arabian horses has not been analyzed yet, it is impossible to decide whether the current Arabian horse mitochondrial haplotypes were present in the oldest ancestral material or not. In horses, a high number of maternal

mitochondrial lines leading to a high diversity is a common case. That makes it difficult to find ancestors and subdivisions.

5.2. Genetic variations at microsatellite loci in three Arabian horse strains

In this study, we used a set of twelve recommended equine microsatellites to investigate the genetic structure of 84 Arabian horses samples descended from the three main Arabian horse strains in Syria, which are Hamdani(n = 26), Kahlawi(n = 30) and Saglawi(n = 28).

The average number of alleles (N_a), and polymorphic information content (PIC) estimates per locus per strain showed that all microsatellites were polymorphic. Our average number of alleles N_a values ranged between 6.333 and 7.583 which were close to those detected in Middle-Eastern Arabian horses (5.130 to 8.470) (Khanshour *et al.* 2013), African Arab-Barb horses (7.86) (Berber *et al.* 2014), and differ slightly from those reported in Arabian horses from Iran and Charmahal-va- bakhtiari province (3 to 9) (Moshkelani *et al.* 2011; Mostafa *et al.* 2011). The PIC values reported in the current study (0.383 to 0.890) were also close to the reported in the Iranian Arabian horses (0.402 to 0.764) (Moshkelani *et al.* 2011). The slight differences can be due to the differences of population structure, number of samples, as well as the chosen microsatellites panel.

Our studied loci were more polymorphic in Saglawi strain (N_a= 7.583, PIC=0.743) which means that Saglawi retained the greatest amount of genetic variation compared to Kahlawi and Hamdani strains. Additionally, Saglawi strain possessed the highest number of private alleles (N_{pa}=12) which is noticeably different from the observed numbers in Kahlawi strain (N_{pa}= 9) and Hamdani strain (N_{pa}=5). This shows that Saglawi contributes positively to the total polymorphic information content and more than the other two strains.

Whereas, the average effective number of alleles (N_e) for the three strain per loci (3.797 to 4.582) falls in the range reported by Khanshour *et al.* (2013) for Middle-Eastern and Western Arabian horses (3.51 to 4.23), but were higher than the N_e values reported by Rukavina *et al.* (2015) for Arabian horses in Bosnia and Herzegovina (2.04, 4.08).

The expected (0.722) and observed (0.680) heterozygosity values of the studied breeds did not differ significantly from those reported in previous studies reported heterozygosity values for horse breeds share origin with Arabian horses, for example in Barb horses(H_o= 0.75 and H_e= 0.75), Arab-Barb horses (H_o= 0.73 and H_e= 0.77), English thoroughbred (H_o =0.71 and H_e = 0.71), Akhal Tek (H_o =0.72 and H_e = 0.65), Iranian Arabian (H_o =0.70 and H_e = 0.71), Polish Arabian (H_o =0.69 and H_e = 0.68) (Cho 2006; Khanshour *et al.* 2013; Berber *et al.* 2014).

Results showed that H_o estimates were less than the H_e, which is rather expected, because the traditional breeders avoid strictly outcrossing with foreign or unknown horses. This suggests that our studied Arabian horses have received less external gene flow (Crawford 2007), but fortunately, still have high heterozygosity. Both, alleles polymorphisms and high heterozygosity which are essential for the future evolution and selection of the Arabian horse population (Hill & Rasbash 1986).

The inbreeding coefficients (F_{IS}) were lower than zero, implying some relatedness among ancestors. This is most evident in Saglawi strain (0.123). Overall F_{IS} reported in this study (0.053) is a bit higher than that reported in Syrian Arabian horses (-0.007), and Iranian Arabian horses (0.017), but close to that reported in Egyptian Arabian horses in the USA (0.047) (Khanshour *et al.* 2013).

AMOVA results revealed that only 0.045 of the genetic variation was attributed to strain differences, as evidenced by the low pairwise F_{ST} values across the three strains. F_{ST} values were close to each other and to the overall value (0.019) showing a low level of population stratification of the three strains.

The Neighbor-joining dendrogram shows Kahlawi strain (POP1) clustering with Saglawi strain (POP3), while Hamdani (POP2) branched off slightly away from Kahlawi and Saglawi strains. Interestingly, in a former study of maternal lineages diversity of the same strains (Almarzook *et al.* 2017b), showed that the lowest pairwise F_{ST} value was between Saglawi and Kahlawi (0.098), while a comparison of Hamdani with Saglawi and Kahlawi showed higher pairwise F_{ST} values (0.205 and 0.138, respectively), which suggested that individuals of Kahlawi and Saglawi strains are sharing maternally closely related ancestors. However, the genetic distances suggest that the three studied strains are closely related.

The distribution of individuals in the PCoA figures indicates high similarities between the three strains, even though, the greatest amount of variation (8%) was captured on axis 1, but no clear pattern of subdivision was observed. This result is in agreement with the pairwise F_{ST} estimates. Furthermore, the samples were collected from different geographical regions in Syria, but the figures represent a possible blood accessibility outwards and inwards strains due to mating horses from different strains.

We think that this method did not identify a uniform of each of the studied strains as the legends of horse history assumed. Increasing the number of markers (Koskinen et al. 2004), or using other method (e.g. SNP array) with better cover of the whole genome are beneficial for better population characterization.

5.3. Genetic diversity in using medium density genome-wide SNP array

For this study, 48 Syrian Arabian horses representing the three strains Saglawi, Kahlawi and Hamdani were genotyped using SNP70 BeadChip. Additional genotypes of 24 Arabian horses from the USA, and three Przewalski`s horses were available for comparison. The quality check of the total 65.156 SNPs, ended with 38.671 informative SNPs.

The filtered genome-wide SNPs enabled assessment of genetic diversity status of the studied strains. The observed heterozygosity (H_o) and the expected heterozygosity (He) values of the three strains are in agreement with recently reported SNP chip data for Arabian, Akhal Teke, Andalusian and Tuva horses (H_e=0.31), for example (Petersen *et al.* 2013a). However, the results obtained in our study using 38.671 informative SNP markers indicate less heterozygosity (expected and observed) than estimated for Syrian and Egyptian Arabian horses which were genotyped for 5 to 15 microsatellite markers, where observed and expected heterozygosity ranged between 0.70 and 0.81 and between 0.69 and 0.75, respectively (Mahrous *et al.* 2011; Khanshour *et al.* 2013).

The inbreeding coefficients F_{IS} are showing that individuals within a strain are less related than under a model of random mating.

The F_{ST} values of the same magnitude were also reported for geographically divided populations like Thoroughbred in UK/Ireland and in USA (0.004), for closely related breeds like Tuva and Mongolian (0.006) as well as for Swiss Warmblood and Hanoverian (0.008) (Petersen *et al.* 2013a). Slightly higher F_{ST} values, but still close to zero, were found between the three Syrian strains and the examined American Arabian horses, where F_{ST} values ranged between 0.025 and 0.040 (Table 1). This is in agreement with F_{ST} values of 0.050 for the population differentiation between Syrian and Egyptian Arabian horses from USA based on mitochondrial D-loop sequencing (Khanshour & Cothran 2013).

In the hierarchical clustering analyses using Manhattan distance, in contrast to our expectations, the three Syrian strains Saglawi, Kahlawi and Hamdani did not cluster separately. All three strains were spread across all clusters containing Syrian horses.

Despite of the high marker density across the whole genome, clear differences between the strains could not be identified. The high density data just further strengthen the findings obtained from analyses of mitochondrial DNA on the maternal side of inheritance as shown for the same three Syrian strains (Almarzook *et al.* 2017b) and seven Middle Eastern strains (Khanshour & Cothran 2013).

Moreover, the hierarchical clustering also showed that one group of American Arabian horses was distinct from the examined Syrian strains while the others were genetically closely related to Syrian horses and likely have their origin in Syria, which is a commonly accepted theory (Bowling *et al.* 2000). Group II comprised only Syrian Arabian horses, which, therefore, provides a genetic reservoir for genetic diversity and it is an important source for worldwide horse breeding.

In the STRUCTURE analyses, when K=2, not more than 10% of alleles affiliated with Przewalski's horses were found in the cluster of Arabian horse strains. This small amount of shared alleles was expected and mirrors common ancestral alleles (Cieslak *et al.* 2010).

With respect to the myth of three founders of the three Syrian strains, we expected that the best STRUCTURE model (K=4) would show three clusters, one for each of the Syrian Arabian horse strains and a fourth cluster for Przewalski's horses as an out-group. Indeed, K=4 revealed three clusters among the examined Syrian horses. However, the assignment of individuals to distinct clusters was not consistent with the current strain designation.

We like to add that the SNPs on the chip, as well as the decision to remove SNPs with minor allele frequencies < 0.05 from the analyses could eventually affect the outcome of the

study, since such SNPs could have contributed to the selection of the strains. Nevertheless, this is very unlikely given the large data set used in this analysis.

5.4. Genome-wide association study of morphological traits in Arabian horses

Using 14.920 informative SNPs and data of twelve corrected morphological traits, we did the genome wide association study of 37 Syrian Arabian horses.

The significance of strain effects on body length (BL), chest depth (ChD), chest width (ChW), neck girth (NG) and throat girth (TG) support the common knowledge of phenotypical variations among Arabian horse strains (Raswan et al. 1981), although we did not detect genetic differentiation between the three main Syrian strains in our previous analyses (Almarzook *et al.* 2017a; Almarzook *et al.* 2017b).

During the course of this study, we pursued many different strategies to get a consistent view on the identified suggestive genomic associations such as pruning of LD. For example, when we lowered the accepted LD threshold from r^2=0.5 to r^2= 0.2, more markers were removed during the LD pruning step, which led to the suggestive associations reported here to become significant. However, we chose to not massage our data too much, and keep a balance between inclusion of potentially interesting markers and exclusion of markers with a high probability of creating false associations.

The CC genotype on this SNP BIEC2_772752 ($MAF_{(T)}$ of 0.49) seems to have a dominant effect of 6.0 cm on the neck girth. Interestingly, the SNP associated with the neck girth in our study is located on ECA 3 at 13.763.517 bp. This locus is not linked to a region around 105 Mb, for which previously the *LCORL/NCAPG* locus has been reported as a candidate locus for body size and withers height in Tennessee Walking horses, German Warm Blood

horses, Franches-Montagnes, and Thoroughbreds (Makvandi-Nejad *et al.* 2012; Signer-Hasler *et al.* 2012; Tetens *et al.* 2013; Staiger *et al.* 2016).

Chest width was suggestively (LOD = 5.10, P = $6.95x10^{-5}$) associated with the SNP BIEC2_444806 ($MAF_{(A)}$= 0.34) on chromosome 19 at 50.879.577 bp. The genetic effect of this SNP showed an underdominant behavior where the heterozygous genotype is associated with a smaller chest width compared to the two homozygous genotypes ($median_{(AA)}$ = 52.3 cm, $median_{(AG)}$ = 50.0 cm, $median_{(GG)}$ = 53.9 cm).

We checked, if functional candidate genes reside in a region of 1 Mb in both directions around the identified genomic loci for neck girth and chest width. However, in the neighborhood of the detected SNPs no candidate genes of interest are known.

5.5. Candidate genes polymorphisms in endurance Arabian horses

5.5.1. Mitochondrial genes

It is widely known that mitochondria affect the skeletal muscles metabolism efficiency, and consequently, mutations in the mitochondrial genes are involved in the athletic performance (Hood 2009). In a study by Harrison and Turrion-Gomez (2006), the *MT-COX3* gene showed no allelic variation in Thoroughbred racing horses. In the current study, five *MT-COX3* synonymous variants were detected in the Syrian Arabian horses. Previously, the high frequency variant mt:9241A>G (rs394787910) has been validated in Quarter American horses (dpSNP, NCBI, DINDOTLAB). The variants mt:8857C>T, mt:9088C>T, mt:9205T>C, and mt:9280T>C were detected in different horse breeds including China native horse, Thoroughbred, and Przewalski horse. All of *MT-COX3* variants found in this study are reported in unpublished researches and have no registration IDs yet. The variants mt:8857C>T, mt:9088C>T, mt:9205T>C showed no specific affiliation to high or low

endurance performance groups, while mt:9280T>C (Ser212Ser) occurred only in the high performing group of Syrian Arabian horses (0.13).

Two of the three variants detected in *MT-CYB* (mt:14628G>A and mt:14827G>A) have been recently identified in a genetic diversity study of the Hungarian Gidran horse (Sziszkosz *et al.* 2016). It is known that Hungarian Gidran is a variety of the Anglo-Arab horse descending from crossing an Arabian stallion with an Andalusian mare in the nineteenth century (Rousseau 2017). Therefore, both variants mt:14628G>A and mt:14827G>A could be original polymorphisms in the Arabian horse breed.

