

THE CONSERVATION OF GENETIC DIVERSITY IN TURKISH NATIVE HORSE
BREEDS

by

Özgür Aslan

BS. in BIO., Boğaziçi University, 2007

Submitted to the Institute of Environmental Sciences in partial fulfillment of

the requirements for the degree of

Master of Science

in

Environmental Sciences

Boğaziçi University

2009

ACKNOWLEDGMENTS

I would profoundly like to express my sincere gratitude to my supervisor Assoc. Prof. Sezen Arat and Prof. Haydar Bağış for letting me be a part of this project and for their precious guidance, advices and encouragements throughout my studies. I would deeply like to express my sincere appreciation to my thesis advisor Assist. Prof. Raşit Bilgin for his invaluable expertise, encouragements, and great patience.

My appreciation is also extended to the members of my thesis committee Prof. Orhan Yenigün and Prof. Işıl Balcıoğlu for devoting their time to evaluate this work and their helpful comments and criticisms on the manuscript.

No product is the effort of one person. Through the assistance of others together with their subtle encouragement and support, an effort produces a result. I would like to express my grateful thanks to Dr. Evren Koban for her substantial contributions, valuable criticism and continuous guidance which greatly improved my skills in the laboratory and in the field. Additionally, I would like to give my sincere thanks to Dr. Diğdem Aktopraklıgil, Dr. Soner Aksu and Gazi Turgut for their substantial contributions and meritorious helps to my study. I would like state my special thanks to all those people including the veterinarians, veterinary technicians and all the farmers who helped us for the collection of the samples. I am grateful to each member of my laboratory friends especially Melis Denizci for their valuable and wonderful friendship and creating the best laboratory environment ever.

I would like to express my great thanks to my mother Fatma Aslan, my sister Ayşegül Candemir, my brother in law Erhan Candemir, Berra and Betül Candemir who are the cutest nieces of the world and my best friends for standing by my side at all time. I am grateful to Ayşe Ezgi Akkaya for being there whenever I needed help and for making the life easier for me with her endless love and patience.

This study is a part of TURKHAYGEN-1 project supported by Turkish Scientific and Technical Research Council (grant no:KAMAG 106G005).

ABSTRACT

THE CONSERVATION OF GENETIC DIVERSITY IN TURKISH NATIVE HORSE BREEDS

In the last two decade, the loss of animal genetic resources, as an inevitable result of the decrease both in the number of species and in the population size, has become an important issue and many countries have prepared their own action plan concerning the conservation strategies on the basis of FAO's global management program. Although, wild species generally have been prioritized in the conservation studies, domesticated animals should also be considered as a part of the present biodiversity. Identifying genetic compositions of species is the first step of subsequent breeding programs and conservation studies. The domestication events of several animal species namely sheep, cattle and goat were carried out in Anatolia. Even though the Anatolian peninsula is located within several immigration and trade routes between the east and the west, and horses were used for transportation and warfare, there have been no studies concerning the origins and the evolutionary processes of Turkish native livestock horse breeds based on DNA markers. In this study, 319 base pairs of the mitochondrial DNA (mtDNA) D-loop region of 317 horse samples representing five phenotypically distinct Turkish native horse breeds was analyzed, genetically characterized and to set up the basis for a conservation program. According to the number of variable sites identified within D-loop region, Turkish native horse breeds have high haplotype, and low nucleotide diversities. In addition, because of the high motility of horses, inappropriate horse trading habits, and lack of proper breeding strategies for horses in Turkey, significant genetic separations were not observed among horse breeds. A phylogenetic network constructed based on mtDNA sequences including various Turkish native livestock horses and GenBank sequences representing seven previously published haplogroups showed that Turkish horse breeds clustered into these seven clades and one additional unique clade which could be named as an eighth haplogroup. mtDNA analysis of Turkish horses also indicates the presence of many ancient maternal lineages within.

ÖZET

YERLİ TÜRK ATLARINDA GENETİK ÇEŞİTLİLİĞİN KORUNMASI

Son yıllarda, hayvan türü sayısındaki ve populasyon ve populasyon büyüklüğündeki azalmanın kaçınılmaz sonucu olan genetik kaynakların kaybedilmesi, önemli bir konu haline gelmiştir ve birçok ülke FAO'nun hazırladığı global yönetim programına uygun olarak kendi koruma stratejilerini oluşturmuştur. Vahşi hayvanlara genel olarak koruma programlarında öncelik verilmesine rağmen, evcil hayvanlar da biyo-çeşitliliğin parçası olarak kabul edilmeli ve koruma programlarına eklenmelidir. Türlerin genetik yapılarının incelenmesi koruma programlarının ve tohumlama stratejilerinin ilk adımıdır. Koyun, keçi ve sığır türlerinin evcilleştirilmesi Anadolu'da gerçekleştirilmiştir. Geçmişte Anadolu yarım adası birçok göç ve ticaret yolunda bulunmasına ve atların taşımada ve savaşta kullanılmalarına rağmen, Türk yerli at ırkları hakkında genetik işaretleyicileri kullanarak şu ana kadar herhangi bir çalışma yapılmamıştır. Bu çalışmada, ırkların genetik karakterizasyonunun yapılması ve koruma programının temellendirilmesi için beş adet Türk at ırkını temsil eden 317 at örneğinin 319 baz çiftlik mitokondriyal D-loop bölgesi kullanılmıştır. Farklılık gösteren D-loop bölgelerinin analizi sonucunda, yerli Türk at ırklarının haplotip çeşitliliğinin yüksek olduğu ve nükleotid bilgisinin düşük olduğu bilgisine ulaşılmıştır. Bunun yanı sıra, uygun olmayan yüksek at hareketliliği ve atlar için tohumlama stratejisinin olamaması sebebiyle, ırklar arasında geçerli bir ayrılma gözlemlenmemiştir. Yerli Türk atları ve daha önce yayınlanmış yedi adet haplogrubu temsil eden 22 örnek kullanılarak bir tane filogenetik ağ oluşturulmuştur. Türk atları, A-G olarak isimlendirilen bu yedi grupta ve sekizinci grup olarak adlandırılacak olan farklı bir grupta toplanmıştır. Ek olarak, Türk atlarının mtDNA analizleri, atların evcilleştirilmesi sürecine bir çok atasal kısrak ırkının dahil olduğunu doğrulamıştır.

TABLE OF CONTENTS

ACKNOWLEDGMENTS	iii
ABSTRACT	iv
ÖZET	v
TABLE OF CONTENTS	vi
LIST OF FIGURES	viii
LIST OF TABLES	xi
LIST OF SYMBOLS/ABBREVIATIONS	xii
1. INTRODUCTION	1
1.1. Literature Review	5
2. MATERIALS AND METHODS	9
2.1. Sample Collection	9
2.2. Laboratory Protocols	11
2.2.1. DNA Extraction	11
2.2.2. PCR Amplification and PCR Product Purification	12
2.2.3. Cycle Sequencing	14
2.3. Data Analysis	14
3. RESULTS AND DISCUSSION	18
3.1. PCR Analysis And Data Collection	18
3.2. Descriptive Data Analysis and Within Population Differentiation	18
3.3. Between Population Differentiation	24
3.4. Phylogenetic Reconstruction and Network Analysis	28
4. CONCLUSIONS AND RECOMMENDATIONS	33
REFERENCES	35
APPENDIX A: 319 BP LONG mtDNA SEQUENCES OF 114 DIFFERENT HAPLOTYPES	46
APPENDIX B: LIST OF HAPLOTYPES INDICATING VARIABLE NUCLEOTIDES IN A 478 BP FRAGMENT OF THE MITOCHONDRIAL DNA D-LOOP	76

APPENDIX C: MISMATCH DISTRIBUTION GRAPHS OF EACH POPULATION BASED ON BOTH POPULATION EXPANSION AND CONSTANT POPULATION SIZE	85
---	----

LIST OF FIGURES

Figure 2.1. Map of Turkey indicating the sampling locations. Each code number on the figure represents the sampling city given in Table 2.1	11
Figure 2.2. The agarose gel photograph of 1:10 diluted DNA samples after electrophoresed at 90 V for 30 minutes. Mass Ruler DNA ladder mix (SM 0313) was used as a marker	12
Figure 2.3. The view of PCR product after electrophoresed at 90 V and for 40 minutes. MassRuler DNA ladder mix (SM 0313) was used as a marker	13
Figure 2.4. The agarose gel photograph of purified DNA samples after they were electrophoresed at 90 V for 60 min. MassRuler DNA ladder mix (SM 0313) was used	13
Figure 2.5. The view of aligned consensus sequences	15
Figure 3.1. The chromatogram and its sequence from one sample	18
Figure 3.2. The mismatch graph of all samples based on the theory that a population expansion has occurred	25
Figure 3.3. The mismatch graph of all samples based on the theory of a constant-sized population	25
Figure 3.4. Mitochondrial DNA haplogroup distributions in Eurasian horse populations Haplogroups are colored as follows: dark blue =A, light blue=B, red=C, green=D, brown=E, orange=F and pink=G	28
Figure 3.5. Mitochondrial DNA haplogroup distributions among Turkish native horse breeds and representative Genbank sequences. Haplogroups are	

colored as follows: Red=A, Green=B, Yellow=C, Blue=D, Purple=E, Turquoise=F, Grey=G, Teal=Ancient Samples, Olive= Pleistocene samples

29

Figure 3.6. Median network of Turkish horse haplotypes including 22 GenBank sequences. Haplotypes were visualized with pie-charts and the size of charts is proportional to the number of individuals showing a particular haplotype. One single mutation separates nodes connected by a straight segment. Each breed has a unique color as follows: Green=AP, Blue=CNK, Pink=MLK, Orange=HH, Black= EAST ANATOLIAN, Yellow=CKR, Brown=ERZ, Purple=A, Dark Green=B, Turquoise=C, White=D, Light Pink=E, Dark Blue=F