The variant mt:14827G>A resulted in Aspartate to Asparagine substitution at the amino acid position 214. Aspartate and Asparagine are both polar amino acids. Asparagine differs from Aspartate only in an amino group instead of oxygen found in Aspartate. As Asparagine and Aspartate have similar properties (Betts & Russell 2003), we do not think that this variant (mt:14827G>A) could lead to any remarkable effect on the protein function.

Interestingly, no *MT-ND4* sequence of the sequenced Syrian Arabian horses was identical to the reference of the Thoroughbred horse. The identified five variants mt:10450C>T, mt:10828G>C, mt:10829G>C, mt:10861 C>T, mt:11242C>T were almost equally distributed between the high and low performance groups. The two transversions mt:10828G>C and mt:10829G>C were fixed in all Syrian Arabian horses.

The variant mt:10829G>C led to substitution of Leucine to Valine (Leu209Val) which is predicted as a missense variant. The Leucine and Valine are both hydrophobic amino acids, which have binding ability of hydrophobic ligands such as lipids (Betts & Russell 2003; Creixell et al. 2012). Due to their very similar chemical properties, we do not think that the Leucine to Valine exchange does dramatically alter the gene function.

In a former study with the same Syrian Arabian horses, we identified 20 mitochondrial haplotypes for the mitochondrial D-loop region (Almarzook *et al.* 2017b). The current study, where we sequenced three mitochondrial genes of these horses, showed that horses carrying the same haplotype in a mitochondrial gene are not necessarily having the same D-loop haplotype. This observation is supported by Harrison and Turrion-Gomez (2006), who analyzed a set of mitochondrial genes including *MT-CYB and* D-loop in Thoroughbred horses representing 33 world-wide distributed maternal lines.

We do not know, which mitochondrial mutations can have a considerable effect on the endurance performance. Studies in human endurance performance provided controversial results where some studies suggest that sequence variation in mitochondrial DNA may contribute to individual difference in endurance capability (Dionne *et al.* 1991). In contrast, Rivera *et al.* (1998) found no association between high endurance status and mtDNA polymorphisms.

5.5.2. Autosomal genes

If we underlay one mutation every 644 to 891 bp in horses (Orlando et al. 2013), we would expect 10 to 14 variants in the *ACTN3* gene. Although *ACTN3* as a functional gene is expected to be conserved, previous studies revealed variants which infer possible functional changes (Thomas et al. 2014). In the current study in Syrian Arabian horses, we detected 12 variants, which is consistent with our expectation assuming an average mutation rate.

The *ACTN3* 5' UTR variant 12:26511704G>A showed significant frequencies differences between four equine phenotypes including endurance, sprint, pace, and strength. The A-allele is overrepresented in the strength (Clydesdale and Shire breeds, frequency=77%) and

pace horses (Standardbred breed, frequency=69%), compared to sprint (Thoroughbreds, frequency=17%) and endurance horses (American Arabian, 38%) (Thomas *et al.* 2014). The A-allele frequency in our current study of Syrian Arabian horses (frequency=42%) is consistent with their findings.

The variant 12:26511704G>A is predicted to cause gain of a binding site for the E4 promoter–binding protein 4 (E4BP4), a basic leucine zipper transcription factor. E4BP4 regulates circadian rhythm by competing for DNA binding with a member of the related PAR family of basic leucine zipper transcription factors. E4BP4, also known as nuclear factor interleukin 3 (NFIL3) is thought to affect exercise in the skeletal muscles (Bottinelli & Reggiani 2007).

Based on findings by Thomas et al. (2014), the *ACTN3* exonic variants 12:26515942C>T (Exon3), 12:26519406A>G (Exon10), and 12:26524717A>G (Exon21) are assigned to three conserved domains (Calpomin homology, Spectrin repeats, and the two EF-hands, respectively), which have an important role in calcium ion binding which supports the protein structure (Djinovic-Carugo *et al.* 2002; Parry & Squire 2005). The 3´ UTR variants 12:26524894T>C and 12:26524930T>C have no direct effect on the gene function. No *ACTN3* allele frequency differences were detected in our study between high and low endurance performance horses. However, all variants in the 5´ and 3´ UTR were analyzed for their potential effects on miRNA binding. Interestingly, one variant in the 3´ UTR of *ACTN3* (12:26524930T>C) is located within two predicted miRNA target sites (eca-miR-1296 and eca-miR-326). The variant C is responsible for abrogation of these miRNA sites, which might affect their post-transcriptional regulation and consequently the gene expression.

In the *MSTN* gene, the promoter variant 18:66495696A>G has been previously reported in different horse breeds including Arabian, Thoroughbred, Polish Konik and Hucul horses (Binns et al. 2010; Dall'Olio et al. 2010; Hill et al. 2010c; Tozaki et al. 2012; Stefaniuk et al. 2014). The G-allele has a frequency of 0.23 in Arabian horses in Poland (Stefaniuk et al. 2014). Furthermore, 18:66495696A>G is suggested to be associated with height at withers in Arabian horses and Uruguayan Creolo horses (Dall'Olio *et al.* 2012; Stefaniuk *et al.* 2016). In the current study, the frequency of the alternative allele at this SNP was 0.09, and it did not differ considerably between high and low performing groups.

Different studies reported promoter variants of the equine *MSTN* gene. Among those promoter variants, 18:66495826A>G was also found in Hucul, Polish Heavy Draft, and Thoroughbred horses (Stefaniuk *et al.* (2014). Studies in Thoroughbred horses showed an association between racing performance phenotypes and a promoter insertion polymorphisms (227 bp SINE insertion located at -373/-147 bp from the start codon ATG). Animal homozygous for the SINE insertion allele were most frequent short distance racing, heterozygous allele carriers were more frequent in horses performing at middle-distance, while homozygous carriers for the wild type allele were most often found in long-distance endurance races(Hill *et al.* 2010c; Dall'Olio *et al.* 2014). The SINE insertion is suggested to affect the *MSTN* gene expression (Santagostino *et al.* 2015b). None of the above promoter variants are observed in the Syrian Arabian horses. Interestingly, the intronic variant 18:66493737T>C has been widely known for its strong positive association with short racing distance, speed and body composition in different racehorse breeds (e.g. Thoroughbred) (Binns *et al.* 2010; Hill *et al.* 2010a; Hill *et al.* 2010c; Tozaki *et al.* 2011; Hill *et al.* 2012b; Tozaki *et al.* 2012). In all examined Syrian Arabian horses of the current study, this locus was homozygous for the reference genotype (TT). Homozygosity for the T allele

was also found in Arabian horses sampled in Poland (Stefaniuk *et al.* 2016). In another study, the variant 18:66492906A>C was found in the second exon of *MSTN* within a unique haplotype (EU241341) present in 12 of 19 Arabian horses (Baron et al. 2012). In our Syrian Arabian horses, the three *MSTN* exons were conserved and identical to the reference. Same to our results, *MSTN* exons were conserved in 96 Arabian horses from Poland (Stefaniuk *et al.* 2016).

The TFBSs analysis of the *MSTN* promoter substitution of A to G at the position 18:66495696 predicted losing The Helix factor hepatocyte nuclear factor-3 homologue 1 (HFH-1) and the sex-determining region SRY-box 5 (Sox-5) transcription factor binding sites (UniProt) and gaining both of Hepatocyte Nuclear Factor 3 Forkhead Homolog 3 (HFH-3) and E4 promoter–binding protein 4 (E4BP4). HFH-1 is one of the nuclear factors that share a conserved DNA-binding domain (winged helix domain). It is thought to have an impact on the nerve and skeletal systems in mammals, particularly during the early stages of the embryo development (Altaba et al. 1993; Hong et al. 1999; Hoggatt et al. 2000). HFH-1 can act as inhibitor for the abundant protein found in smooth muscle-specific promoters (Hoggatt *et al.* 2000). Smooth muscles are not directly involved in the athletic performance, but their contractions are regulating the internal organs (e.g. blood vessels).

The Sox family of transcription factors is involved in regulating cell development and tissue regeneration (Sarkar & Hochedlinger 2013). The transcription factor Sox-5 (sex-determining region SRY-box 5) is considered to be an enhancer of chondrogenesis; as such it controls the correct development of the skeleton (Smits et al. 2001).

Hepatocyte Nuclear Factor 3 Forkhead Homolog 3 (HFH-3 or FoxO1) is a member of the forkhead domain transcription factors family, which are all expressed in skeletal muscle

(Sanchez *et al.* 2014) and involved in a wide range of cellular functions including energy metabolism (Ogg et al. 1997). HFH-3 has a vital role in development the sense of balance, mediating the formation of fatty acids and glycerol to be consumed by muscle cells under exercise, regulating of glucose metabolism in skeletal muscle, as well as, muscle energy homeostasis (Furuyama *et al.* 2003; Sanchez *et al.* 2014).

6. Conclusions and outlook

The present study illustrates how variable the three major Syrian Arabian horse strains are in light of the mitochondrial D-loop variability. The high number of mitochondrial haplotypes suggests that the Syrian Arabian horse has more than five maternal ancestors, which opposed the myth of the five founder mares. The phylogenetic analyses show a mixed clustering of the mitochondrial haplotypes, which does not support the proposed sub-division based on the strains designation system. The dominance of Kahlawi haplotypes in all clusters can be the result of introgression, which has not been documented. Although the Syrian Arabian horse strains are still classified according to the historical designation system, our findings of the mitochondrial variability suggest a new definition of strains.

In the second part of the study, the twelve microsatellites markers analyses confirmed a general genetic feature of the three strains analyzed, suggesting a low level of population differentiation and the three Syrian Arabian horses strains appeared to be genetically related. To ascertain the results, it will be necessary to collate our findings with the historical pedigrees. Using the Equine SNP70 BeadChip, we conclude from the analysis of autosomal SNPs that the hypothesis of three founders in the population of the examined Syrian Arabian horse strains can not be held and must be rejected. However, the results do

not support the one strain–one founder assumption. The genomic composition of individuals within every strain results from a mating system by which stallions are mated independently of their strain origin. This is the major reason why the current strain designation system is not supported by genetic distances.

In the genome-wide association study of twelve morphological traits in Arabian horses, we show that using proper quality control of SNPs and morphological data led to the detection of suggestive associations for neck girth and chest width in Syrian Arabian horses. Future investigations have to be done using a larger number of measurements data to show if these SNPs are real markers for the selected traits.

The genetic background of endurance performance of Syrian Arabian horses is not determined yet. This research shows polymorphisms in three mitochondrial (*MT-COX3, MT-CYB* and *MT-ND4)* and two autosomal genes (*ACTN3* and *MSTN*) which are thought to be involved in the endurance performance. The comparative sequencing identified 13 variants in the analyzed mitochondrial genes and 13 variants in the autosomal genes. Only one mitochondrial variant mt:9280T>C in *MT-COX3* occurred in the high performing horses. We could not detect any significant difference in allele frequencies between the two performance groups for sequence variants of the *ACTN3* and *MSTN* genes.

Due to the importance of the TFBSs as key regulators of genes expression, studies underscored their power in enhancing the biological functions. Endurance is a complex trait, which is controlled by hundreds of genes that are working in concert mediated by transcription factors. It is not known to which extent the gain of the E4BP4, as well as the substitution of HFH-1 and Sox-5 with HFH-3 and E4BP4 transcriptions factors, could affect the endurance performance of the Arabian horses in the current study. However, this study

contributes to the knowledge of candidate genes variants that are related to endurance performance in Syrian Arabian horses. Detection of the genetic background of the endurance-related genes in Arabian horses is still a question to be answered. This can be achieved if we test the association between polymorphisms of a multi-gene panel in a larger group of endurance Arabian horses using the appropriate endurance data. To generate the appropriate data, we have to collect standard endurance records of multiple years of horses that performed endurance over endurance distances. The information on the identified polymorphisms of the candidate genes could be beneficial in improving selecting and breeding programs.

7. Bibliography

Achilli A., Olivieri A., Soares P., Lancioni H., Kashani B.H., Perego U.A., Nergadze S.G., Carossa V., Santagostino M. & Capomaccio S. (2012b) Mitochondrial genomes from modern horses reveal the major haplogroups that underwent domestication. *Proceedings of the National Academy of Sciences* **109**, 2449-54.

Adamu L., Adzahan N.M., Rasedee A. & Ahmad B. (2014) Physical parameters and risk factors associated with the elimination of arabian and crossed arabian endurance horses during a 120-km endurance race. *Journal of Equine Veterinary Science* **34**, 494-9.

Adrian J.A.L. (2013) Equine Color Genetics and Deoxyribonucleic Acid Testing. *Journal of Veterinary Science & Technology* **2013**.

AHA (2015) Arabian Conformation. Arabian Horse Association, Aurora, Colorado,USA.

Ahmed M.M., Amer S.A. & Sayed S.M. (2011) Molecular investigation of the Arabian horse breeds of racing, productivity and longevity. *Advances in Bioscience and Biotechnology* **2**, 450.

AL Shaieb M.G. (2016) Syrian Arabian horse strains in numbers. Arab Horses Office, Syrian Ministry of Agriculture and Agrarian Reform Syria.