32

Figure C.1. The graph of AP based on the theory of a constant-sized population

86

Figure C.2. The graph of AP based on the theory that a population expansion has occurred

86

Figure C.3. The graph of CKR based on the theory that a population expansion has occurred

87

Figure C.4. The graph of CKR based on the theory of a constant-sized population

87

Figure C.5. The graph of CNK based on the theory that a population expansion has occurred

88

Figure C.6. The graph of CNK based on the theory of a constant-sized population

88

Figure C.7. The graph of HH based on the theory that a population expansion has occurred

89

Figure C.8. The graph of HH based on the theory of a constant-sized population

89

Figure C.9. The graph of ERZ based on the theory that a population expansion has occurred	90
Figure C.10. The graph of ERZ based on the theory of a constant-sized population	90
Figure C.11. The graph of MLK based on the theory that a population expansion has occurred	91
Figure C.12. The graph of MLK based on the theory of a constant-sized population	91
Figure C.13. The graph of EAST ANATOLIAN horses based on the theory that a population expansion has occurred	92
Figure C.14. The graph of EAST ANATOLIAN horses based on the theory of a constant-sized population	92

LIST OF TABLES

Table 2.1. List of sampling location, breed name, abbreviation and number of samples collected	9
Table 3.1. List of name samples grouped in haplotypes and frequencies of haplotypes	19
Table 3.2. Population diversity parameters	23
Table 3.3. The average number of (Dxy) and the number of net (Da) nucleotide substitution per site, lower and upper diagonals respectively, between the populations	24
Table 3.4. Fu's, Tajima's, Ramos-Onsis and Rozas's Neutrality Test values	26
Table 3.5. Population pairwise ϕ_{st} values between mitochondrial sequences of breeds are below diagonal. The probabilities of the estimated ϕ_{st} values are given above diagonal	26

LIST OF SYMBOLS/ABBREVIATIONS

Symbol	Explanation
AP	Ayvacık pony
B.C.	Before christ
B.P.	Before the present
bp	Base pair
CKR	Çukurova horse
CNK	Canik horse
Da	Average number of nucleotide substitution
ddATP	Dideoxyadenosine Triphosphate
ddCTP	Dideoxycytidine Triphosphate
ddGTP	Dideoxyguanosine Triphosphate
ddUTP	Dideoxyuridine Triphosphate
D-loop	Displacement loop
DNA	Deoxyribonucleic Acid
DNTP	Deoxyribonucleotide triphosphate
Dxy	Number of net nucleotide substitution
ERZ	Erzurum horse
FAO	Food and agriculture organization
Hap	Haplotype
H _d	Haplotype diversity
HH	Hınıs horse
MLK	Malakan horse
mtDNA	Mitochondrial deoxyribonucleic acid
PCR	Polymerase chain reaction
TNHB	Turkish native horse breeds
π	Nucleotide diversity
Φ_{st}	Coancestry coefficient

1. INTRODUCTION

The current global rates of species loss and extinction levels are so high that it is considered to represent the “sixth mass extinction” (Leakey and Lewin, 1995). The first five extinction events included those such as the Permian extinction, where 95 % of all marine invertebrates went extinct, and the K-T extinction resulting in the demise of dinosaurs 65 million years ago. The main causes of the current extinction crisis is related to human impacts on the environment, such as deforestation, over-harvesting, and pollution (Furman and Yenigün, 1998) that causes decrease in the size of the populations below the minimum viable number of individuals to ensure the survival of species leading to a growing decrease in the number of species present as well as a decrease in the genetic diversity present in the species. The loss of genetic diversity ultimately leads to inbreeding depression, which manifests itself in decreasing the fitness of individuals at various levels (Charlesworth and Charlesworth, 1987). Hence, the loss of genetic diversity is a major concern in the current extinction crisis and needs to be investigated to be able to prevent it.

Animal genetic resources are considered as a part of the present biodiversity, which comprises domestic animals having a vital role in regulation and improvement of social and economic life. Loss of animal genetic resources has also become a global concern because undocumented genetic diversity at present may contain valuable data for future studies. Turkey, which has a great number of endemic species, has an important share in the global genetic resources. The loss of genetic diversity is also seen within domesticated species, as well as wild ones.

In human history, radical restructuring of human societies, worldwide alterations in biodiversity, and significant changes in the Earth's landforms and its atmosphere are associated with the transition from foraging and hunting, to farming and herding. Because of this relationship and outcomes of domestication events, identification of origins and spread of domesticates is proceeding interest for scientific community In the past decades, tracking changes in human prey strategies, in order to mark the transition from hunting to herding was mainly done by using morphometric methods. Besides, genetic analyses and

sample atomic mass spectrometry radiocarbon dating serve essential new approaches into investigating original livestock domestication and their dispersal (Zeder, 2008).

Until the late 1990s, a putative sharp and rapid reduction in overall body size in domesticates was used to identify where and when domestication events took place (Uerpmann, 1979; Meadow, 1989). However, both modern skeletal collections and archaeological caprine (sheep and goat) provided little support to accept that domestication event results in an automatic overall reduction in body size in managed animals (Bar and Meadow, 1995). Environmental conditions, of increasing heat and aridity, for instance, seem to be more positively correlated with smaller size. Body size reduction at the initial stage of livestock domestication is mostly related to differences in the culling strategies of herders as opposed to hunters (Zeder, 2008).

On the other hand, the systematic difference between individuals of different sexes in caprines has been used as another approach to identify the transition from hunting to herding. The resolution of harvest profiles of male and female caprines are different between the prey strategies of hunters from the harvest strategies of herders. On the basis of sequence and timing of long bone fusion, sizes of male and female skeletal elements help to separate archaeological assemblages. As a result of several studies in according to this approach, harvesting of young males and prolonged survivorship of females was observed (Zeder, 2006).

In southeastern Anatolia at *ca.* 10,500 B.P. both alterations in the age of harvested caprines and morphological size changes in the initial stage of herd management were observed. On the other hand, in Central Anatolia between 10,400 and 9,400 B.P., several demographic evidences for the management of morphologically unaltered caprines were found. Thus, there occurred a transition from hunting to herding in this region. On the basis of these findings, probably independently of one another and possibly multiple times, domestication of both sheep and goats took place in the region between northern Zagros and southeastern Anatolia in 11,000 and 10,500 B.P. (Peters *et al.*, 2005; Vigne *et al.*, 1999).

Several recent investigations shed light the context of the domestication of major livestock species, specifically pigs and cattle. According to archaeological evidence, pigs were first domesticated somewhere in southeastern Anatolia by 10,500-10,000 B.P. and taurine cattle were initially domesticated somewhere in the upper Euphrates Valley between *ca.* 11,000 and 10,000 B.P. (Helmer *et al.*, 2005). Sheep, pigs and cattle arrived relatively late in more distant parts of the Fertile Crescent.

In addition, according to several archaeo-zoological studies, three geographical areas for domestication events are accepted. Cattle, sheep, goats, pigs and buffalo were domesticated in Asia (the Fertile Crescent and its eastern margin, towards the Indus valley region) (Salamini *et al.*, 2002) and East Asia (China and countries south of China). On the other hand, llamas and alpacas were domesticated in the Andean chain of South America 8000-10000 years before present (Bruford *et al.*, 2003).

However, horse is a notable exception because there are lots of areas in which domestication events took place. Besides, horse was domesticated relatively recently when compared to other animal species (6,000 years before present) (Vila *et al.*, 2001).

Comprising a part of the Silk Road, many armies and tradesmen used Anatolian peninsula as the major bridge when moving between east and west in the past. As they enhanced transportation and warfare capabilities, horses played a remarkable role throughout all this movement. In this regard, Persian, Arabian and Thracian horses, as well as Russian, Mongolian and Caucasian horse breeds have contributed to the formation of Anatolian horse breeds (Batu, 1938). Among these breeds, Anatolian, Çukurova Horse, Ayvacık Pony breeds are classified as major breeds. On the other hand, Çamardı Kulası Horse, Malakan Horse, Canik Horse, Hınıs Horse and others are known as some local breeds which are present in Anatolia (Batu, 1938; Arpacık, 1994). After the establishment of Turkish Republic, in order to improve their quality, Anatolian horse breeds were crossed with Nonius, Ardene and Halflinger breeds from Europe.

In addition to those, Thoroughbred and Arabian horses were also brought to Turkey for the improvement of the native horses, creating half-breed populations (Batu, 1938). However, the need for horse power was dramatically reduced in the recent years (the last

two-three decades) due to the modernization of farming and transportation. Hence horses have not been as much utilized and/or raised in Anatolia as before and there is a continuous decline in the number of people breeding horses. Consequently, they face the threat of extinction. The size of the horse population decreased from 1.1 million in 1960 to 300.000 in 2006 (Anonym, 2006).

On the basis of these facts, Turkish livestock breeds (including Turkish native horse breeds) should be under conservation because they are on the edge of extinction, as mentioned above, due to the mechanization of agriculture and increasing tendency for inbreeding and artificial selection to increase productivity. Hence, characterizing and also showing the variability of Anatolian breeds and conservation by construction of DNA banks are necessary to avoid loss of this animal diversity (Bruford *et al.*, 2003).

The domestication of horses has been widely affected by the course of civilization. In 1998, Lister and his friends suggested that “the extent of modern haplotype diversity probably reflects an input of wild animals of different areas”. In order to understand the basis and processes of domestication and evolutionary history of horse lineages, several studies in different regions have been performed (Bruford *et al.*, 2003). However, there have been no genetic studies based on mtDNA diversity on Anatolian domestic horse livestock. Since Anatolia is the birthplace of several ancient societies and contains the first human settlements, characterization and investigation of Anatolian domestic horse breeds is a vital part to complete the whole evolutionary picture of horse domestication and diversification.