AL Shaieb M.G. (2017) SYRIAN ARABIAN HORSES OFFICE REPORT In: *WAHO Conference* Office of Arabian Horses in Syria, Bahrain

Albrechtsen A., Nielsen F.C. & Nielsen R. (2010) Ascertainment biases in SNP chips affect measures of population divergence. *Molecular biology and evolution*, msq148.

Almarzook S., Reissmann M., Arends D. & Brockmann G. (2017a) Genetic diversity of Syrian Arabian horses. *Animal genetics* **48**, 486–9.

Almarzook S., Reissmann M. & Brockmann G. (2017b) Diversity of mitochondrial DNA in three Arabian horse strains. *Journal of applied genetics* **58**, 273–6.

Altaba A.R., Prezioso V., Darnell J. & Jessell T. (1993) Sequential expression of HNF-3β and HNF-3α by embryonic organizing centers: the dorsal lip/node, notochord and floor plate. *Mechanisms of development* **44**, 91-108.

Anderson C.A., Pettersson F.H., Clarke G.M., Cardon L.R., Morris A.P. & Zondervan K.T. (2010) Data quality control in genetic case-control association studies. *nature protocols* **5**, 1564-73.

Andreu A., Tanji K., Bruno C., Hadjigeorgiou G., Sue C., Jay C., Ohnishi T., Shanske S., Bonilla E. & DiMauro S. (1999) Exercise intolerance due to a nonsense mutation in the mtDNA ND4 gene. *Annals of neurology* **45**, 820-3.

Anthony D.W. (2013) Horses, ancient Near East and Pharaonic Egypt. *The Encyclopedia of Ancient History*.

Bakhtiari J. & Heshmat G. (2009) Estimation of genetic parameters of conformation traits in Iranian Thoroughbred horses. *Livestock Science* **123**, 116-20.

Bandelt H.-J., Forster P., Sykes B.C. & Richards M.B. (1995) Mitochondrial portraits of human populations using median networks. *Genetics* **141**, 743-53.

Baron E., Lopes M., Mendonça D. & da Câmara Machado A. (2012) SNP identification and polymorphism analysis in exon 2 of the horse myostatin gene. *Animal genetics* **43**, 229-32.

Barrey E. (2010) Reviewe: Genetics and genomics in equine exercise physiology: an overview of the new applications of molecular biology as positive and negative markers of performance and health. *Equine Veterinary Journal* **42**, 561-8.

Barrey E., Mucher E., Jeansoule N., Larcher T., Guigand L., Herszberg B., Chaffaux S., Guérin G., Mata X. & Benech P. (2009) Gene expression profiling in equine polysaccharide storage myopathy revealed inflammation, glycogenesis inhibition, hypoxia and mitochondrial dysfunctions. *BMC veterinary research* **5**, 29.

Benjamini Y. & Hochberg Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300.

Bennett D. (2008) Introduction - Part 2: The Origin and Relationships of the Mustang, Barb, and Arabian Horse. In: *The Horse of the Americas Registry*. The Spanish Mustang.

Berber N., Gaouar S., Leroy G., Kdidi S., Tabet Aouel N. & Saïdi Mehtar N. (2014) Molecular characterization and differentiation of five horse breeds raised in Algeria using polymorphic microsatellite markers. *Journal of Animal Breeding and Genetics* **131**, 387-94.

Bergero D., Assenza A. & Caola G. (2005) Contribution to our knowledge of the physiology and metabolism of endurance horses. *Livestock Production Science* **92**, 167-76.

Bernardo R. (2008) Molecular markers and selection for complex traits in plants: learning from the last 20 years. *Crop Science* **48**, 1649-64.

Betts M.J. & Russell R.B. (2003) Amino acid properties and consequences of substitutions. *Bioinformatics for geneticists* **317**, 289.

Bigi D., Perrotta G. & Zambonelli P. (2014) Genetic analysis of seven Italian horse breeds based on mitochondrial DNA D-loop variation. *Animal genetics* **45**, 593-5.

Binns M., Boehler D. & Lambert D. (2010) Identification of the myostatin locus (MSTN) as having a major effect on optimum racing distance in the Thoroughbred horse in the USA. *Animal genetics* **41**, 154-8.

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

Blakely E.L., Mitchell A.L., Fisher N., Meunier B., Nijtmans L.G., Schaefer A.M., Jackson M.J., Turnbull D.M. & Taylor R.W. (2005) A mitochondrial cytochrome b mutation causing severe respiratory chain enzyme deficiency in humans and yeast. *FEBS journal* **272**, 3583-92.

Bodó I., Alderson L. & Langlois B. (2005) *Conservation genetics of endangered horse breeds*. Wageningen Academic Pub.

Bokonyi S. (1987) History of horse domestication. *Animal Genetic Resources Information* **6**, 29-34.

Bottinelli R. & Reggiani C. (2007) *Skeletal muscle plasticity in health and disease: from genes to whole muscle*. Springer.

Bowling A., Del Valle A. & Bowling M. (2000) A pedigree-based study of mitochondrial d-loop DNA sequence variation among Arabian horses. *Animal genetics* **31**, 1-7.

Brard S. & Ricard A. (2015) Genome-wide association study for jumping performances in French sport horses. *Animal genetics* **46**, 78-81.

Bray M., Hagberg J., Perusse L., Rankinen T., Roth S., Wolfarth B. & Bouchard C. (2009) The human gene map for performance and health-related fitness phenotypes: the 2006-2007 update. *Medicine+ Science in Sports+ Exercise* **41**, 35.

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

Brooks S., Makvandi-Nejad S., Chu E., Allen J., Streeter C., Gu E., McCleery B., Murphy B., Bellone R. & Sutter N. (2010a) Morphological variation in the horse: defining complex traits of body size and shape. *Animal genetics* **41**, 159-65.

Brooks S.A., Gabreski N., Miller D., Brisbin A., Brown H.E., Streeter C., Mezey J., Cook D. & Antczak D.F. (2010b) Whole-genome SNP association in the horse: identification of a deletion in myosin Va responsible for Lavender Foal Syndrome. *PLoS Genet* **6**, e1000909.

Cardinali I., Lancioni H., Giontella A., Capodiferro M.R., Capomaccio S., Buttazzoni L., Biggio G.P., Cherchi R., Albertini E., Olivieri A., Cappelli K., Achilli A. & Silvestrelli M. (2016) An Overview of Ten Italian Horse Breeds through Mitochondrial DNA. *PLoS One* **11**, e0153004.

Carnac G., Ricaud S., Vernus B. & Bonnieu A. (2006) Myostatin: biology and clinical relevance. *Mini reviews in medicinal chemistry* **6**, 765-70.

Castejon F., Trigo P., Muñoz A. & Riber C. (2006) Uric acid responses to endurance racing and relationships with performance, plasma biochemistry and metabolic alterations. *Equine Veterinary Journal* **38**, 70-3.

Cervantes I., Baumung R., Molina A., Druml T., Gutiérrez J., Sölkner J. & Valera M. (2009) Size and shape analysis of morphofunctional traits in the Spanish Arab horse. *Livestock Science* **125**, 43-9.

Charlesworth D. & Willis J.H. (2009) The genetics of inbreeding depression. *Nature reviews genetics* **10**, 783-96.

Chauhan M., Gupta A.K. & Dhillon S. (2011) Genetic diversity and population structure of three Indian horse breeds. *Mol Biol Rep* **38**, 3505-11.

Chen W.-M. & Abecasis G.R. (2007) Family-based association tests for genomewide association scans. *The American Journal of Human Genetics* **81**, 913-26.

Cho G. (2006) Genetic relationship among the Korean native and alien horses estimated by microsatellite polymorphism. *ASIAN AUSTRALASIAN JOURNAL OF ANIMAL SCIENCES* **19**, 784.

Cho H.-W., Shin S., Song K.-D., Park J.-W., Choi J.-Y., Lee H.-K. & Cho B.-W. (2015) Molecular characterization and expression analysis of adrenergic receptor beta 2 (ADRB2) gene before and after exercise in the horse. *Asian-Australasian Journal of Animal Sciences (AJAS)* **28**, 686-90.

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

Çilek S. (2012) Heritability parameters for some body measurements in Turkish Arabian foals. *Iranian Journal of Veterinary Research* **13**, 323-9.

Clarkson P.M., Devaney J.M., Gordish-Dressman H., Thompson P.D., Hubal M.J., Urso M., Price T.B., Angelopoulos T.J., Gordon P.M. & Moyna N.M. (2005) ACTN3 genotype

is associated with increases in muscle strength in response to resistance training in women. *Journal of Applied Physiology* **99**, 154-63.

Collins A., Lonjou C. & Morton N. (1999) Genetic epidemiology of single-nucleotide polymorphisms. *Proceedings of the National Academy of Sciences* **96**, 15173-7.

Coogle L., Bailey E., Reid R. & Russ M. (1996) Equine dinucleotide repeat polymorphisms at loci LEX002,-003,-004,-005,-007,-008,-009,-010,-011,-013 and-014. *Animal genetics* **27**, 126-7.

Cooper T., Wiggans G. & VanRaden P. (2013) Short communication: Relationship of call rate and accuracy of single nucleotide polymorphism genotypes in dairy cattle. *Journal of dairy science* **96**, 3336-9.

Cottin F., Metayer N., Goachet A., Julliand V., Slawinski J., Billat V. & Barrey E. (2010) Oxygen consumption and gait variables of Arabian endurance horses measured during a field exercise test. *Equine Veterinary Journal* **42**, 1-5.

Crawford M.H. (2007) *Anthropological genetics: theory, methods and applications*. Cambridge University Press.

Creixell P., Schoof E.M., Tan C.S.H. & Linding R. (2012) Mutational properties of amino acid residues: implications for evolvability of phosphorylatable residues. *Phil. Trans. R. Soc. B* **367**, 2584-93.

Cucchi T., Mohaseb A., Peigné S., Debue K., Orlando L. & Mashkour M. (2017) Detecting taxonomic and phylogenetic signals in equid cheek teeth: towards new palaeontological and archaeological proxies. *Royal Society Open Science* **4**, 160997.

DAD_IS (2011) DAD-IS, Food and Agriculture Organization of the United Nations. Accessed September 2011.

Dall'Olio S., Fontanesi L., Nanni Costa L., Tassinari M., Minieri L. & Falaschini A. (2010) Analysis of horse myostatin gene and identification of single nucleotide polymorphisms in breeds of different morphological types. *BioMed Research International* **2010**.

Dall'Olio S., Fontanesi L., Antonelli C., Nanni Costa L., Tassinari M. & Falaschini A. (2012) Association study between a SNP of the myostatin gene promoter and morphological traits in Uruguayan Creole horse. *Proc. Soc. Ital. Sci. Vet* **66**, 412-4.

Dall'Olio S., Scotti E., Fontanesi L. & Tassinari M. (2014) Analysis of the 227 bp short interspersed nuclear element (SINE) insertion of the promoter of the myostatin (MSTN) gene in different horse breeds. *Veterinaria italiana* **50**, 193-7.

Dario C., Carnicella D., Dario M. & Bufano G. (2006) Morphological evolution and heritability estimates for some biometric traits in the Murgese horse breed. *Genetics and Molecular Research* **5**, 309-14.

Davie A., Qi A. & Zhou S. (2011) Mitogenomics of thoroughbred horses.

De Coo I., Renier W., Ruitenbeek W., Ter Laak H., Bakker M., Schägger H., Van Oost B. & Smeets H. (1999) A 4–base pair deletion in the mitochondrial cytochrome b gene associated with parkinsonism/MELAS overlap syndrome. *Annals of neurology* **45**, 130-3.

de Vicente M.C., Guzman F.A., Engels J. & Rao V.R. (2006) Genetic characterization and its use in decision-making for the conservation of crop germplasm. *The role of biotechnology in exploring and protecting agricultural genetic resources*, 129.

Devi K.M. & Ghosh S.K. (2013) Molecular phylogeny of Indian horse breeds with special reference to Manipuri pony based on mitochondrial D-loop. *Mol Biol Rep* **40**, 5861-7.

Dionne F.T., Turcotte L., Thibault M.-c., Boulay M.R., Skinner J.S. & Bouchard C. (1991) Mitochondrial DNA sequence polymorphism, VO2max, and response to endurance training. *Medicine and science in sports and exercise* **23**, 177-85.

Djinovic-Carugo K., Gautel M., Ylänne J. & Young P. (2002) The spectrin repeat: a structural platform for cytoskeletal protein assemblies. *FEBS letters* **513**, 119-23.

Ehrlich P.R. (1988) *Biodiversity, Chapter 2:The Loss of Diversity Causes and Consequences*. National Academies Press (US), Washington (DC).

Eivers S., McGivney B., Gu J., MacHugh D., Katz L. & Hill E. (2012) PGC-1α encoded by the PPARGC1A gene regulates oxidative energy metabolism in equine skeletal muscle during exercise. *Animal genetics* **43**, 153-62.

Ellegren H. (2004) Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics* **5**, 435-45.

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

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

Excoffier L. & Lischer H.E. (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular ecology resources* **10**, 564-7.

Eynon N., Meckel Y., Alves A.J., Yamin C., Sagiv M., Goldhammer E. & Sagiv M. (2009) Is there an interaction between PPARD T294C and PPARGC1A Gly482Ser polymorphisms and human endurance performance? *Experimental physiology* **94**, 1147-52.

Eynon N., Morán M., Birk R. & Lucia A. (2011) The champions' mitochondria: is it genetically determined? A review on mitochondrial DNA and elite athletic performance. *Physiological genomics* **43**, 789-98.

Fanelli H. (2005) Coat colour dilution lethal ('lavender foal a'): syndrome tetany syndrome of Arabian foals. *Equine Veterinary Education* **17**, 260-3.

FAO (2011) MOLECULAR GENETIC CHARACTERIZATION OF ANIMAL GENETIC RESOURCES. In: *FAO Animal Production and Health Guideline*. COMMISSION ON GENETIC RESOURCES FOR FOOD AND AGRICULTURE FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS, Rome.

FAO (2012a) PHENOTYPIC CHARACTERIZATION OF ANIMAL GENETIC RESOURCES. FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS, Rome.

FAO (2012b) STATUS AND TRENDS OF ANIMAL GENETIC RESOURCES. p. FAO. Commission on Genetic Resources for Food and Agriculture.

Fernández M.E., Goszczynski D.E., Lirón J.P., Villegas-Castagnasso E.E., Carino M.H., Ripoli M.V., Rogberg-Muñoz A., Posik D.M., Peral-García P. & Giovambattista G. (2013) Comparison of the effectiveness of microsatellites and SNP panels for genetic identification, traceability and assessment of parentage in an inbred Angus herd. *Genetics and molecular biology* **36**, 185-91.

Forstén A. (1991) Mitochondrial-DNA time-table and the evolution of Equus: comparison of molecular and paleontological evidence. In: *Annales Zoologici Fennici*, pp. 301-9. JSTOR.

Fox-Clipsham L.Y., Carter S.D., Goodhead I., Hall N., Knottenbelt D.C., May P.D., Ollier W.E. & Swinburne J.E. (2011) Identification of a mutation associated with fatal Foal Immunodeficiency Syndrome in the Fell and Dales pony. *PLoS Genet* **7**, e1002133.

François L., Fegraeus K.J., Eriksson S., Andersson L.S., Tesfayonas Y.G., Viluma A., Imsland F., Buys N., Mikko S. & Lindgren G. (2016) Conformation traits and gaits in the Icelandic horse are associated with genetic variants in Myostatin (MSTN). *Journal of Heredity*, esw031.

Frankham R., Briscoe D.A. & Ballou J.D. (2002) *Introduction to conservation genetics*. Cambridge University Press.

Furuyama T., Kitayama K., Yamashita H. & Nozomu M. (2003) Forkhead transcription factor FOXO1 (FKHR)-dependent induction of PDK4 gene expression in skeletal muscle during energy deprivation. *Biochemical Journal* **375**, 365-71.

Gärke C., Ytournel F., Bed'hom B., Gut I., Lathrop M., Weigend S. & Simianer H. (2012) Comparison of SNPs and microsatellites for assessing the genetic structure of chicken populations. *Animal genetics* **43**, 419-28.

GeneCards, www.genecards.org.

Georgescu S.E., Condac E., Rebedea M., DumitruTesio C., Dinischiotu A. & Costache M. (2005) Arabian horses genotyping using seventeen microsatellites. *Archiva Zootechnica* **8**, 169-75.

Gharahveysi S., Kashan N., Gerami A. & Torshizi R.V. (2008) Estimation of genetic parameters on conformation traits of the Iranian Arab horses population. *Pakistan journal of biological sciences: PJBS* **11**, 280-4.

Gharahveysi S., Manafiazar G. & Merhabibi S. (2010) Genetic correlation of conformationand race performance traits in Iranian Arab horse population *Global Veterinaria*.

Giacomoni E., Fernández-Stolz G. & Freitas T. (2008) Genetic diversity in the Pantaneiro horse breed assessed using microsatellite DNA markers. *Genet. Mol. Res* **7**, 261-70.

Gibbs R.A. (1990) DNA amplification by the polymerase chain reaction. *Analytical chemistry* **62**, 1202-14.

Głażewska I. (2010) Speculations on the origin of the Arabian horse breed. *Livestock Science* **129**, 49-55.

Głażewska I., Anna W., Barbara G., Renata P. & Jerzy S. (2007) A new view on dam lines in Polish Arabian horses based on mtDNA analysis. *Genet. Sel. Evol* **39**, 609-19.

Goodman H.M. (2010) *Basic medical endocrinology*. Academic Press.

Gorbach D.M., Fan B., Onteru S.K., Zhao X., Du Z.-Q., Garrick D.J., Dekkers J. & Rothschild M.F. (2010) Genome-Wide Association Studies for Important Economic Traits in Domestic Animals Using High Density SNP Genotyping. *Animal Industry Report* **656**, 61.

Gorman G.S., Blakely E.L., Hornig-Do H.-T., Tuppen H.A., Greaves L.C., He L., Baker A., Falkous G., Newman J. & Trenell M.I. (2015) Novel MTND1 mutations cause isolated exercise intolerance, complex I deficiency and increased assembly factor expression. *Clinical Science* **128**, 895-904.

GOV.UK (2014) £10 million Chinese thoroughbred horse deal announced. (ed. by Department for Environment FRAaTRHET). All content is available under the Open Government Licence v3.0, UK.

Griffiths-Jones S., Grocock R.J., Van Dongen S., Bateman A. & Enright A.J. (2006) miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic acids research* **34**, D140-D4.

Griffiths-Jones S., Saini H.K., van Dongen S. & Enright A.J. (2007) miRBase: tools for microRNA genomics. *Nucleic acids research* **36**, D154-D8.

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

Gu J., MacHugh D., McGivney B., Park S., Katz L. & Hill E. (2010) Association of sequence variants in CKM (creatine kinase, muscle) and COX4I2 (cytochrome c oxidase, subunit 4, isoform 2) genes with racing performance in Thoroughbred horses. *Equine veterinary journal* **42**, 569-75.

Gu J., Orr N., Park S.D., Katz L.M., Sulimova G., MacHugh D.E. & Hill E.W. (2009) A genome scan for positive selection in thoroughbred horses. *PLoS One* **4**, e5767.

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

Guo Y., He J., Zhao S., Wu H., Zhong X., Sheng Q., Samuels D.C., Shyr Y. & Long J. (2014) Illumina human exome genotyping array clustering and quality control. *nature protocols* **9**, 2643-62.

Haeringen H.v., Bowling A., Stott M., Lenstra J. & Zwaagstra K. (1994) A highly polymorphic horse microsatellite locus: VHL20. *Animal genetics* **25**, 207-.

Hall T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. In: *Nucleic acids symposium series*, pp. 95-8.

Hanna M. & Nelson I. (1999) Genetics and molecular pathogenesis of mitochondrial respiratory chain diseases. *Cellular and Molecular Life Sciences CMLS* **55**, 691-706.

Harrison R.G. (1989) Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. *Trends in Ecology & Evolution* **4**, 6-11.

Harrison S.P. & Turrion-Gomez J.L. (2006) Mitochondrial DNA: an important female contribution to thoroughbred racehorse performance. *Mitochondrion* **6**, 53-66.

Hartl D.L., Clark A.G. & Clark A.G. (1997) *Principles of population genetics*. Sinauer associates Sunderland.

Hasler H., Flury C., Menet S., Haase B., Leeb T., Simianer H., Poncet P. & Rieder S. (2011) Genetic diversity in an indigenous horse breed–implications for mating strategies and the control of future inbreeding. *Journal of Animal Breeding and Genetics* **128**, 394-406.

Hatefi Y. (1985) The mitochondrial electron transport and oxidative phosphorylation system. *Annual review of biochemistry* **54**, 1015-69.

Heaton M.P., Harhay G.P., Bennett G.L., Stone R.T., Grosse W.M., Casas E., Keele J.W., Smith T.P., Chitko-McKown C.G. & Laegreid W.W. (2002) Selection and use of SNP markers for animal identification and paternity analysis in US beef cattle. *Mammalian Genome* **13**, 272-81.

Hill E., Bradley D., Al-Barody M., Ertugrul O., Splan R., Zakharov I. & Cunningham E. (2002) History and integrity of thoroughbred dam lines revealed in equine mtDNA variation. *Animal genetics* **33**, 287-94.

Hill E., Gu J., McGivney B. & MacHugh D. (2010a) Targets of selection in the Thoroughbred genome contain exercise-relevant gene SNPs associated with elite racecourse performance. *Animal genetics* **41**, 56-63.

Hill E., W, Ryan D., P & MacHugh D., E (2012a) Horses for courses: a DNA-based test for race distance aptitude in thoroughbred racehorses. *Recent patents on DNA & gene sequences* **6**, 203-8.

Hill E.W., Fonseca R.G., McGivney B.A., Gu J., MacHugh D.E. & Katz L.M. (2012b) MSTN genotype (g. 66493737C/T) association with speed indices in Thoroughbred racehorses. *Journal of Applied Physiology* **112**, 86-90.

Hill E.W., Gu J., Eivers S.S., Fonseca R.G., McGivney B.A., Govindarajan P., Orr N., Katz L.M. & MacHugh D. (2010b) A sequence polymorphism in MSTN predicts sprinting ability and racing stamina in thoroughbred horses. *PLoS One* **5**, e8645.

Hill E.W., McGivney B.A., Gu J., Whiston R. & MacHugh D.E. (2010c) A genome-wide SNP-association study confirms a sequence variant (g. 66493737C> T) in the equine myostatin (MSTN) gene as the most powerful predictor of optimum racing distance for Thoroughbred racehorses. *BMC genomics* **11**, 1.

Hill W.G. & Rasbash J. (1986) Models of long term artificial selection in finite population. *Genetics Research* **48**, 41-50.

Hirschhorn J.N. & Daly M.J. (2005) Genome-wide association studies for common diseases and complex traits. *Nature reviews genetics* **6**, 95-108.

Hoggatt A.M., Kriegel A.M., Smith A.F. & Herring B.P. (2000) Hepatocyte nuclear factor-3 homologue 1 (HFH-1) represses transcription of smooth muscle-specific genes. *Journal of Biological Chemistry* **275**, 31162-70.

Hong H.-K., Lass J.H. & Chakravarti A. (1999) Pleiotropic skeletal and ocular phenotypes of the mouse mutation congenital hydrocephalus (ch/Mf1) arise from a winged helix/forkhead transcription factor gene. *Human molecular genetics* **8**, 625-37.

Hood D.A. (2009) Mechanisms of exercise-induced mitochondrial biogenesis in skeletal muscle This paper is one of a selection of papers published in this Special Issue, entitled 14th International Biochemistry of Exercise Conference–Muscles as Molecular and Metabolic Machines, and has undergone the Journal's usual peer review process. *Applied Physiology, Nutrition, and Metabolism* **34**, 465-72.

Horvath R., Schoser B., Müller-Höcker J., Völpel M., Jaksch M. & Lochmüller H. (2005) Mutations in mtDNA-encoded cytochrome c oxidase subunit genes causing isolated myopathy or severe encephalomyopathy. *Neuromuscular Disorders* **15**, 851-7.

Hrouda B. & Bottéro J. (1991) *Der alte Orient: Geschichte und Kultur des alten Vorderasien*. Bertelsmann, Gütersloh.

Hu J., Gui S., Zhu Z., Wang X., Ke W. & Ding Y. (2015) Genome-wide identification of SSR and SNP markers based on whole-genome re-sequencing of a Thailand wild sacred lotus (Nelumbo nucifera). *PLoS One* **10**, e0143765.

Huebner A.K., Schinke T., Priemel M., Schilling S., Schilling A.F., Emeson R.B., Rueger J.M. & Amling M. (2006) Calcitonin deficiency in mice progressively results in high bone turnover. *Journal of Bone and Mineral Research* **21**, 1924-34.

Hurst C.C. (1906) On the inheritance of coat colour in horses. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character* **77**, 388-94.

ISAG (2014) Equine Genetics and Thoroughbred Parentage Testing Workshop. In: *ISAG* (ed. by Committee IS). International Society for Animal Genetics, Xian, China.

Ishida N., Oyunsuren T., Mashima S., Mukoyama H. & Saitou N. (1995) Mitochondrial DNA sequences of various species of the genus Equus with special reference to the phylogenetic relationship between Przewalskii's wild horse and domestic horse. *Journal of molecular evolution* **41**, 180-8.

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

Jónsson H., Schubert M., Seguin-Orlando A., Ginolhac A., Petersen L., Fumagalli M., Albrechtsen A., Petersen B., Korneliussen T.S. & Vilstrup J.T. (2014) Speciation with gene flow in equids despite extensive chromosomal plasticity. *Proceedings of the National Academy of Sciences* **111**, 18655-60.