As a rescue plan a project named “In vitro Conservation and Preliminary Molecular Identification of Some Turkish Native Domestic Animal Genetic Resources-1” has been carried out with the involvement of TUBITAK and 10 universities from all regions of Turkey by the contribution of Turkish Ministry of Agriculture and Rural Affairs. This project aims to establish banks (Gene banks) which preserve animal genetic resources, to start characterization of endemic domestic animal species to obtain information to be used in registration studies, to build national researcher capacity in animal genetics and biotechnology areas, and to combine and disseminate knowledge. Samples from 1500 individuals belong to total five species would be preserved.

1.1. Literature Review

Horse domestication which remains an enduring hot topic should be considered a process flowing through wild, captive, tame, and domesticated horses (Olsen, 2006). As horses have been used for riding, some clues about horseback riding such as, bit wear first appeared in the steppes east of the Ural Mountains (Brown and Anthony, 1998). Way of circle and semicircular horse corrals in Eneolithic Botai settlements is accepted to be another indirect evidence for horse domestication (Olsen, 2006; Stiff *et al.*, 2006). However, the observed age and sex structure of horse population fit the 'hunting model', which is proposed by Levine (2005). According to the hunting model, wild horses were captured and tamed for meat.

Dateable textual and artistic evidence left at the end of the third millennium B.C. could be considered to be the direct evidence for horse domestication (Levine, 2005). Several artifacts obviously symbolizing horse riding and traction is even more recent, dating to no later than the beginning of the second millennium B.C. At the middle of the second millennium horse usage for pulling chariots was common in the Near East, Greece and on the Eurasian steppe (Levine, 2005).

As mentioned before, Reduction in overall animal size and an increasing in variability are classic indicators of domestication (Uerpmann, 1990). However, there is almost no effect of domestication on the equine anatomy, in other words; there are no indisputable osteological differences between wild and domesticated horses (Levine, 2006).

There are two models about horse domestication (Clutton-Brock, 1999):

- Model I: Wild horse stock, from which all domestic horses descended, inhabited in the region classified from the Ukraine to the region of Turkestan. Domesticated horses were spread from this land, and certain breeds that are known today are as a result of artificial selection, in combination with natural selection for adaptation to local environmental conditions.

- Model II: According to the alternative hypothesis, horse domestication took place in different times and regions. The second model is based on the presence of a geographical cline in the population of wild horses. Horses in the northern part of the range are smaller and sturdier than those found in the south. These observed differences are present even in the earliest findings of the recently domesticated horses. There are considerable differences in the size and proportions of the bones from different regions (Kavar and Dovc, 2008).

Development of firstly domesticated horse would have taken relatively long time and arbitrary genetic alterations, which would have predisposed some horses to be bred in captivity (Levine, 2005). Besides, after domestication event, wild horses could, and probably would, have been introduced into domestic herds (Levine, 2006).

Mitochondrial DNA (mtDNA) is widely used as a molecular tool, to study the history of maternal lines, owing to lack of recombination and its maternal inheritance. The offspring has same haplotype or mtDNA type as their mother, if there are no mutations. Besides occasionally occurred mutations are fixed in the maternal (female) lines, therefore, distinct maternal lines can be followed back to track the population history of various species. The most variable part of mitochondrial DNA in mammals is the control (or D-loop) region, which consists of a non-coding region that is about 1200 bp long region (Xu and Arnason, 1994; Ishida *et al.*, 1994).

As a result of several studies concerning various breeds all over the world, more than 100 distinct nucleotide sequences, in other words mtDNA haplotypes, have been obtained. Lipizzans (Kavar *et al.*, 1999; Kavar *et al.*, 2002) and Arabians (Bowling *et al.*, 2000; Jansen *et al.*, 2002) breeds are the most thoroughly investigated horse breeds. High intra-breed diversity, and roughly the same clusters were observed in all studied breeds (Kavar and Dovc, 2008). In addition, coalesce of modern horse lineages occurred long before the first domestic horse appeared at about 320,000 to 630,000 years ago (Vila *et al.*, 2001). The majority of haplotypes are shared by different breeds. Thus, the majority of present nucleotide differences are before the domestication event (Lister, *et al.*, 1998).

mtDNA control regions of ancient horses from the Viking age from archeological sites in southern Sweden and Estonia dating to 1000 to 2000 years ago (Vila *et al.*, 2001), from Kwakji (Korea) dating to A.D. 700 to A.D. 800 (Jung *et al.*, 2002), from the Pompeii and Herculaneum dating to 79 B.C. (Di Bernardo *et al.*, 2004), from a Scythian princely tomb in Kazakhstan dating from the beginning of the 3rd century B.C. (Keyser-Tracqui, *et al.*, 2005), from Yakutia (Russia) from the 17th or 18th century A.D. (Keyser-Tracqui, *et al.*, 2005), from Siberia and Ural dating to 2200 B.P. (Weinstock *et al.*, 2005), from Ireland and England dating to 1244 to 1595 B.P. (McGahern *et al.*, 2006), from Bronze Age horses recovered from the Chifeng region of the 'Inner Mongolia', China dating to 4000- 2000 B.P. (Cai *et al.*, 2007), were sequenced. Haplotypes of ancient domestic horses are unambiguously similar to haplotypes of extant domestic horses because very low number of mutations (novel mutations) has occurred in the mtDNA control region during the last few thousand years. Besides, many maternal lines vanished owing to random genetic drift and natural and artificial selection. Since firstly domesticated founder mares had such haplotypes, frequent haplotypes of extant domestic horses are also widely observed among ancient horses.

According to the previously discussed model I of domestication (Clutton-Brock, 1999), extant horse breeds had originated from highly variable, but not phylogeographically structured wild horse populations. However, the main reason of lack of phylogeographic structure is gene flow by migration. Since horses are highly mobile animals, identical haplotypes might be seen in geographically different breeds. For example, identical haplotypes of wild horses from Germany (DQ007590) were found in ones from Siberia (DQ007573) in the late Pleistocene (Kavar and Dovc, 2008). Furthermore, another possible reason of the lack of consistency between mtDNA, extant breeds and/or geographic regions might be the use of the same founders in different breeds (Luís *et al.* 2006).

Horse domestication model II proposes that distinct geographical cline or presences of different subspecies are the most plausible explanation of differentiation process of wild horses. However, Lister *et al.* (1998) performed the first publication concerning horse domestication predicated on genetic data and suggested an intermediate model, which might be the most likely likely explanation of horse domestication event. As to this

intermediate model; domesticated horses were partially distributed in an extensive geographical region, and they had large enough population sizes to contain pre-existing haplotype diversity.

To understand the evolutionary history of horse livestock breeds better, molecular markers have been used. These markers should have certain characteristics. For instance, they should be variable and also should evolve rapidly, although in a constant rate. On the other hand, they should be conserved evolutionarily to allow identification of populations that are closely related. Mitochondrial DNA (mtDNA) meets these combinations of characteristics. Mutation rate of mtDNA is high because of free radicals produced during cellular respiration and due to lack of repair mechanism. Because i) mitochondrial inheritance is mediated via matrilineal transmission, ii) recombination does not occur during mitochondrial DNA duplication, and iii) displacement loop (D-loop) region of mtDNA has evolved rapidly compared to other nuclear genomic DNA (Wolstenholme, 1992), the region has been used in identification of levels of genetic diversity, and undertaking phylogenetic analysis within closely related species (Ishida *et al.*, 1995), within a species (Okumura *et al.*, 1996) and even within a breed (Bowling *et al.*, 1998). However, it should be noted that as mtDNA inherited maternally as it is without recombination, it is accepted as a one locus in the genotyping studies.

The aim of this study is to understand differences and relations of genetic structure, and complex diversification processes of Anatolian domestic horse breeds by investigating the variability of the mitochondrial D-loop region. D-loop region of mitochondrial DNA is well suited for examining phylogenetic relationships and characterization of interbreed and intrabreed variation due to its high level of sequence variation. Therefore, within the context of this study, genetic characterization, which is a prerequisite for managing genetic resources (Bjornstad and Roed, 2002), will be performed. In addition, in conjunction with the molecular analyses, two gene banks will be established to conserve the genetic diversity of Turkish native horse breeds.

2. MATERIALS AND METHODS

2.1. Sample Collection

A total 317 horse samples for five distinct horse breeds, namely Ayvacık Pony, Hınıs Horse, Malakan Horse, Canik Horse and Çukurova Horse were collected throughout ten different locations in Turkey (Table 2.1, Figure 2.1) for TÜRKHAYGEN-1 project. Distinctive phenotypes of horse breeds were known from old text books, academicians and veterinarians. On the basis of the phenotypic characters, animal scientists helped evaluating the phenotypic characteristics of the individuals sampled from these breeds. Only samples possessing phenotypic distinctness were included in the study as much as possible.

Some horses that were from Erzurum, Ağrı, Kars, Ardahan, Iğdır and Van, which did not have the characteristic phenotype of the studies breeds, were sampled in order for comparison. Samples, collected from Ağrı, Kars, Ardahan, Iğdır and Van were grouped as East Anatolian horses in the analysis.