Joyner M.J. & Coyle E.F. (2008) Endurance exercise performance: the physiology of champions. *The Journal of physiology* **586**, 35-44.

Kader A., Li Y., Dong K., Irwin D.M., Zhao Q., He X., Liu J., Pu Y., Gorkhali N.A. & Liu X. (2016) Population Variation Reveals Independent Selection toward Small Body Size in Chinese Debao Pony. *Genome biology and evolution* **8**, 42-50.

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* **45**, 375-95.

Kavar T., Brem G., Habe F., Solkner J. & Dovc P. (2002) History of Lipizzan horse maternal lines as revealed by mtDNA analysis. *Genetics Selection Evolution* **34**, 635.

Keightley J.A., Anitori R., Burton M.D., Quan F., Buist N.R. & Kennaway N.G. (2000) Mitochondrial encephalomyopathy and complex III deficiency associated with a stop-codon mutation in the cytochrome b gene. *The American Journal of Human Genetics* **67**, 1400-10.

Keyser-Tracqui C., Blandin-Frappin P., Francfort H.P., Ricaut F.X., Lepetz S., Crubezy E., Samashev Z. & Ludes B. (2005) Mitochondrial DNA analysis of horses recovered from a frozen tomb (Berel site, Kazakhstan, 3rd Century BC). *Anim Genet* **36**, 203-9.

Khadka R. (2015) Global horse population with respect to breeds and risk status.

Khanshour A., Conant E., Juras R. & Cothran E.G. (2013) Microsatellite analysis of genetic diversity and population structure of Arabian horse populations. *J Hered* **104**, 386-98.

Khanshour A.M. & Cothran E.G. (2013) Maternal phylogenetic relationships and genetic variation among Arabian horse populations using whole mitochondrial DNA D-loop sequencing. *BMC genetics* **14**, 83.

Korte A. & Farlow A. (2013) The advantages and limitations of trait analysis with GWAS: a review. *Plant methods* **9**, 29.

Koskinen M.T., Hirvonen H., LANDRY P.A. & Primmer C.R. (2004) The benefits of increasing the number of microsatellites utilized in genetic population studies: an empirical perspective. *Hereditas* **141**, 61-7.

Kreuzer S., Reissmann M. & Brockmann G.A. (2013) Gene test to elucidate the ETEC F4ab/F4ac receptor status in pigs. *Veterinary microbiology* **162**, 293.

Kühlbrandt W. (2015) Structure and function of mitochondrial membrane protein complexes. *BMC biology* **13**, 1.

Kulbrock M., Lehner S., Metzger J., Ohnesorge B. & Distl O. (2013) A genome-wide association study identifies risk loci to equine recurrent uveitis in German warmblood horses. *PLoS One* **8**, e71619.

Larson G. & Fuller D.Q. (2014) The evolution of animal domestication. *Annual Review of Ecology, Evolution, and Systematics* **45**, 115-36.

Laurie C.C., Doheny K.F., Mirel D.B., Pugh E.W., Bierut L.J., Bhangale T., Boehm F., Caporaso N.E., Cornelis M.C. & Edenberg H.J. (2010) Quality control and quality assurance in genotypic data for genome-wide association studies. *Genetic epidemiology* **34**, 591-602.

Laval G., SanCristobal M. & Chevalet C. (2002) Measuring genetic distances between breeds: use of some distances in various short term evolution models. *Genetics Selection Evolution* **34**, 1.

Lenstra J., Groeneveld L., Eding H., Kantanen J., Williams J., Taberlet P., Nicolazzi E., Sölkner J., Simianer H. & Ciani E. (2012) Molecular tools and analytical approaches for the characterization of farm animal genetic diversity. *Animal genetics* **43**, 483-502.

Leroy G., Callède L., Verrier E., Mériaux J.-C., Ricard A., Danchin-Burge C. & Rognon X. (2009) Genetic diversity of a large set of horse breeds raised in France assessed by microsatellite polymorphism. *Genetics Selection Evolution* **41**, 1.

Leutenegger A.-L., Prum B., Génin E., Verny C., Lemainque A., Clerget-Darpoux F. & Thompson E.A. (2003) Estimation of the inbreeding coefficient through use of genomic data. *The American Journal of Human Genetics* **73**, 516-23.

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

Ling Y., Ma Y., Guan W., Cheng Y., Wang Y., Han J., Mang L., Zhao Q., He X. & Pu Y. (2011) Evaluation of the genetic diversity and population structure of Chinese indigenous horse breeds using 27 microsatellite markers. *Animal genetics* **42**, 56-65.

Lips P. & van Schoor N. (2011) The effect of vitamin D on bone and osteoporosis. *Best practice & research Clinical endocrinology & metabolism* **25**, 585-91.

Lopes M., Mendonça D., Cymbron T., Valera M., Costa-Ferreira D. & da Câmara Machado A. (2005) The Lusitano horse maternal lineage based on mitochondrial D-loop sequence variation. *Animal genetics* **36**, 196-202.

Lucia A., Gómez-Gallego F., Barroso I., Rabadán M., Bandrés F., San Juan A.F., Chicharro J.L., Ekelund U., Brage S. & Earnest C.P. (2005) PPARGC1A genotype (Gly482Ser) predicts exceptional endurance capacity in European men. *Journal of Applied Physiology* **99**, 344-8.

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

MacArthur D.G. & North K.N. (2004) A gene for speed? The evolution and function of α-actinin-3. *Bioessays* **26**, 786-95.

MacArthur D.G., Seto J.T., Chan S., Quinlan K.G., Raftery J.M., Turner N., Nicholson M.D., Kee A.J., Hardeman E.C. & Gunning P.W. (2008) An Actn3 knockout mouse provides mechanistic insights into the association between α-actinin-3 deficiency and human athletic performance. *Human molecular genetics* **17**, 1076-86.

Mackowski M., Mucha S., Cholewinski G. & Cieslak J. (2015) Genetic diversity in Hucul and Polish primitive horse breeds. *Archiv fuer Tierzucht* **58**, 23.

Maffetone P. (2010) *The big book of endurance training and racing*. Skyhorse Publishing, Inc.

Mahrous K., Hassanane M., Abdel Mordy M., Shafey H. & Hassan N. (2011) Genetic variations in horse using microsatellite markers. *Journal of Genetic Engineering and Biotechnology* **9**, 103-9.

Makvandi-Nejad S., Hoffman G.E., Allen J.J., Chu E., Gu E., Chandler A.M., Loredo A.I., Bellone R.R., Mezey J.G. & Brooks S.A. (2012) Four loci explain 83% of size variation in the horse. *PLoS One* **7**, e39929.

Markert J.A., Champlin D.M., Gutjahr-Gobell R., Grear J.S., Kuhn A., McGreevy T.J., Roth A., Bagley M.J. & Nacci D.E. (2010) Population genetic diversity and fitness in multiple environments. *BMC evolutionary biology* **10**, 1.

Marklund L., Moller M.J., Sandberg K. & Andersson L. (1996) A missense mutation in the gene for melanocyte-stimulating hormone receptor (MCIR) is associated with the chestnut coat color in horses. *Mammalian genome* **7**, 895-9.

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

Martin W. & Mentel M. (2010) The origin of mitochondria. *Nature Education* **3**, 58.

Mata X., Vaiman A., Ducasse A., Diribarne M., Schibler L. & Guerin G. (2012) Genomic structure, polymorphism and expression of the horse alpha-actinin-3 gene. *Gene* **491**, 20-4.

Maudet C., Luikart G. & Taberlet P. (2002) Genetic diversity and assignment tests among seven French cattle breeds based on microsatellite DNA analysis. *Journal of Animal Science* **80**, 942-50.

Mayouf N.K., Jadaan B. & Alfares A. (2011) *Arabian Horse Adjudication, between originality and beauty.* . Orient printing and publishing.

McCue M.E., Bannasch D.L., Petersen J.L., Gurr J., Bailey E., Binns M.M., Distl O., Guerin G., Hasegawa T., Hill E.W., Leeb T., Lindgren G., Penedo M.C., Roed K.H., Ryder O.A., Swinburne J.E., Tozaki T., Valberg S.J., Vaudin M., Lindblad-Toh K., Wade C.M. & Mickelson J.R. (2012) A high density SNP array for the domestic horse and extant Perissodactyla: utility for association mapping, genetic diversity, and phylogeny studies. *PLoS Genet* **8**, e1002451.

Medugorac I., Veit-Kensch C.E., Ramljak J., Brka M., Marković B., Stojanović S., Bytyqi H., Kochoski L., Kume K. & Grünenfelder H.P. (2011) Conservation priorities of genetic diversity in domesticated metapopulations: a study in taurine cattle breeds. *Ecology and evolution* **1**, 408-20.

Meira C.T., Farah M.M., Fortes M.R., Moore S.S., Pereira G.L., Silva J.A.I.V., da Mota M.D.S. & Curi R.A. (2014) A Genome-Wide Association Study for Morphometric Traits in Quarter Horse. *Journal of Equine Veterinary Science* **34**, 1028-31.

Meleleo D. & Picciarelli V. (2016) Effect of calcium ions on human calcitonin. Possible implications for bone resorption by osteoclasts. *BioMetals* **29**, 61-79.

Metayer N., Biau S., Cochet J. & Barrey E. (2004) Stduy of locomotion and morphological factors in the performance of the horse specialized in endurance tests. In: *30ème journée de la recherche équine, 3 mars 2004.*, pp. 67-76. Les Haras Nationaux Direction du Développement.

Metzger J., Karwath M., Tonda R., Beltran S., Águeda L., Gut M., Gut I.G. & Distl O. (2015) Runs of homozygosity reveal signatures of positive selection for reproduction traits in breed and non-breed horses. *BMC genomics* **16**, 764.

Metzger J., Ohnesorge B. & Distl O. (2012) Genome-wide linkage and association analysis identifies major gene loci for guttural pouch tympany in Arabian and German warmblood horses. *PLoS One* **7**, e41640.

Miller S., Dykes D. & Polesky H. (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic acids res* **16**, 1215.

Molina A., Valera M., Dos Santos R. & Rodero A. (1999) Genetic parameters of morphofunctional traits in Andalusian horse. *Livestock Production Science* **60**, 295-303.

Moritz C., Dowling T. & Brown W. (1987) Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annual review of ecology and systematics*, 269-92.

Mosher D.S., Quignon P., Bustamante C.D., Sutter N.B., Mellersh C.S., Parker H.G. & Ostrander E.A. (2007) A mutation in the myostatin gene increases muscle mass and enhances racing performance in heterozygote dogs. *PLoS genetics* **3**, e79.

Moshkelani S., Rabiee S. & Javaheri-Koupaei M. (2011) DNA fingerprinting of Iranian Arab horse using fourteen microsatellites marker. *Res J Biol Sci* **6**, 402-5.

Mostafa F., Mohammad J.K., Abbas D. & Saadat M. (2011) Identity testing with fourteen genetic markers of Arabian horses in Charmahal-va-bakhtiari province. *J Hered* **88**, 335-42.

Nagylaki T. (1998) Fixation indices in subdivided populations. *Genetics* **148**, 1325-32.

Nei M. (1972) Genetic distance between populations. *American naturalist*, 283-92.

Nei M. (1973) Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences* **70**, 3321-3.

Nei M. (1977) F-statistics and analysis of gene diversity in subdivided populations. *Annals of human genetics* **41**, 225-33.

Nei M. (1987) Genetic distance and molecular phylogeny. *Population Genetics and Fishery Management. Seattle. USA*, 193-223.

Nei M. & Kumar S. (2000) *Molecular evolution and phylogenetics*. Oxford university press.

Nei M. & Tajima F. (1981) DNA polymorphism detectable by restriction endonucleases. *Genetics* **97**, 145-63.

Nei M., Tajima F. & Tateno Y. (1983) Accuracy of estimated phylogenetic trees from molecular data. *Journal of molecular evolution* **19**, 153-70.

Niemi A.-K. (2005) *Mitochondrial DNA Variation in Extremely Selected Traits: Longevity and Elite Athletic Performance*. University of Oulu.

Niemi A.-K. & Majamaa K. (2005) Mitochondrial DNA and ACTN3 genotypes in Finnish elite endurance and sprint athletes. *European Journal of Human Genetics* **13**, 965-9.

Ning T., Xiao H., Li J., Hua S. & Zhang Y. (2010) Adaptive evolution of the mitochondrial ND6 gene in the domestic horse. *Genet Mol Res* **9**, 144-50.

Ogg S., Paradis S., Gottlieb S., Patterson G.I., Lee L., Tissenbaum H.A. & Ruvkun G. (1997) The Fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in C. elegans. *Nature* **389**, 994-9.

Ojango J., Mpofu N., Marshall K. & Andersson-Eklund L. (2011) Quantitative methods to improve the understanding and utilization of animal genetic resources. *Animal Genetics Training Resources. International Livestock Research Institute, Nairobi, Kenya and Swedish University of Agricultural Science, Uppsala, Sweden*.

Orlando L., Ginolhac A., Zhang G., Froese D., Albrechtsen A., Stiller M., Schubert M., Cappellini E., Petersen B. & Moltke I. (2013) Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. *Nature* **499**, 74-8.

Orr N., Back W., Gu J., Leegwater P., Govindarajan P., Conroy J., Ducro B., Van Arendonk J., MacHugh D. & Ennis S. (2010) Genome-wide SNP association–based localization of a dwarfism gene in Friesian dwarf horses. *Animal genetics* **41**, 2-7.

Orysiak J., Busko K., Michalski R., Mazur-Różycka J., Gajewski J., Malczewska-Lenczowska J., Sitkowski D. & Pokrywka A. (2014) Relationship between ACTN3 R577X polymorphism and maximal power output in elite Polish athletes. *Medicina* **50**, 303-8.

Ostrander E.A., Huson H.J. & Ostrander G.K. (2009) Genetics of athletic performance. *Annual review of genomics and human genetics* **10**, 407-29.

Park C.B. & Larsson N.-G. (2011) Mitochondrial DNA mutations in disease and aging. *The Journal of cell biology* **193**, 809-18.

Parry D.A. & Squire J.M. (2005) *Fibrous proteins: Coiled-coils, collagen and elastomers*. Gulf Professional Publishing.

Patnala R., Clements J. & Batra J. (2013) Candidate gene association studies: a comprehensive guide to useful in silico tools. *BMC genetics* **14**, 1.

Peakall R. & Smouse P.E. (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics* **28**, 2537-9.

Pearson K. & Lee A. (1900) Mathematical Contributions to the Theory of Evolution. VIII. On the Inheritance of Characters not Capable of Exact Quantitative Measurement. Part I. Introductory. Part II. On the Inheritance of Coat-Colour in Horses. Part III. On the Inheritance of Eye-Colour in Man. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* **195**, 79-150.

Pearson T.A. & Manolio T.A. (2008) How to interpret a genome-wide association study. *Jama* **299**, 1335-44.

Petersen J.L., Mickelson J.R., Cothran E.G., Andersson L.S., Axelsson J., Bailey E., Bannasch D., Binns M.M., Borges A.S., Brama P., da Camara Machado A., Distl O., Felicetti M., Fox-Clipsham L., Graves K.T., Guerin G., Haase B., Hasegawa T., Hemmann K., Hill E.W., Leeb T., Lindgren G., Lohi H., Lopes M.S., McGivney B.A., Mikko S., Orr N., Penedo M.C., Piercy R.J., Raekallio M., Rieder S., Roed K.H., Silvestrelli M., Swinburne J., Tozaki T., Vaudin M., C M.W. & McCue M.E. (2013a) Genetic diversity in the modern horse illustrated from genome-wide SNP data. *PLoS One* **8**, e54997.

Petersen J.L., Mickelson J.R., Rendahl A.K., Valberg S.J., Andersson L.S., Axelsson J., Bailey E., Bannasch D., Binns M.M., Borges A.S., Brama P., da Camara Machado A., Capomaccio S., Cappelli K., Cothran E.G., Distl O., Fox-Clipsham L., Graves K.T., Guerin G., Haase B., Hasegawa T., Hemmann K., Hill E.W., Leeb T., Lindgren G.,

Lohi H., Lopes M.S., McGivney B.A., Mikko S., Orr N., Penedo M.C., Piercy R.J., Raekallio M., Rieder S., Roed K.H., Swinburne J., Tozaki T., Vaudin M., Wade C.M. & McCue M.E. (2013b) Genome-wide analysis reveals selection for important traits in domestic horse breeds. *PLoS Genet* **9**, e1003211.

Petersen J.L., Valberg S.J., Mickelson J.R. & McCue M.E. (2014) Haplotype diversity in the equine myostatin gene with focus on variants associated with race distance propensity and muscle fiber type proportions. *Animal genetics* **45**, 827-35.

Polasik D., Rogers B., Bobrowska-Chwat A. & Pikuła R. (2017) DETECTION AND ANALYSIS OF POLYMORPHISM IN THE PROMOTER REGION OF EQUINE PPARCG1A GENE. *The Journal of Animal & Plant Sciences* 691-5.

Prado R. & Mota M. (2008) Genetic parameters for biometric traits in Mangalarga horses. *REDVET. Revista Electrónica de Veterinaria* **9**, 1-15.

Pretorius S., Van Marle-Koster E. & Mostert B. (2004) Description of the Friesian horse population of South Africa and Namibia. *South African Journal of Animal Science* **34**, 149-57.

Pritchard J., Wen X. & Falush D. (2010) Documentation for structure software: Version 2.3. *University of Chicago, Chicago, IL*.

Rankinen T., Fuku N., Wolfarth B., Wang G., Sarzynski M.A., Alexeev D.G., Ahmetov I.I., Boulay M.R., Cieszczyk P. & Eynon N. (2016) No evidence of a common DNA variant profile specific to world class endurance athletes. *PLoS One* **11**, e0147330.

Raswan C.R., Wenzler G. & Seydel H. (1981) *Der Araber und sein Pferd*. Georg Olms Verlag, Schickhardt&Ebner (Konrad Wittwer), Stuttgart, Deutschland.

Raudsepp T., Santani A., Wallner B., Kata S.R., Ren C., Zhang H.-B., Womack J.E., Skow L.C. & Chowdhary B.P. (2004) A detailed physical map of the horse Y chromosome. *Proc Natl Acad Sci U S A* **101**, 9321-6.

Reissmann M., Musa L., Zakizadeh S. & Ludwig A. (2016) Distribution of coat-color-associated alleles in the domestic horse population and Przewalski's horse. *Journal of applied genetics*, 1-7.

Reynolds J., Weir B.S. & Cockerham C.C. (1983) Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics* **105**, 767-79.

Ricard A., Danvy S. & Legarra A. (2013) Computation of deregressed proofs for genomic selection when own phenotypes exist with an application in French show-jumping horses. *Journal of Animal Science* **91**, 1076-85.

Ricard A., Robert C., Blouin C., Baste F., Torquet G., Morgenthaler C., Rivière J., Mach N., Mata X. & Schibler L. (2017) Endurance exercise ability in the horse: a trait with complex polygenic determinism. *Frontiers in genetics* **8**, 89.

Ricard A. & Touvais M. (2007) Genetic parameters of performance traits in horse endurance races. *Livestock Science* **110**, 118-25.

Rieder S. (2009) Molecular tests for coat colours in horses. *Journal of Animal Breeding and Genetics* **126**, 415-24.

Rieder S., Taourit S., Mariat D., Langlois B. & Guérin G. (2001) Mutations in the agouti (ASIP), the extension (MC1R), and the brown (TYRP1) loci and their association to coat color phenotypes in horses (Equus caballus). *Mammalian Genome* **12**, 450-5.

Rivera M.A., Wolfarth B., Dionne F.T., Chagnon M., Simoneau J.-A., Boulay M.R., Song T.M., Perusse L., Gagnon J. & Leon A.S. (1998) Three mitochondrial DNA restriction polymorphisms in elite endurance athletes and sedentary controls. *Medicine and science in sports and exercise* **30**, 687-90.

Rivero J.-L.L. & Barrey E. (2001) Heritabilities and genetic and phenotypic parameters for gluteus medius muscle fibre type composition, fibre size and capillaries in purebred Spanish horses. *Livestock Production Science* **72**, 233-41.

Rivero J., Galisteo A., Agüera E. & Miro F. (1993a) Skeletal muscle histochemistry in male and female Andalusian and Arabian horses of different ages. *Research in veterinary science* **54**, 160-9.

Rivero J., Serrano A.L., Henckel P. & Aguera E. (1993b) Muscle fiber type composition and fiber size in successfully and unsuccessfully endurance-raced horses. *Journal of Applied Physiology* **75**, 1758-66.

Rothfels K., Axelrad A., Siminovitch L., McCulloch E. & Parker R. (1959) The origin of altered cell lines from mouse, monkey, and man, as indicated by chromosome and transplantation studies. In: *Can. Cancer Conf*, pp. 189-214.

Rousseau É. (2017) *Horses of the World*. Princeton University Press.

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

Ruane J. (1999) A critical review of the value of genetic distance studies in conservation of animal genetic resources. *Journal of Animal Breeding and Genetics* **116**, 317-23.

Rukavina D., Hasanbašić D., Durmić-Pašić A., Kalamujić B., Zahirović A., Ramić J. & Pojskić N. (2015) Genetic Diversity of Arabian Horse from Stud "Borike" (Bosnia and Herzegovina) Using Microsatellite Markers. *Research & Reviews: Journal of Veterinary Sciences* **1**.

Sadek M. (2006) Estimation of genetic and phenotypic parameters of body measurements and conformation traits in Arabian horses. *Egyptian J. Anim. Prod. ESAP* **43**, 153-62.

Sadek M., Al-Aboud A. & Ashmawy A. (2006) Factor analysis of body measurements in Arabian horses. *J. Anim. Breed. Genet.*

Saitou N. & Nei M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular biology and evolution* **4**, 406-25.

Samoré A., Pagnacco G. & Miglior F. (1997) Genetic parameters and breeding values for linear type traits in the Haflinger horse. *Livestock Production Science* **52**, 105-11.

Samper J.C. (2009) *Equine breeding management and artificial insemination*. Elsevier Health Sciences.

Sanchez A.M., Candau R.B. & Bernardi H. (2014) FoxO transcription factors: their roles in the maintenance of skeletal muscle homeostasis. *Cellular and molecular life sciences* **71**, 1657-71.

Sánchez M.J., Gómez M.D., Peña F., Monterde J.G., Morales J.L., Molina A. & Valera M. (2013) Relationship between conformation traits and gait characteristics in Pura Raza Español horses. *Archiv Tierzucht* **56**, 137-48.

Santagostino M., Khoriauli L., Gamba R., Bonuglia M., Klipstein O., Piras F.M., Vella F., Russo A., Badiale C. & Mazzagatti A. (2015a) Genome-wide evolutionary and functional analysis of the Equine Repetitive Element 1: an insertion in the myostatin promoter affects gene expression. *BMC genetics* **16**, 1.

Santagostino M., Khoriauli L., Gamba R., Bonuglia M., Klipstein O., Piras F.M., Vella F., Russo A., Badiale C. & Mazzagatti A. (2015b) Genome-wide evolutionary and functional analysis of the Equine Repetitive Element 1: an insertion in the myostatin promoter affects gene expression. *BMC genetics* **16**, 126.

Sargious M.A., Bakry H., El-Shawarby R. & Ahmed H.A. (2014) Parentage testing of Arabian horse in Egypt using microsatellite DNA typing.

Sarkar A. & Hochedlinger K. (2013) The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. *Cell stem cell* **12**, 15-30.

Sarpeshkar V. & Bentley D.J. (2010) Adrenergic-β2 receptor polymorphism and athletic performance. *Journal of human genetics* **55**, 479-85.

Schaefer R.J., Schubert M., Bailey E., Bannasch D.L., Barrey E., Bar-Gal G.K., Brem G., Brooks S.A., Distl O. & Fries R. (2017) Development of a high-density, 2M SNP genotyping array and 670k SNP imputation array for the domestic horse. *bioRxiv*, 112979.

Schiettecatte J. & Zouache A. (2017) The Horse in Arabia and the Arabian Horse: Origins, Myths and Realities. *Arabian Humanities. Revue internationale d'archéologie et de sciences sociales sur la péninsule Arabique/International Journal of Archaeology and Social Sciences in the Arabian Peninsula*.

Schinke T., Liese S., Priemel M., Haberland M., Schilling A.F., Catala-Lehnen P., Blicharski D., Rueger J.M., Gagel R.F. & Emeson R.B. (2004) Decreased Bone Formation and Osteopenia in Mice Lacking α-Calcitonin Gene-Related Peptide. *Journal of Bone and Mineral Research* **19**, 2049-56.

Schmitt A.O., Bortfeldt R.H. & Brockmann G.A. (2010) Tracking chromosomal positions of oligomers-a case study with Illumina's BovineSNP50 beadchip. *BMC genomics* **11**, 1.

Schröder W., Klostermann A. & Distl O. (2011) Candidate genes for physical performance in the horse. *The Veterinary Journal* **190**, 39-48.

Schröder W., Klostermann A., Stock K. & Distl O. (2012) A genome-wide association study for quantitative trait loci of show-jumping in Hanoverian warmblood horses. *Animal genetics* **43**, 392-400.

Senju N., Tozaki T., Kakoi H., Shinjo A., Matsuyama R., Almunia J. & Takasu M. (2017) Genetic diversity of the Yonaguni horse based on polymorphisms in microsatellites and mitochondrial DNA. *Journal of Veterinary Medical Science* **79**, 425-31.