Table 2.1. List of sampling location, breed name, abbreviation and number of samples collected

Code	The Name of Sampling Location	The Name Of Breed	The Abbreviation Of Breeds	The Number Of Samples	Photos
1	Çanakkale	Ayvacık Pony	AP	48	
2	Erzurum	Hınıs Horse	HH	60	
2	Erzurum	Erzurum Horse	ERZ	17	

Code	The Name of Sampling Location	The Name Of Breed	The Abbreviation Of Breeds	The Number Of Samples	Photos
3	Kars	Malakan Horse	MLK	32	
	Ardahan				
	Iğdır				
3	Kars	East Anatolian Horse	KRS	47	
4	Ardahan		ARD		
5	Iğdır		IGD		
6	Ağrı		AGR		
7	Van		VAN		
8	Samsun	Canik Horse	CNK	55	
9	Adana	Çukurova Horse	CKR	58	
10	Osmaniye				



Figure 2.1. Map of Turkey indicating the sampling locations. Each code number on the figure represents the sampling city given in Table 2.1

2.2. Laboratory Protocols

2.2.1. DNA Extraction

10 ml of blood sample was labeled for each sample. Phenol-isoamyl alcohol DNA isolation protocol (Sambrook *et al.*, 1989) was applied to these samples as follows:

1. 10 ml blood sample was put into a tube including 0.5 ml EDTA (0.5 M, pH 8.0) and mixed with 40 ml of 2X lysis buffer (Lysis Buffer: 770 mM NH_4Cl , 46 mM KHCO_3 , 10 mM EDTA).
2. Tubes were mixed vigorously for 10 minutes.
3. Tubes were centrifuged for 10 minutes at 3,000 rpm, 4°C.
4. Supernatant phase was discarded and 3 ml of salt/EDTA (75 mM NaCl, 25 mM EDTA) was added and mixed.
5. 0.3 ml of 10 % SDS and 150 μl of proteinase K (10 mg/ml) were added and samples were incubated at 55°C for one-three hours.
6. After incubation period, 3 ml of phenol (pH 8.0) was added to each tube. Tubes were shaken vigorously for 20 seconds and mixed well for 5 minutes.
7. Tubes were centrifuged for 10 minutes at 3,000 rpm, 4°C.

8. Supernatant phase was transferred to new tubes and 3 ml of phenol: chloroform: isoamyl alcohol (25:24:1) was added to each tube. Tubes were shaken vigorously for 20 seconds and mixed well for 5 minutes.
9. Tubes were centrifuged for 10 minutes at 3,000 rpm, 4°C.
10. Supernatant phase was transferred to 15 ml falcon tubes. Ice cold EtOH (kept at -20°C) was added to each tube, twice the volume of the supernatant. Tubes were shaken well and the extracted DNA was precipitated.
11. DNA was taken by micropipette help and transferred into 1.5 ml tubes containing 0.3 ml of 10 mM Tris-HCl (pH 8.00) solution.

The isolated DNA samples were 1:10 diluted with Tris-EDTA or Tris HCl and run in a 0.8 % agarose gel in order to observe the quality and quantity of samples (Figure 2.2).

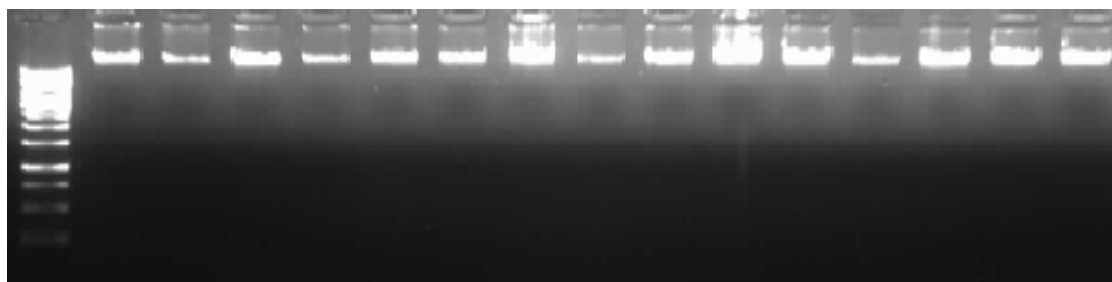


Figure 2.2. The agarose gel photograph of 1:10 diluted DNA samples after electrophoresed at 90 V for 30 minutes. Mass Ruler DNA ladder mix (SM 0313) was used as a marker

2.2.2. PCR Amplification and PCR Product Purification

Primers were designed from D-loop region of mitochondrial DNA of *Equus caballus* species. 478 bp region of D-loop was amplified with primers 5'-CCCAAGGACTATCAAGGAAG-3' and 5'-GGAATGGCCCTGAAGAAAGA-3'. For each PCR, a final volume of 50 µl was prepared including 100 ng DNA template, 1X PCR Buffer, 2.5 mM MgCl₂, 0.2 mM dNTP mix, 20 pmol of forward and reverse primers, 1 U Taq DNA polymerase and dH₂O for volume adjustment. The PCR was performed with the following program: 94°C for 3 minutes, 10 cycles of 94°C for 20 seconds, 58°C for 20 seconds, 72°C for 40 seconds. This was followed by 30 cycles of 94°C for 20 seconds, 58°C for 20 seconds, 72°C for 40 seconds and a final elongation cycle at 72°C for 10

minutes. The results of the PCRs were checked by electrophoresis using 1.0 % agarose gel (1X TAE buffer) and the products were separated based on their length (Figure 2.3).

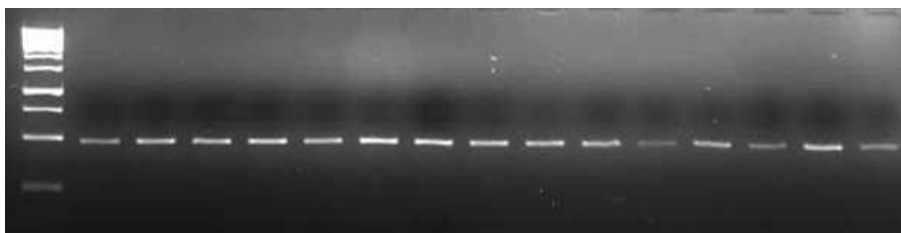


Figure 2.3. The view of PCR product after electrophoresed at 90 V and for 40 minutes.

MassRuler DNA ladder mix (SM 0313) was used as a marker

PCR products of desired length (as compared to DNA ladder) and without any non-specific amplification bands were cleaned up for further use in the sequencing reaction by using kits. Three kits were used in the study: (i) Wizard SV Gel and PCR Clean-up System (Promega) (ii) Roche Agarose Gel Extraction Kit (iii) Roche Column Kit. After clean up, the concentration of the purified PCR products were estimated by comparison using the known quantity of DNA bands of the DNA size ladder based on band thickness and shininess on 1 % agarose gel. The rough estimation of the concentration of the purified PCR products is used in determination of the volume of template to be used in the cycle-sequencing reaction (Figure 2.4).

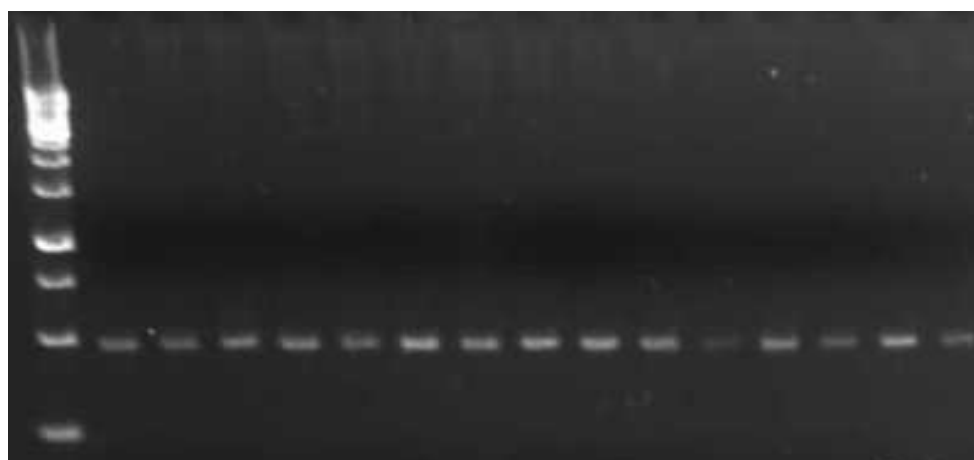


Figure 2.4. The agarose gel photograph of purified DNA samples after they were electrophoresed at 90 V for 60 min. MassRuler DNA ladder mix (SM 0313) was used

2.2.3. Cycle Sequencing

MtDNA D-loop sequencing was performed in terms of Sanger method in which 2', 3'-dideoxynucleotide triphosphates (ddNTPs) is utilized to terminate DNA chain elongation because they cannot form a phosphodiester bond with the next deoxynucleotide (Sanger, 1977).

Genome Lab Dye Terminator Cycle Sequencing with Quick Start Kit (Beckman Coulter) and a Beckman Coulter 8800 sequencer was used for the sequencing reactions. 1.6 pmol/ μ l primers, 50 fmol template (which equals to 15.5 ng for 478 bp DNA), 8 μ l DNTP premix including ddUTP dye terminator, ddGTP dye terminator, ddCTP dye terminator and ddATP dye terminator at different volumes were added. The amount of water was determined depending on the amount of PCR template used, with a final reaction volume of 20 μ l. The cycle sequencing was done as 30 cycles of 20 seconds at 96°C, 20 seconds at 50°C and 60°C for 4 minutes.

After sequencing cycle, samples are solidified by following ethanol precipitation method. Finally, samples are solved in sample loading solution and then loaded into Beckman Coulter 8800.

2.3. Data Analysis

First, forward and reverse sequences of each sample were aligned to each other forming contigs for each sequence using CLUSTALX 1.8 (Thompson *et al.*, 1997). Then, chromatograms of both forward and reverse sequences in the contigs were compared to each other, manual editing was performed where necessary. Afterwards, the consensus sequences of all the samples were taken in fasta format, ready to be aligned.

The consensus sequences were aligned using ClustalW software integrated into BIOEDIT Sequence Alignment Editor 7.0.5.2 (Hall, 1999) (Figure 2.5). The aligned sequences were trimmed from both ends in order for each consensus sequence to start from and end at the same nucleotide position with respect to the reference sequence (X79547).

Finally, the trimmed sequences were saved in fasta format to be analysed by bioinformatics softwares.

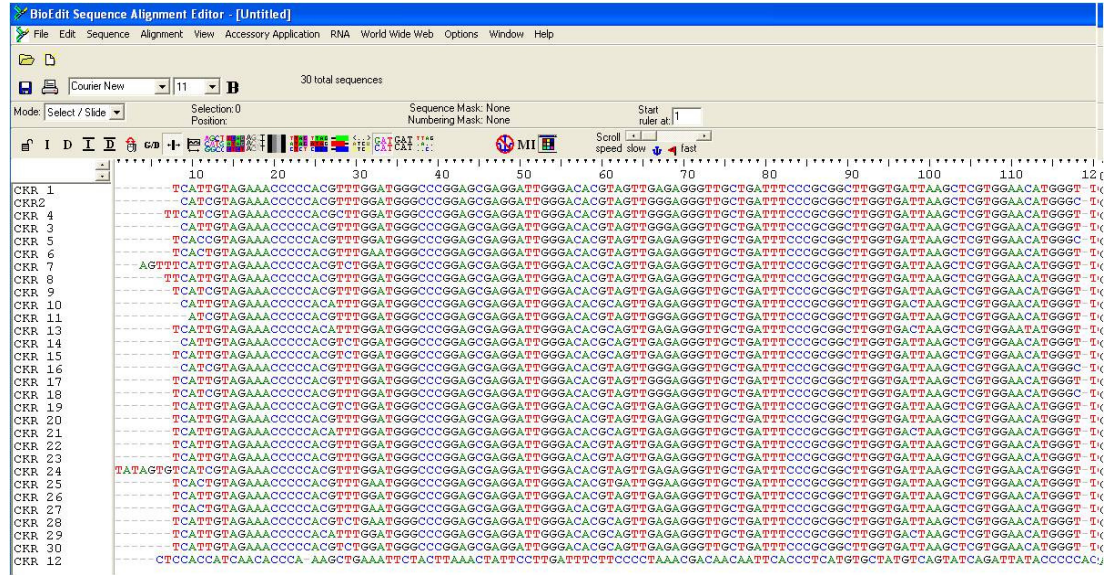


Figure 2.5. The view of aligned consensus sequences

Haplotype diversity (H_d) is uniqueness of a particular haplotype in a given population is measured by haplotype diversity. The haplotype diversity was based on the formula $h = (1 - \sum x_i^2) n / (n - 1)$, where x_i is the frequency of a haplotype and n is the sample size (Nei and Tajima, 1981).

Nucleotide diversity is denoted by π , and is a measure of genetic variation. The nucleotide diversity was first introduced by Nei and Li (1979) and is given by the Equation 2.1.

$$\pi = \sum_{ij} x_i x_j \pi_{ij} = \sum_{i=1}^n \sum_{j=1}^i x_i x_j \pi_{ij} \quad (2.1)$$

where x_i and x_j are the respective frequencies of the i^{th} and j^{th} sequences, π_{ij} is the number of nucleotide differences per nucleotide site between the i^{th} and j^{th} sequences and n is the number of sequences in the sample. The summation is taken over all distinct pairs i, j , without repetition.

Haplotype diversity (H_d), nucleotide diversity (π), number of polymorphic sites, average number of nucleotide differences, D_a and D_{xy} values among breeds, and number of polymorphic sites (Nei and Tajima, 1981; Graur and Li, 2000), were calculated using the software DNASP v4.0 (Rozas *et al.*, 2003). Genetic structure of populations was examined by the Analysis of Molecular Variance (AMOVA), and ϕ_{ST} calculations, to measure differences among populations, were performed with ARLEQUIN V. 3.11 (Excoffier *et al.*, 2005).

The genetic relationships between Turkish native horse breeds and other selected breeds were also investigated by building haplotype median networks, which were constructed with the software NETWORK 4.1.1.2. (Bandelt *et al.*, 2002) with 22 representative of the six main clades (A–F) (Vila *et al.*, 2001), (G) (Jansen *et al.*, 2002) and three ancient and three Pleistocene GenBank sequences (GenBank accession numbers; AF326676, AF326677, AF326678, AF326672, A326674, AF326669, AF064628, AF014416, AF014410, AF072996, AF14413, AF072990, AF072988, AF072992, AF072987, AF014409, AF064632, AF072977, D23666, AF168698, AF431969, X79547).

In order to assess the genetic relationships between the Turkish native horses and the other selected breeds from the rest of the world, a neighbor joining tree was constructed with software Mega 4.0 (Tamura *et al.*, 2007). 22 GenBank sequences representative of the seven main, globally recognized clades (A–G), described for the horse mitochondrial D-loop by Vila *et al.* (2001) and Jansen *et al.* (2002) were added to analysis (Figure 3.4). In fact, mitochondrial lineages are not directly used in horse breed identification because their diversification is thought to have occurred prior to domestication (Vila *et al.*, 2001; Jansen *et al.*, 2002). However, relevant information about the origins and relationships of the breeds can be obtained from detailed review of their phylogenetic trees (Perez *et al.*, 2008).

In addition, Fu's F_s (Fu and Li, 1993), Tajima's D (Tajima, 1989) and R_2 (Raos-Onsins and Rozas, 2002) neutrality test were made using DNASP v4.0 (Rozas *et al.*, 2003) to examine whether populations underwent any selection or recent demographic expansion.

In the estimation of the population parameters values, several factors including demographic changes, selection, and recombination affecting the distribution of the number of pairwise differences between DNA should be considered (Mousset *et al.*, 2004). Based on weak or absent recombination, the effects of population expansion on distribution of pairwise differences have been studied (Rogers and Harpending 1992). Rogers and Harpending (1992) indicated histograms showing the relative frequencies of pairs of individuals which differ by i sites, where $i = 0, 1, \dots, n$. Characteristic distributions of nucleotide (or restriction) site differences between pairs of individuals appear in those histograms. A wave that travels to the right, traversing one unit of the horizontal axis in each $1 / 2u$ generations is produced by an episode of growth, where u is the mutation rate. Based on the size of population, the vertical intercepts of distribution is generated, where “The larger the increase in population size, the smaller will be the distribution’s vertical intercept” (Rogers and Harpending, 1992). Generated waves of both implication of continued expansion and a sudden burst of population growth are same. Besides, bottlenecks in population size are almost indistinguishable from those produced by a sudden expansion. The sole difference is elevated upper tail probabilities. On the other hand, reductions in population size initially generate L-shaped distributions. The pairwise nucleotide (mismatch) differences graphs between haplotypes were plotted for each breed and also for total data with the software DNASP v4.0 (Rozas *et al.*, 2003).

3. RESULTS AND DISCUSSION

3.1. PCR Analysis And Data Collection

The primers used amplified 478 bp D-loop region of mtDNA as a maximum. The chromatograms of samples were firstly analyzed with Beckman Coulter 8800 software system. Figure 3.1 shows an example of one of the sequenced samples. After alignment of the whole data set and trimming, 319 bp region of each individual was used for the subsequent analyses.



Figure 3.1. The chromatogram and its sequence from one sample

3.2. Descriptive Data Analysis and Within Population Differentiation

114 haplotypes were observed in 317 horse samples, and 319 bp sequences gave 56 polymorphic sites in which there were ten singleton and 46 parsimony informative sites. Total numbers of mutations were 57. List of samples for each haplotype is given in Table 3.1.

Table 3.1. List of name samples grouped in haplotypes and frequencies of haplotypes

Name of Haplotype	Frequency	Name of Samples
Hap_1	16	AP1 AP11 AP45 AP48 CNK1 CNK49 CNK51 CNK55 CNK58 HH6 HH32 HH38 HH39 HH40 HH57 IGD3
Hap_2	5	AP2 AP47 IGD12 CKR23 CKR26
Hap_3	12	AP3 AP5 AP23 AP46 CNK40 ERZ18 HH1 HH11 HH41 AGR3 AGR11 CKR58
Hap_4	3	AP4 AP12 HH53
Hap_5	9	AP6 AP14 CNK47 ERZ12 HH18 HH22 HH58 MLK12 CKR39
Hap_6	1	AP7
Hap_7	2	AP8 HH28
Hap_8	5	AP9 AP10 AP16 AP32 HH2
Hap_9	4	AP13 AP28 AP41 CNK46
Hap_10	6	AP17 AP49 KRSKIR4 CKR55 CKR56 CKR57
Hap_11	8	AP18 AP25 AP34 AP37 AP38 CNK27 MLK27 CKR37
Hap_12	3	AP19 AP39 AP44
Hap_13	9	AP20 HH24 HH37 HH43 MLK32 IGD7 AGR6 KRSKIR2 CKR47
Hap_14	1	AP21
Hap_15	2	AP22 AGR4
Hap_16	2	AP26 AP33
Hap_17	11	AP27 CNK19 CNK22 ERZ44 HH8 HH30 HH44 HH48 MLK20 CKR6 CKR32
Hap_18	16	AP29 AP35 AP36 CNK6 CNK54 CNK63 HH52 MLK4 MLK17 IGD13 KRSKULA1 KRSYAGIZ3 CKR2 CKR16 CKR18 CKR44
Hap_19	11	AP30 CNK11 CNK59 HH56 MLK7 IGD14 AGR5 AGR13 KRSDORU6 CKR31 CKR52
Hap_20	13	AP31 AP40 AP43 AP50 HH46 MLK21 KRSKIR5 CKR7 CKR12 CKR15 CKR28 CKR30 CKR38
Hap_21	1	AP42
Hap_22	2	CNK2 CNK29
Hap_23	2	CNK4 CNK28
Hap_24	2	CNK9 CKR42
Hap_25	6	CNK10 CNK26 CNK35 CNK42 CNK56 ERZ14
Hap_26	3	CNK12 HH49 MLK14
Hap_27	3	CNK13 CNK48 HH36
Hap_28	2	CNK14 MLK1
Hap_29	5	CNK17 CNK52 CNK53 CNK57 CNK61
Hap_30	1	CNK18
Hap_31	5	CNK20 CNK24 CNK31 CNK41 HH25
Hap_32	1	CNK21
Hap_33	3	CNK23 ERZ15 MLK2