Seyedabadi H., Amirinia C., Banabazi M.H. & Emrani H. (2006) Parentage verification of Iranian Caspian horse using microsatellites markers. *Iranian Journal of Biotechnology* **4**, 260-4.

Shokolenko I.N. & Alexeyev M.F. (2015) Mitochondrial DNA: A disposable genome? *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease* **1852**, 1805-9.

Sievers F. & Higgins D.G. (2014) Clustal omega. *Current Protocols in Bioinformatics*, 3.13. 1-3.. 6.

Signer-Hasler H., Flury C., Haase B., Burger D., Simianer H., Leeb T. & Rieder S. (2012) A genome-wide association study reveals loci influencing height and other conformation traits in horses. *PLoS One* **7**, e37282.

Silva M.S., Bolani W., Alves C.R., Biagi D.G., Lemos Jr J.R., da Silva J.L., de Oliveira P.A., Alves G.B., de Oliveira E.M. & Negrão C.E. (2015) Elimination of Influences of the ACTN3 R577X Variant on Oxygen Uptake by Endurance Training in Healthy Individuals. *IJSPP* **10**.

Sjöblom B., Salmazo A. & Djinović-Carugo K. (2008) α-Actinin structure and regulation. *Cellular and molecular life sciences* **65**, 2688-701.

Smits P., Li P., Mandel J., Zhang Z., Deng J.M., Behringer R.R., De Crombrugghe B. & Lefebvre V. (2001) The transcription factors L-Sox5 and Sox6 are essential for cartilage formation. *Developmental cell* **1**, 277-90.

Stachurska A., Pięta M., Łojek J. & Szulowska J. (2007) Performance in racehorses of various colours. *Livestock Science* **106**, 282-6.

Staiger E.A., Al Abri M., Pflug K.M., Kalla S., Ainsworth D., Miller D., Raudsepp T., Sutter N.B. & Brooks S.A. (2016) Skeletal variation in Tennessee Walking Horses maps to the LCORL/NCAPG gene region. *Physiological genomics* **48**, 325-35.

Stefaniuk M., Kaczor U., Augustyn R., Gurgul A., Kulisa M. & Podstawski Z. (2014) Identification of a New Haplotype within the Promoter Region of the MSTN Gene in Horses from Five of the most Common Breeds in Poland. *Folia biologica* **62**, 219-22.

Stefaniuk M., Ropka-Molik K., Piórkowska K., Kulisa M. & Podstawski Z. (2016) Analysis of polymorphisms in the equine MSTN gene in Polish populations of horse breeds. *Livestock Science* **187**, 151-7.

Sziszkosz N., Mihók S., Jávor A. & Kusza S. (2016) Genetic diversity of the Hungarian Gidran horse in two mitochondrial DNA markers. *PeerJ* **4**, e1894.

Tabangin M.E., Woo J.G. & Martin L.J. (2009) The effect of minor allele frequency on the likelihood of obtaining false positives. In: *BMC proceedings*, p. S41. BioMed Central Ltd.

Takezaki N., Nei M. & Tamura K. (2010) POPTREE2: Software for constructing population trees from allele frequency data and computing other population statistics with Windows interface. *Molecular biology and evolution* **27**, 747-52.

Tamura K., Peterson D., Peterson N., Stecher G., Nei M. & Kumar S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution* **28**, 2731-9.

Tanaka M., Takeyasu T., Fuku N., LI-JUN G. & Kurata M. (2004) Mitochondrial genome single nucleotide polymorphisms and their phenotypes in the Japanese. *Annals of the New York Academy of Sciences* **1011**, 7-20.

Tapio M., Ozerov M., Tapio I., Toro M.A., Marzanov N., Ćinkulov M., Goncharenko G., Kiselyova T., Murawski M. & Kantanen J. (2010) Microsatellite-based genetic diversity and population structure of domestic sheep in northern Eurasia. *BMC genetics* **11**, 1.

Tetens J., Widmann P., Kühn C. & Thaller G. (2013) A genome-wide association study indicates LCORL/NCAPG as a candidate locus for withers height in German Warmblood horses. *Animal genetics* **44**, 467-71.

Thomas K.C., Hamilton N.A., North K.N. & Houweling P.J. (2014) Sequence analysis of the equine ACTN3 gene in Australian horse breeds. *Gene* **538**, 88-93.

Toro M. & Caballero A. (2005) Characterization and conservation of genetic diversity in subdivided populations. *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**, 1367-78.

Tozaki T., Hill E., Hirota K., Kakoi H., Gawahara H., Miyake T., Sugita S., Hasegawa T., Ishida N. & Nakano Y. (2012) A cohort study of racing performance in Japanese Thoroughbred racehorses using genome information on ECA18. *Animal genetics* **43**, 42-52.

Tozaki T., Sato F., Hill E.W., Miyake T., Endo Y., Kakoi H., Gawahara H., Hirota K.-i., Nakano Y. & Nambo Y. (2011) Sequence variants at the myostatin gene locus influence the body composition of Thoroughbred horses. *Journal of Veterinary Medical Science* **73**, 1617-24.

Tsianos G.I., Evangelou E., Boot A., Zillikens M.C., van Meurs J.B., Uitterlinden A.G. & Ioannidis J.P. (2010) Associations of polymorphisms of eight muscle-or metabolism-related genes with performance in Mount Olympus marathon runners. *Journal of Applied Physiology* **108**, 567-74.

Turner S., Armstrong L.L., Bradford Y., Carlson C.S., Crawford D.C., Crenshaw A.T., Andrade M., Doheny K.F., Haines J.L. & Hayes G. (2011) Quality control procedures for genome-wide association studies. *Current protocols in human genetics*, 1.19. 1-1.. 8.

UniProt The Universal Protein Resource. The UniProt Consortium, http://www.uniprot.org.

Untergasser A., Cutcutache I., Koressaar T., Ye J., Faircloth B.C., Remm M. & Rozen S.G. (2012) Primer3—new capabilities and interfaces. *Nucleic acids research* **40**, e115-e.

USEF (2016) *RULE BOOK*. United States Equestrian Federation,,

Vignal A., Milan D., SanCristobal M. & Eggen A. (2002) A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution* **34**, 1.

Visscher P.M., Brown M.A., McCarthy M.I. & Yang J. (2012) Five years of GWAS discovery. *The American Journal of Human Genetics* **90**, 7-24.

Votion D.-M., Gnaiger E., Lemieux H., Mouithys-Mickalad A. & Serteyn D. (2012) Physical fitness and mitochondrial respiratory capacity in horse skeletal muscle. *PLoS One* **7**, e34890.

Wade C., Giulotto E., Sigurdsson S., Zoli M., Gnerre S., Imsland F., Lear T., Adelson D., Bailey E. & Bellone R. (2009) Genome sequence, comparative analysis, and population genetics of the domestic horse. *Science* **326**, 865-7.

WAHO (2015) Arabian horse definition. World Arabian Horse Organization

Wallace D.C., Ye J., Neckelmann S.N., Singh G., Webster K.A. & Greenberg B.D. (1987) Sequence analysis of cDNAs for the human and bovine ATP synthase β subunit: mitochondrial DNA genes sustain seventeen times more mutations. *Current genetics* **12**, 81-90.

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**, 158-64.

Wallner B., Vogl C., Shukla P., Burgstaller J.P., Druml T. & Brem G. (2013) Identification of genetic variation on the horse y chromosome and the tracing of male founder lineages in modern breeds. *PLoS One* **8**, e60015.

WCMC (1992) Global Diversity. (ed. by Groombridge B), pp. 13-4. The World Conservation Monitoring Centre, UK.

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

Wiggans G., VanRaden P. & Cooper T. (2011) The genomic evaluation system in the United States: Past, present, future. *Journal of dairy science* **94**, 3202-11.

Williams L.M., Ma X., Boyko A.R., Bustamante C.D. & Oleksiak M.F. (2010) SNP identification, verification, and utility for population genetics in a non-model genus. *BMC genetics* **11**, 1.

Winton C.L., Plante Y., Hind P., McMahon R., Hegarty M.J., McEwan N.R., Davies-Morel M.C., Morgan C.M., Powell W. & Nash D.M. (2015) Comparative genetic diversity in a sample of pony breeds from the U.K. and North America: a case study in the conservation of global genetic resources. *Ecol Evol* **5**, 3507-22.

Wolfarth B., Rankinen T., Mühlbauer S., Scherr J., Boulay M.R., Pérusse L., Rauramaa R. & Bouchard C. (2007) Association between a β 2-adrenergic receptor polymorphism and elite endurance performance. *Metabolism* **56**, 1649-51.

Wollstein A., Herrmann A., Wittig M., Nothnagel M., Franke A., Nürnberg P., Schreiber S., Krawczak M. & Hampe J. (2007) Efficacy assessment of SNP sets for genome-wide disease association studies. *Nucleic acids research* **35**, e113.

Wood C.H. & Jackson S.G. (1989) *Horse Judging Manual*. University of Kentucky Cooperative Extension Service.

Wright S. (1951) The genetical structure of populations. *Ann. Eugen.* **15**, 323-54.

Wright S. (1969) Evolution and the Genetics of Populations: The Theory of Gene Frequencies, vol. 2University of Chicago Press. *Chicago, USA*.

Wright S. (1978) *Evolution and the Genetics of Populations: A Treatise in Four Volumes: Vol. 4: Variability Within and Among Natural Populations*. University of Chicago Press.

Xiufeng X. & Árnason Ú. (1994) The complete mitochondrial DNA sequence of the horse, Equus caballus: extensive heteroplasmy of the control region. *Gene* **148**, 357-62.

Xu L., Yang S., Lin R., Yang H., Li A. & Wan Q. (2012) Genetic diversity and population structure of Chinese pony breeds using microsatellite markers. *Genetics and Molecular Research* **11**, 2629-40.

Yang N., MacArthur D.G., Gulbin J.P., Hahn A.G., Beggs A.H., Easteal S. & North K. (2003) ACTN3 genotype is associated with human elite athletic performance. *The American Journal of Human Genetics* **73**, 627-31.

Yang Y., Zhu Q., Liu S., Zhao C. & Wu C. (2016) The origin of Chinese domestic horses revealed with novel mtDNA variants. *Animal Science Journal*.

Yue X., Qin F., Campana M., Liu D., Mao C., Wang X., Lan X., Chen H. & Lei C. (2012) Characterization of cytochrome b diversity in Chinese domestic horses. *Animal genetics* **43**, 624-6.

Zanoteli E., Lotuffo R.M., Oliveira A.S., Beggs A.H., Canovas M., Zatz M. & Vainzof M. (2003) Deficiency of muscle α-actinin-3 is compatible with high muscle performance. *Journal of molecular neuroscience* **20**, 39-42.

Zhang H., Wang Z., Wang S. & Li H. (2012) Progress of genome wide association study in domestic animals. *Journal of animal science and biotechnology* **3**, 1.

8. Appendix

Table S3.1. Markers per chromosomes before and after quality control

ECA	Before	After	%		ECA	Before	After	%
1	4889	2851	58.3		18	2170	1353	62.4
2	3453	2039	59.1		19	1724	1057	61.3
3	3041	1847	60.7		20	1793	1031	57.5
4	2966	1805	60.9		21	1691	1035	61.2
5	2783	1683	60.5		22	1440	893	62.0
6	2424	1434	59.2		23	1516	913	60.2
7	2641	1595	60.4		24	1410	881	62.5
8	2680	1611	60.1		25	1087	641	59.0
9	2352	1405	59.7		26	970	517	53.3
10	2338	1363	58.3		27	1086	636	58.6
11	1799	1062	59.0		28	1257	772	61.4
12	903	482	53.4		29	812	538	66.3
13	1167	626	53.6		30	828	504	60.9
14	2683	1623	60.5		31	640	365	57.0
15	2561	1522	59.4		X	3409	1678	49.2
16	2437	1480	60.7		Y	2	0	0.0
17	2204	1429	64.8					

Table S4.1. Distribution pattern of 38 variations of 353 bp segment of mtDNA D-loop in three strains of Arabian horses