Name of Haplotype	Frequency	Name of Samples
Hap_34	1	CNK25
Hap_35	1	CNK30
Hap_36	3	CNK32 CNK34 ARD3
Hap_37	2	CNK33 HH7
Hap_38	1	CNK36
Hap_39	2	CNK39 CNK62
Hap_40	2	CNK43 HH55
Hap_41	2	CNK45 MLK3
Hap_42	2	CNK50 HH31
Hap_43	1	CNK60
Hap_44	1	CNK65
Hap_45	1	ERZ2
Hap_46	1	ERZ4
Hap_47	1	ERZ6
Hap_48	1	ERZ9
Hap_49	1	ERZ10
Hap_50	9	ERZ13 ERZ37 HH15 HH35 MLK16 MLK18 IGD4 AGR2 CKR4
Hap_51	2	ERZ16 HH16
Hap_52	1	ERZ31
Hap_53	2	ERZ34 HH12
Hap_54	1	ERZ39
Hap_55	1	ERZ46
Hap_56	7	HH3 MLK13 MLK29 AGR10 AGR12 KRSKIR3 CKR1
Hap_57	2	HH4 CKR11
Hap_58	2	HH5 CKR8
Hap_59	1	HH9
Hap_60	3	HH10 VAN14 CKR35
Hap_61	1	HH13
Hap_62	2	HH14 HH20
Hap_63	2	HH17 CKR10
Hap_64	3	HH19 HH34 HH42
Hap_65	1	HH21
Hap_66	3	HH23 HH27 KRSDORU3
Hap_67	1	HH26
Hap_68	2	HH29 CKR49
Hap_69	1	HH33
Hap_70	2	HH45 HH50
Hap_71	2	HH47 IGD5
Hap_72	5	HH51 IGD15 AGR7 AGR8 CKR20
Hap_73	1	HH54
Hap_74	1	HH59
Hap_75	1	HH60

Name of Haplotype	Frequency	Name of Samples
Hap_76	5	MLK5 MLK22 MLK23 MLK24 MLK31
Hap_77	1	MLK6
Hap_78	4	MLK8 MLK10 MLK26 KRSDORU1
Hap_79	7	MLK9 MLK28 MLK30 IGD11 ARD2 ARD4 KRSDORU5
Hap_80	1	MLK11
Hap_81	1	MLK15
Hap_82	1	MLK19
Hap_83	1	MLK25
Hap_84	2	IGD1 IGD9
Hap_85	1	IGD2
Hap_86	3	IGD6 AGR14 KRSYAGIZ1
Hap_87	2	IGD8 AGR9
Hap_88	1	IGD10
Hap_89	1	ARD1
Hap_90	1	KRSDORU2
Hap_91	1	KRSDORU4
Hap_92	2	KRSKIR6 CKR13
Hap_93	1	CKR3
Hap_94	1	CKR5
Hap_95	2	CKR9 CKR41
Hap_96	2	CKR14 CKR48
Hap_97	1	CKR17
Hap_98	1	CKR19
Hap_99	1	CKR21
Hap_100	1	CKR22
Hap_101	1	CKR24
Hap_102	2	CKR25 CKR51
Hap_103	1	CKR27
Hap_104	1	CKR29
Hap_105	1	CKR33
Hap_106	1	CKR40
Hap_107	1	CKR43
Hap_108	1	CKR45
Hap_109	1	CKR46
Hap_110	1	CKR50
Hap_111	1	CKR53
Hap_112	1	CKR54
Hap_113	1	CKR59
Hap_114	1	CKR60

The most frequent haplotype (Hap_1) appeared in sixteen of 317 individuals (5 %). Hap_1, Hap_5, Hap_13, Hap_11, Hap_10, Hap_3, Hap_2, Hap_17, Hap_18, Hap_19, Hap_20, Hap_50, Hap_33, Hap_56 and Hap_72 were shared between different members of breeds. Common founder lineages might be a possible reason behind shared haplotypes among populations (Aberle *et al.*, 2007). Descriptive statistics for the different breeds are shown in Table 3.2. The base sequences of each haplotype can be found in the Appendix A. Haplotype list showing variable nucleotides in a 319 bp fragment of the mitochondrial DNA D-loop is given in Appendix B. The positions of base sequences are given with respect to the whole D-loop as described by Xu and Arnason (1994) in the GenBank accession X79547.

The presence of 114 mtDNA haplotypes in 319 bp fragment in 317 individuals indicates the broad genetic base of founder mares of Turkish native horse breeds' maternal lines (Kavar *et al.*, 2002). The observed result is consistent with earlier horse mtDNA studies. For instance, Hill *et al.* (2002) found 19 haplotypes with 39 polymorphic sites in a 381 bp mtDNA D-loop fragment in 19 female thoroughbred families, which is 10.2 % of the total DNA sequence analyzed. Besides Yang *et al.* (2002) found about 6 % polymorphic sites defining 17 haplotypes in 65 Cheju Horses, whereas Aberle *et al.* (2007) observed 24 mtDNA haplotypes with a total of 10 % polymorphic sites in a 610 bp fragment in 27 horses belonging to nine different heavy horse breeds. The multiple origins of the breeds bred in Turkey may be the reason behind the wide variability of the D-loop sequences among Turkish native horse population. On the other hand, Arabian horses, which comprise a remarkably large population, have only 29 distinct haplotypes (Bowling *et al.*, 2000). Cothran *et al.* (2005) explains the observed differences in diversity being due to the independent development of the maternal lines at several studs from the founder mares originating from many different breeds, and showed high degree of genetic variability.

Haplotypes of TNHB were distributed in the different clades consistent with previous studies (Jansen *et al.*, 2002; Mirol *et al.*, 2002; Cozzi *et al.*, 2004). Several distinct populations participated in the horse domestication process and mtDNA clusters were generated by spreading out of domesticated horses and subsequent incorporation of wild mares (Jansen *et al.*, 2002). Therefore, the modern horse mtDNA sequences do not define

monophyletic groups. The observed differences in TNHB are probably the result of a very ancient mitochondrial DNA diversity. As a conclusion, the study thus supports the multiple origins scenario in the maternal lineage of domestic horse breeds indicated by other researchers (Kavar *et al.*, 1999, 2002; Kim *et al.*, 1999; Bowling *et al.*, 2000; Hill *et al.*, 2002; Yang *et al.*, 2002).

Table 3.2. Population diversity parameters

Population Name	Nucleotide Diversity π	Haplotype diversity H_d	Number of sequences	Number of polymorphic sites	Total number of mutation	Average number of nucleotide differences
AP	0,02494	0,957	48	32	33	7,905
CNK	0,01989	0,968	55	34	34	6,265
CKR	0,02426	0,982	58	40	40	7,689
HH	0,02373	0,979	60	40	40	7,290
ERZ	0,02076	0,993	17	26	26	6,382
MLK	0,02337	0,962	32	37	37	7,161
EAST ANATOLIAN	0,02426	0,971	47	35	35	7,665
TOTAL	0,02252	0,9830	317	56	57	7,079

Turkish native horse breeds have high haplotype diversity (H_d) ranging from 0,957 to 0,993. Haplotype addition from different source of horse populations, and combination of reduced effect of random genetic drift and relaxed selection pressure as a result of rapid population growth would account for the observed high mtDNA haplotype diversities of breeds. HH and CKR breeds had 40 polymorphic sites over 56 sites, and AP had the fewest polymorphic regions.

When compared to haplotype diversity, Turkish native horse breeds have low nucleotide diversity. This is probably due to a small historical effective population size. Only AP, MLK and EAST ANATOLIAN horses have higher nucleotide diversity values. The least number of nucleotide differences is found in the CNK breeds. Combined with the high haplotype diversity, the finding of low nucleotide diversity suggests that most of the genetic diversity is caused by a small number of mutations.

3.3. Between Population Differentiation

Table 3.3 provides the average number of (Dxy) and the number of net (Da) nucleotide substitution per site values between populations. The below diagonal indicates the Dxy and the above diagonal shows the Da values. The highest Dxy value was observed between MLK and AP breeds (0,02495) and the lowest one was found between ERZ and CNK breeds (0,01962). The highest Da value is between MLK and AP breeds (0,00111) and the lowest one is examined between HH and ERZ (-0,00035), which is probably due to the breeds having been sampled from the same location.

Table 3.3. The average number of (Dxy) and the number of net (Da) nucleotide substitution per site, lower and upper diagonals respectively, between the populations

	AP	CNK	CKR	HH	ERZ	MLK	EAST ANATOLIAN
AP		0,00055	0,00002	0,00023	0,00048	0,00111	0,00027
CNK	0,02244		0,00018	0,00016	0,00027	0,00038	0,00053
CKR	0,02462	0,02187		0,00005	0,00018	0,00065	0,00030
HH	0,02419	0,02127	0,02358		0,00035	0,00004	-0,00008
ERZ	0,02308	0,01962	0,02208	0,02128		0,00013	0,00022
MLK	0,02495	0,02141	0,02415	0,02290	0,02130		0,00012
EAST ANATOLIAN	0,02491	0,02210	0,02459	0,02358	0,02245	0,02358	

The pairwise nucleotide (mismatch) differences between haplotypes were plotted for each population and the total data set to see if statistical support for population expansion existed. The expected and observed mismatch distribution plots are given for the whole data set in Figure 3.2 and Figure 3.3. The plots for each breed are given in the Appendix C.

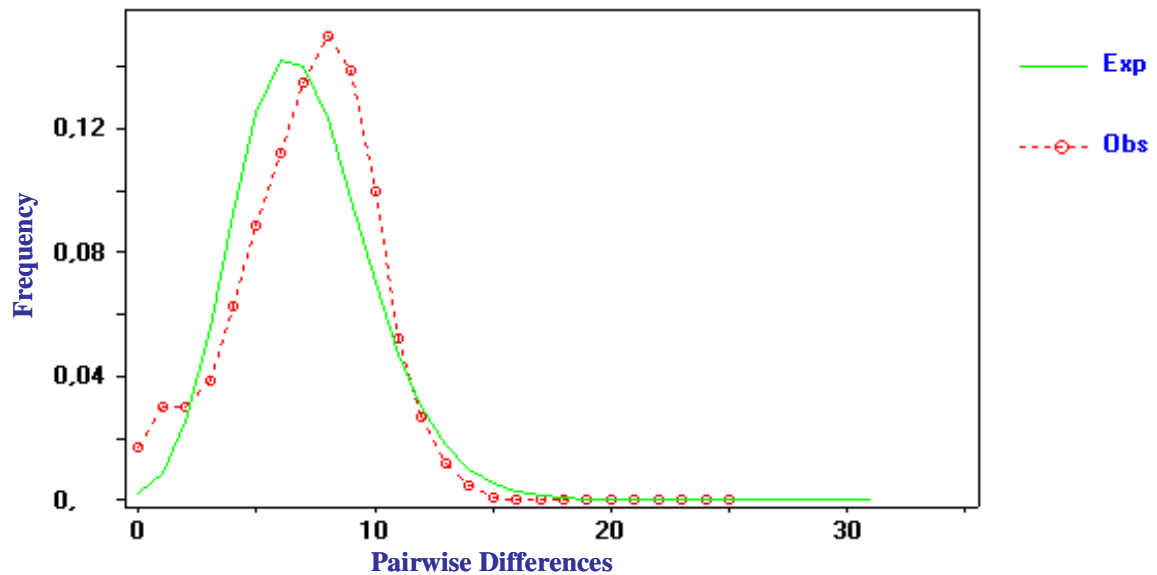


Figure 3.2. The mismatch graph of all samples based on the theory that a population expansion has occurred

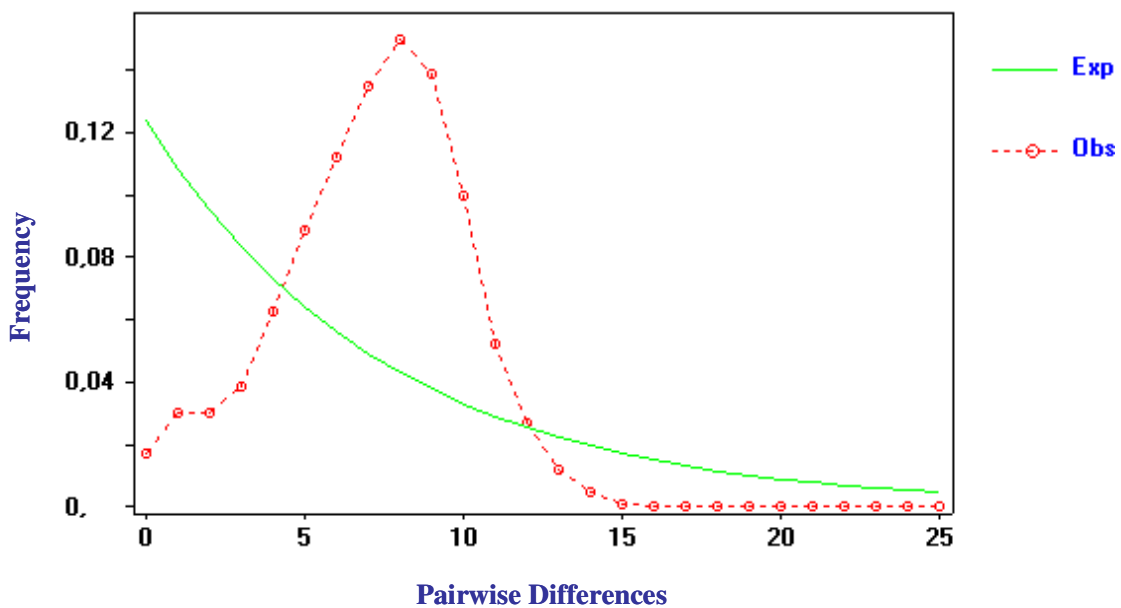


Figure 3.3. The mismatch graph of all samples based on the theory of a constant-sized population

Table 3.4 indicates the results of Fu's, Tajima's, Ramos-Onsis and Rozas's tests for neutrality. Even though, the presence of evidence for population expansions seemed to be construed from the mismatch distribution graphs in certain populations, the values of the three neutrality tests mentioned were not statistically significant; they were within the Confidence Interval of the estimated values from simulated data. In other words, Fu's,

Tajima's, Ramos-Onsis and Rozas's tests for neutrality failed to reject the presence the population expansion for the whole data set. This data also suggests that there might be occurred a selection for the horse breeds, for the investigated genetic region.

Table 3.4. Fu's, Tajima's, Ramos-Onsis and Rozas's Neutrality Test values

Fu's	Tajima's	Ramos-Onsis and Rozas's
(Min; Max) (P; Fs)	(Min; Max) (P; Fs)	(Min; Max) (P; Fs)
-3,57412 ; 5,22233 0,42900 ; 0,42765	-1,73189 ; 1,68608 0,54000 ; -0,7943	0,10379 ; -0,23290 0,00000 ; 0,16113
-3,43102 ; 4,63422 0,45400 ; 0,26587	-1,71310 ; 1,61098 0,53300 ; -0,06388	0,10417 ; 0,23281 0,00000 ; 0,16190
-3,86586 ; 4,88867 0,43100 ; 0,39819	-1,73885 ; 1,71699 0,55400 ; -0,12609	0,10179 ; 0,23320 0,00000 ; 0,16222
-3,69296 ; 4,08385 0,44100 ; 0,26336	-1,73871 ; 1,67205 0,51700 ; -0,03334	0,10375 ; 0,22912 0,00000 ; 0,16032
-3,41421 ; 4,82100 0,45000 ; 0,35814	-1,66908 ; 1,68625 0,54500 ; -0,8124	0,09923 ; 0,23676 0,00000 ; 0,16195
-3,55415 ; 4,97472 0,45900 ; 0,29798	-1,68283 ; 1,70314 0,52800 ; -0,03987	0,10318 ; 0,23096 0,00000 ; 0,16173
-3,71319 ; 4,89963 0,44400 ; 0,33814	-1,65792 ; 1,72067 0,54200 ; -0,03563	0,10555 ; 0,23308 0,00000 ; 0,16183

According to the pairwise ϕ_{st} values between each population indicated in Table 3.5, there are significant genetic differences between AP and CNK, AP and MLK, KRS and CNK. The population pairwise ϕ_{st} values for the other breeds are low and non-significant. Hence, there is no significant genetic differentiation found between the populations, apart from those mentioned above.

Table 3.5. Population pairwise ϕ_{st} values between mitochondrial sequences of breeds are below diagonal. The probabilities of the estimated ϕ_{st} values are given above diagonal

	AP	CNK	HH	ERZ	MLK	EAST ANATOLIAN	CKR
AP	0.00000	0.00901	N.S.	N.S.	0.02703	N.S.	N.S.
CNK	0.03091*	0.00000	N.S.	N.S.	N.S.	0.00901	N.S.
HH	0.00948	0.00992	0.00000	N.S.	N.S.	N.S.	N.S.
ERZ	0.01945	0.01314	0.01431	0.00000	N.S.	N.S.	N.S.
MLK	0.04351*	0.01817	0.00150	0.00497	0.00000	N.S.	N.S.
EAST ANATOLIAN	0.01076	0.02797*	0.00336	0.00895	0.00466	0.00000	0.04505
CKR	0.00083	0.01122	0.00191	0.00732	0.02627	0.01207*	0.00000

*p<0.05

The AMOVA results (Table 3.6), indicate that total differences are mostly due to the variation within populations (98 %). Only 2.5 % of the total variation is present among groups. This result provides another line of evidence that the Turkish native breeds are not genetically differentiated from each other.

Table 3.6. The result of AMOVA of mitochondrial sequences of the analyzed breeds

Source of Variation	Sum Of Squares	Variance Components	Percentage Variation
Among Groups	30.204	0.09478	2.54807
Among Populations Within Groups	2.259	-0.05359	-1.44078
Within Populations	1140.329	3.67848	98.89271
Total	1172.792	3.71967	

Based on mtDNA control region sequence data, clear separations among studied Turkish native horse breeds were not observed. The most likely reason for this lack of differentiation is the extensively active horse trade in Anatolia as well as the multiple origins in horse domestication. In addition, because of a lack of effective breeding strategies, population pairwise ϕ_{st} values between mitochondrial sequences of breeds do not show significant distinctions among most of the highly distinctive breeds. The low D_{xy} , D_a values and most of the variation (98 %) being observed within populations in the AMOVA analysis also point out to the lack of genetic differentiation among various breeds.