ID	HT Code	15495	15521	15532	15534	15538	15542	15544	15585	15598	15600	15602	15603	15604	15615	15616	15617	15635	15649	15650	15659	15666	15667	15703	15709	15720	15726	15740	15770	15771	15775	15776	15777	15806	15807	15809	15811	15826	15827
Ref.	NC001640	T	G	C	C	A	C	T	G	A	G	C	T	G	A	A	T	C	A	A	T	G	A	T	C	G	G	A	C	C	C	T	A	C	C	A	C	A	A
H	H001	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	A	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	G	.
H	H002	C	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	C	.	A	.	G	.	T	.	.	G	.	.	.	T	.	.
H	H003	C	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.
H	H004	C	.	.	.	.	T	.	A	G	.	T	.	.	.	.	.	T	.	G	.	A	.	C	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.
H	H005	C	.	.	.	.	.	.	.	.	.	T	.	A	.	.	.	.	.	.	.	.	.	C	.	A	.	G	.	T	.	.	G	.	.	.	T	.	.
K	K002	C	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	C	.	A	A	G	.	T	.	.	G	.	.	.	T	.	.
K	K003	C	.	.	.	.	T	.	A	.	.	T	.	.	.	.	.	T	.	G	.	A	.	C	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.
K	K004	C	.	_	.	.	.	.	A	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	T	.	.	.	.	T	.	.	.	G
K	K005	C	.	.	.	.	.	.	.	.	.	T	.	.	G	G	.	.	.	.	C	.	.	C	.	A	.	.	T	.	T	C	.	T	.	.	.	.	G
K	K006	C	.	.	.	.	.	.	.	G	.	T	.	A	.	.	.	.	.	.	.	.	.	C	.	A	.	.	.	T	.	.	G	.	.	G	.	.	.
K	K007	C	.	.	.	.	T	.	.	G	.	T	.	.	.	.	.	.	.	G	.	A	.	.	.	A	.	.	.	T	.	.	.	.	.	.	.	.	.
K	K008	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	T	.	.	.	.	.	.	.	.	G	.
K	K009	C	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	G	.	A	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	G	.
K	K010	C	A	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	T	.	.	.	.	.	.	.	.	.	.
K	K011	C	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	T	.	.	.	T	.	.	.	.	G
K	K012	C	.	.	.	G	.	.	A	.	A	T	.	.	.	.	.	.	.	.	.	.	.	.	T	A	.	.	.	.	.	.	.	.	.	.	.	G	.
K	K013	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.
S	S002	C	.	.	.	.	.	.	.	.	.	T	.	A	.	.	.	.	.	.	.	.	.	C	.	A	A	G	.	T	.	.	G	.	.	.	T	.	.
S	S003	C	.	.	.	.	.	C	.	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	G	.
S	S004	C	.	.	.	.	.	.	.	G	.	T	.	A	.	.	.	.	.	.	.	.	G	C	.	A	.	.	.	T	.	.	G	.	.	G	.	.	.
S	S005	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.
S	S006	C	.	.	.	G	.	.	.	G	.	T	.	.	.	.	.	.	.	G	.	.	.	.	T	A	.	.	.	T	.	.	.	.	.	.	.	G	.
S	S007	C	.	.	T	.	.	.	.	.	.	T	C	.	.	.	.	.	G	.	.	.	.	.	.	A	.	.	.	T	.	.	.	.	.	.	.	.	.
S	S008	C	.	.	.	.	.	.	.	.	.	T	.	A	.	.	.	.	.	.	.	.	.	C	.	A	A	G	.	T	.	.	.	.	.	.	T	.	.
S	S009	C	.	.	.	.	.	.	.	.	.	T	.	.	.	.	C	.	.	.	C	.	.	.	.	A	.	.	.	T	C	.	.	T	.	.	.	.	G
S	S010	C	.	.	.	.	.	.	.	.	.	T	.	A	.	.	.	.	.	.	.	.	G	C	.	A	.	.	.	T	C	.	G	.	.	G	.	.	.
SKH	SKH1	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	A	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.
KS	SK01	C	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	G	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	G	.

The identified 28 Arabian horse haplotype sequences of mtDNA D-loop compared to the GenBank horse sequence NC-001640. Dots indicate concordance with the NC-001640 sequence. ID is representing the abbreviation of the three strains: Saglawi (S), Kahlawi (K) and Hamdani (H).

Table S4.2. Mitochondrial haplotypes identified within 192 Arabian horses representing three strains.

HT	Maternal Lineages	Ind. Per maternal lineage	Acc.Nr.
H001	1	1	JN398390
H002	2	30	JN398452
H002	3	3	JN398452
H003	4	20	JN398378
H004	5	1	JN398404
H004	6	1	JN398404
H005	7	1	KF192402
H005	8	1	KF192402
HSK1	9	1	AY805661
HSK1	10	1	AY805661
K002	11	2	JN398450
K002	12	7	JN398450
K002	13	3	JN398450
K003	14	1	AY246236
K003	15	5	AY246236
K004	16	12	KF038163
K004	17	5	KF038163
K004	18	2	KF038163
K004	15	1	KF038163
K005	19	2	AP013078
K005	17	1	AP013078
K005	20	2	AP013078
K006	21	4	**KT724969**
K006	22	3	**KT724969**
K007	23	6	AY246231
K007	24	1	AY246231
K007	25	1	AY246231
K008	26	4	EF437565
K009	27	1	JN398389
K009	28	1	JN398389
K009	29	1	JN398389
K009	30	1	JN398389
K010	31	1	AY246253
K010	32	1	AY246253
K011	33	1	JN398437
K012	34	1	**KT724968**
K012	35	1	**KT724968**
K013	36	1	JN398379
K013	37	1	JN398379
KS01	38	1	JN398395
KS01	39	1	JN398395

HT	Maternal Lineages	Ind. Per maternal lineage	Acc.Nr.
KSH1	40	1	AY805661
KSH1	41	1	AY805661
S002	42	2	AY246174
S003	43	2	**KT724970**
S003	44	2	**KT724970**
S004	45	2	JN398448
S004	46	2	JN398448
S004	47	2	JN398448
S004	48	1	JN398448
S005	49	1	JN398391
S006	50	1	AY805660
S006	51	1	AY805660
S006	52	1	AY805660
S006	53	1	AY805660
S007	54	1	AY519946
S007	55	1	AY519946
S007	56	1	AY519946
S007	57	2	AY519946
S008	58	1	HQ439465
S008	59	1	HQ439465
S008	60	1	HQ439465
S008	61	1	HQ439465
S008	62	1	HQ439465
S008	63	2	HQ439465
S009	64	1	HQ827110
S010	65	1	KF192415
SK01	66	21	JN398395
SKH1	67	3	AY805661
SKH1	68	1	AY805661

The identified 28 Arabian horse haplotypes (HT) which are representing 68 maternal lineages and numbers of individuals per each maternal lineage. The new haplotypes which have been identified and their accession numbers in the GenBank: KT724968 (K012), KT724969 (K006) and KT724970 (S003) (in bold). Strains abbreviations are: Hamdani ''H'', Saglawi ''S'' and Kahlawi ''K''

Table S4.3. Number of polymorphic alleles (N_a), number of effective alleles (N_e), number of private alleles or rare alleles with frequency less than 0.1 (N_{Pa}) and allelic range size (bp) per locus per strain.

Strains	Kahlawi			Hamdani			Saglawi			Allelic range (bp)
Loci	N_a	N_e	N_{pa}	N_a	N_e	N_{pa}	N_a	N_e	N_{pa}	
AHT4	8	4.478	1	5	3.725	-	7	3.689	-	148 - 162
ASB17	11	7.171	1	11	6.531	1	11	9.064	-	91 - 135
ASB23	7	3.020	-	6	3.549	-	7	4.455	1	148 - 170
HMS1	5	3.358	-	7	3.066	2	6	3.350	1	171 - 189
HMS2	11	5.294	4	8	4.333	-	7	4.653	-	218 - 240
HMS3	7	4.905	-	5	3.684	-	10	5.851	3	148 - 170
HMS6	6	3.888	1	6	4.097	1	6	3.548	1	131 - 171
HMS7	6	4.891	-	6	4.212	-	7	4.766	-	169 - 181
HTG4	6	3.659	1	3	2.499	-	5	3.213	-	127 - 137
HTG7	3	1.569	-	4	1.767	1	5	1.620	2	118 - 179
LEX3	8	3.462	1	8	5.121	-	11	6.078	2	190 - 214
VHL20	6	2.839	-	7	2.978	-	9	4.695	2	87 - 105
Overall	84	-	9	76	-	5	91	-	12	-
Mean ± S.E.	7.000 ±0.663	4.044 ±0.414		6.333 ±0.607	3.797 ±0.357		7.583 ±0.621	4.582 ±0.537		

Table S4.4. Observed (H_o) and expected (H_e) heterozygosities and polymorphic information content (PIC) for loci in 12 microsatellites set in three Syrian Arabian horses strains.

Strains	Kahlawi			Hamdani			Saglawi		
Loci	**H_o**	**H_e**	**PIC**	**H_o**	**H_e**	**PIC**	**H_o**	**H_e**	**PIC**
AHT4	0.933	0.790	0.777	0.846	0.746	0.732	0.893	0.742	0.729
ASB17	0.900	0.875	0.861	0.846	0.863	0.847	0.786	0.906	0.890
ASB23	0.567	0.680	0.669	0.654	0.732	0.718	0.464	0.790	0.776
HMS1	0.767	0.714	0.702	0.769	0.687	0.674	0.714	0.714	0.702
HMS2	0.800	0.825	0.811	0.808	0.784	0.769	0.607	0.799	0.785
HMS3	0.700	0.810	0.796	0.615	0.743	0.729	0.786	0.844	0.829
HMS6	0.867	0.755	0.743	0.654	0.771	0.756	0.679	0.731	0.718
HMS7	0.600	0.809	0.796	0.692	0.778	0.763	0.786	0.805	0.790
HTG4	0.733	0.739	0.727	0.577	0.612	0.600	0.643	0.701	0.689
HTG7	0.367	0.369	0.363	0.500	0.443	0.434	0.357	0.390	0.383
LEX3	0.467	0.723	0.711	0.731	0.821	0.805	0.571	0.851	0.835
VHL20	0.667	0.659	0.648	0.654	0.677	0.664	0.464	0.801	0.787
Mean ± S.E.	0.697 ±0.050	0.729 ±0.037	0.717 ±0.037	0.696 ±0.031	0.721 ±0.032	0.707 ±0.031	0.646 ±0.046	0.756 ±0.038	0.743 ±0.037

Acknowledgment

My sincere gratitude and my first thanks go to my supervisor **Prof.Dr. Gudrun Brockmann**, who expertly guided me through my PhD research and who shared me the challenges of four years of discovery. Her unwavering enthusiasm, which kept me strongly engaged with my research, and her personal attitude helped to make my time enjoyable and exciting even in the difficult days. My appreciation also extends to **Dr. Monika Reissmann**, who believed in the importance of this topic during my Master's studies and helped in setting up the experiments of this research. Her knowledge in the horse genetics, advices and early insights lunched the greater part of this dissertation. I am also indebted towards **Dr. Derk Arends** for his support in analyzing the SNP array data and teaching me a lot through our discussions and lectures. He is the statistician who knew how to make sense of data, particularly in the GWAS. Thanks also go to **Mrs. Dorothea Schemmel**, who enlivened the laboratory atmosphere with her outstanding professional experience and good humor. A special thank for the dearest **Mrs. Birgit Jache** who offered all required administration help and always with a smile during the course of this research. I also thank deeply **Dr. Jessica Petersen** who provided, generously, SNP array genotypes for American Arabian horses through the Equine Genetic Diversity Consortium. Loads of love to all my **real friends** everywhere. I am very grateful to my friend and office mate **Dr. Ammar Said Ahmed,** who provided me with a congenial working atmosphere. His wide knowledge, ideas, manners and unconditional support encouraged me to go further in research and life. A huge thank you goes to **Dr. Hamdy Abdel-Shafy,** who shared me the office for six months, enrich my knowledge and supported me consistently with a lot of patience till the last minutes before printing this dissertation. I would like also to express the appreciation to **Dr. Siham Rahmatalla**, **Dr. Susanne Kreuzer-Redmer** and **Dr. Uwe Müller** for their caring, morality,

all great discussions and good times we spent together. Thanks to all my colleagues/friends at our institute who helped me survive all the stress in the last years. I stand firmly in gratitude to **Ramez Alrawas,** who provided his consistent moral support and smart technical solutions. It is a pleasure to express my gratitude to **Wassim Alrez** from the German Archaeological Institute for providing me with historical references. I am deeply grateful to **IDB Merit Scholarship Programme for High Technology** for awarding me and supporting the research logistics generously. Special thanks to **Mr. Ghayath Al Shaieb** from Arabian horse office at the Syrian Ministry of Agriculture and Agrarian Reform, **Prof. Dr. Issam Kasem**, **Prof. Dr. Abdulkader Rahmo** and **Dr. Manar Makhoul** from the Syrian National Commission for Biotechnology, all staff of Alnashabia farm particularly **Dr. Ahmed Shreedy** and **Mr. Ahmed Saleem**, the Syrian Arab Horse Association represented by **Mr. Bassel Jadaan** and all Arabian horse owners in Syria and worldwide for their collaboration, hosting and supporting in sampling and pedigrees validation, particularly, **Mr. Mahmoud Anzarouti** and Sham Stables. To my **committee**, I am extremely grateful for your careful reading of the manuscript, assistance and suggestions. Finally, I am indebted to my parents **Lamia** and **AbdulSamad** (May he rest in peace), my family flowers **Zara**, **Lily**, **Nora** and **Hamody**, whose value to me only grows with age and distance. And finally, I acknowledge my husband, **Christian**, who blessed me with joy and light and brought, with **Heidi** and **Wolfgang**, loads of love to my life**.**

Thanks to **Arabian horses** ….for inspiration!