3.4. Phylogenetic Reconstruction and Network Analysis

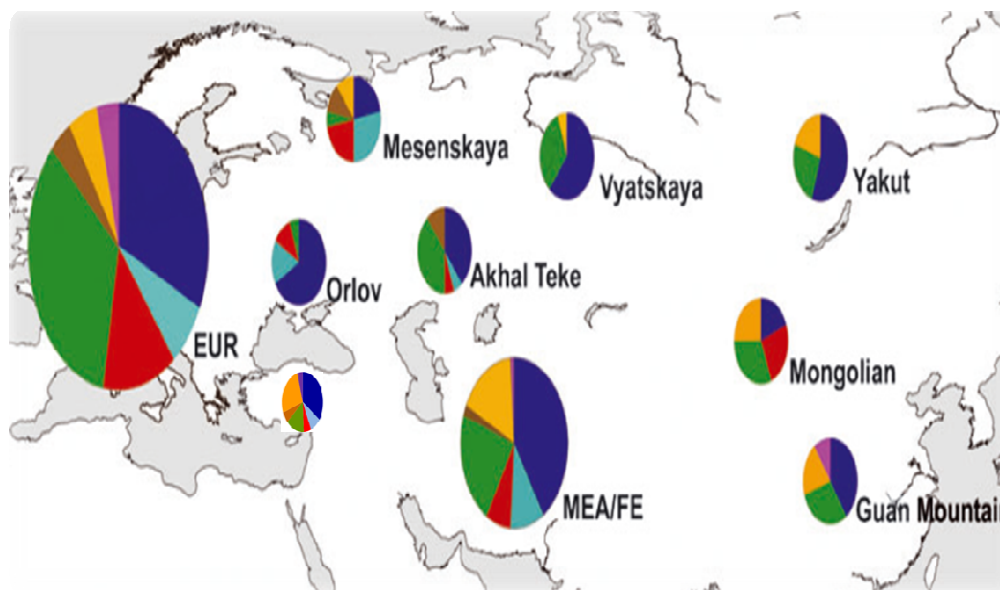


Figure 3.4. Mitochondrial DNA haplogroup distributions in Eurasian horse populations
Haplogroups are colored as follows: dark blue =A, light blue=B, red=C, green=D, brown=E, orange=F and pink=G

Turkish native horse breeds were found to be associated with all the Vila *et al.* (2001) clades, except the poorly represented clades B and E. Additionally, it was observed that, most of the Turkish native horse breeds (TNHB) belonged to the haplogroup A, which is common in Europe and Asia. Interestingly, high numbers of members of TNHB were classified as similar to the Asian and Middle Eastern populations, rather than the European ones. Consistent with other analyzed European and Asian populations, haplogroup E was observed to be rare and poorly represented within TNHB (McGahern *et al.*, 2006). In 2006, McGahern *et al.* (2006) also showed that, haplogroup G is seen rarely within European, Middle Eastern and Far Eastern horse populations and concordantly in this study, only a few individuals of TNHB were classified into haplogroup G.

In addition, TNHB horses possessed many exclusive, low-frequency mitochondrial haplotypes not shared with any other group, namely, HAP_49, HAP_83, HAP_1, HAP_32 and HAP_108. Since these haplotypes may represent the maternal lineages which have not

been analyzed yet, we propose a new classification for these haplotypes within haplogroup H in an addition to the clades A-G (Vila *et al.*, 2001).

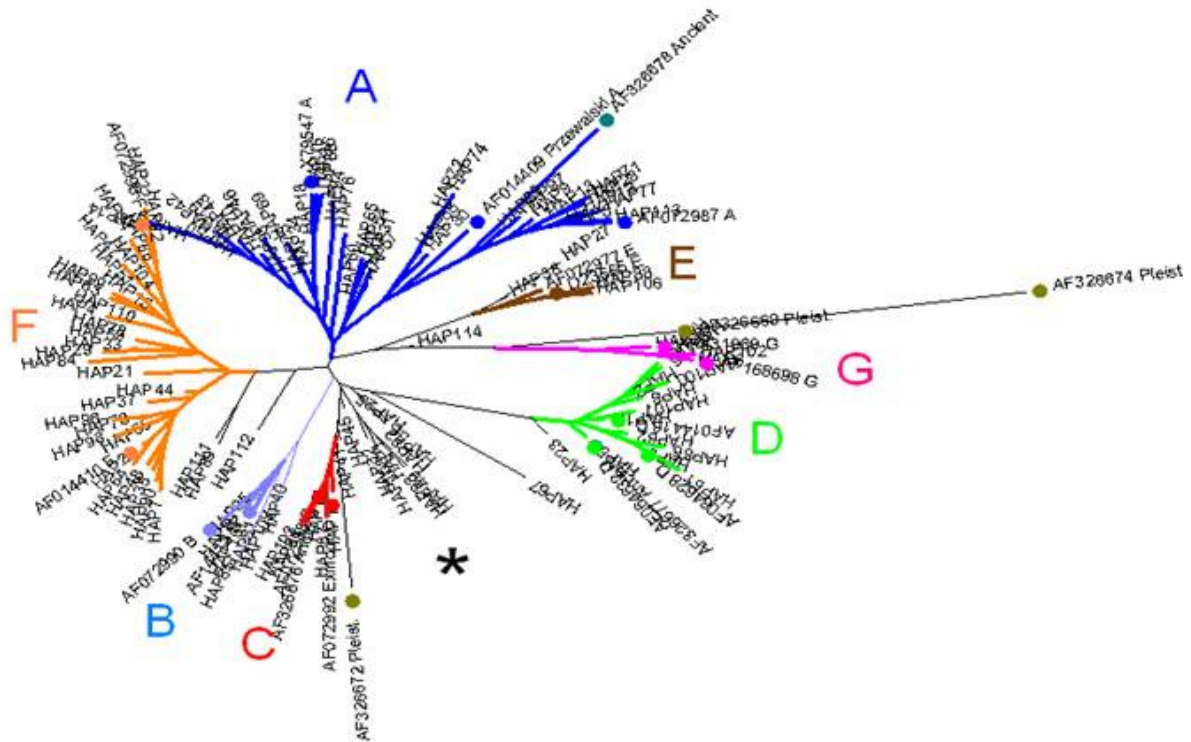


Figure 3.5. Mitochondrial DNA haplogroup distributions among Turkish native horse breeds and representative Genbank sequences. Haplogroups are colored as follows: Red=A, Green=B, Yellow=C, Blue=D, Purple=E, Turquoise=F, Grey=G, Teal=Ancient Samples, Olive= Pleistocene samples

Consistent with Vila *et al.*'s (2001) findings, ancient specimens and Przewalski's horses clustered together with the modern Turkish native horse breeds, due to the high similarity in their sequences. This contrasts with the fact that a high mutation rate is the major reason of high diversity observed in modern horses (Vila *et al.*, 2001). In addition, ancient horse sequences obtained from Genbank clustered in the same group. Therefore, since Turkish native horse sequences would not be defined as monophyletic group with respect to an ancient progenitor, they probably did not originate from the same selected wild stock (Vila *et al.*, 1997; Guiffra *et al.*, 2000). As the diversity of natural populations might have been limited in single natural populations, sequences obtained from the late Pleistocene horses were observed in different groups. Consequently the observed high

matrilineal diversity among Turkish native horses is probably due to the wild horses from a number of different populations being used as founders during domestication processes.

There are two alternative hypotheses for the origin of the domestic horses from wild populations. According to the “restricted origin hypothesis”, domestic horse was firstly developed as a result of complex multi-generation selection on traits breeding, and subsequently distributed to other regions (Vila *et al.*, 2001). Alternatively, domestication process could have started independently at different times and in different regions of the Eurasian range, with involvement of a large number of founders. In this “multiple origins scenario”, since horses would be captured from diverse wild populations and bred in captivity, the effects of selective breeding may not be observed in the early domestic horses. The main difference between these two hypotheses is related to the possible predictions for the origin of the domestic horse on the basis of genetic variation in maternally inherited mtDNA. The former hypothesis explains the diversity of the horse resulted in a few founding lineages and those added subsequently by mutations. On the other hand, a multiple origins hypothesis predicts greater diversity and divergence among populations as a consequence of domestication (Vila *et al.*, 2001).

In addition to the Turkish native horse breeds sampled from different regions of Anatolia, representative Genbank sequences from distinctive geographical regions of the world were also added to analyses. The clustering of geographically divergent sequences in the same group is consistent with the multiple origins scenario, in which the wild founders of domesticated horses would have been recruited from geographically different areas, unlike many other large domesticated animal species which exhibit clear geographic patterns (MacHugh and Bradley, 2001). Domestication of dogs, cattle, sheep, and goats took place several thousand years before the horse (Clutton-Brock, 1999), and in these species fewer founders were present, which were domesticated more than once (Loftus *et al.*, 1994, Bradley *et al.*, 1996, Troy *et al.*, 2001). The geographical spread of these species indicates a few number of domestication centers (Vila *et al.*, 2001). In the light of this observation, horses were domesticated from several wild populations (Lister *et al.*, 1998; Vila *et al.*, 2001), which individually had lower levels of genetic diversity (Jansen *et al.*, 2002).

In the neighbor-joining tree constructed in this study, some haplogroups included representatives from different horse breed types, whereas sometimes haplotypes from the same breed were present in separate haplogroups due to a lack of informative sites in the control region. This same pattern has been seen in other studies of horse mtDNA (Vila *et al.*, 2001; Jansen *et al.*, 2002). In contrast, on the basis of blood and protein typing data, horse breeds clustered corresponding to their known ancestry (Juras *et al.*, 2003). High rate of mutation in which latter mutations neutralize the effect of former mutations by altering back to the original bases (back mutations), or migration of individuals between populations at the initial stage of breed formation are likely reasons behind the lack of breed-related patterns.

Previous studies indicated that there have been only seven haplogroups in horse populations ranked A-G (Vila *et al.*, 2001; Jansen *et al.*, 2002) however, clustering of HAP_49, HAP_83, HAP_1, HAP_32 and HAP_108 in a independent branch of neighbor joining tree may be accepted as a tangible evidence of presence of the eighth haplogroup in the global horse populations. The finding of this unique eighth haplogroup, which we designate as “H”, comprises one of the significant findings of this study. However, this must be checked with the global data.

A median joining network was also generated by using same reference sequences to see whether the clade classification observed in neighbor joining tree was maintained (Figure 3.6). However, median joining network did not provide as much of an obvious resolution within clades, as neighbor joining tree did.

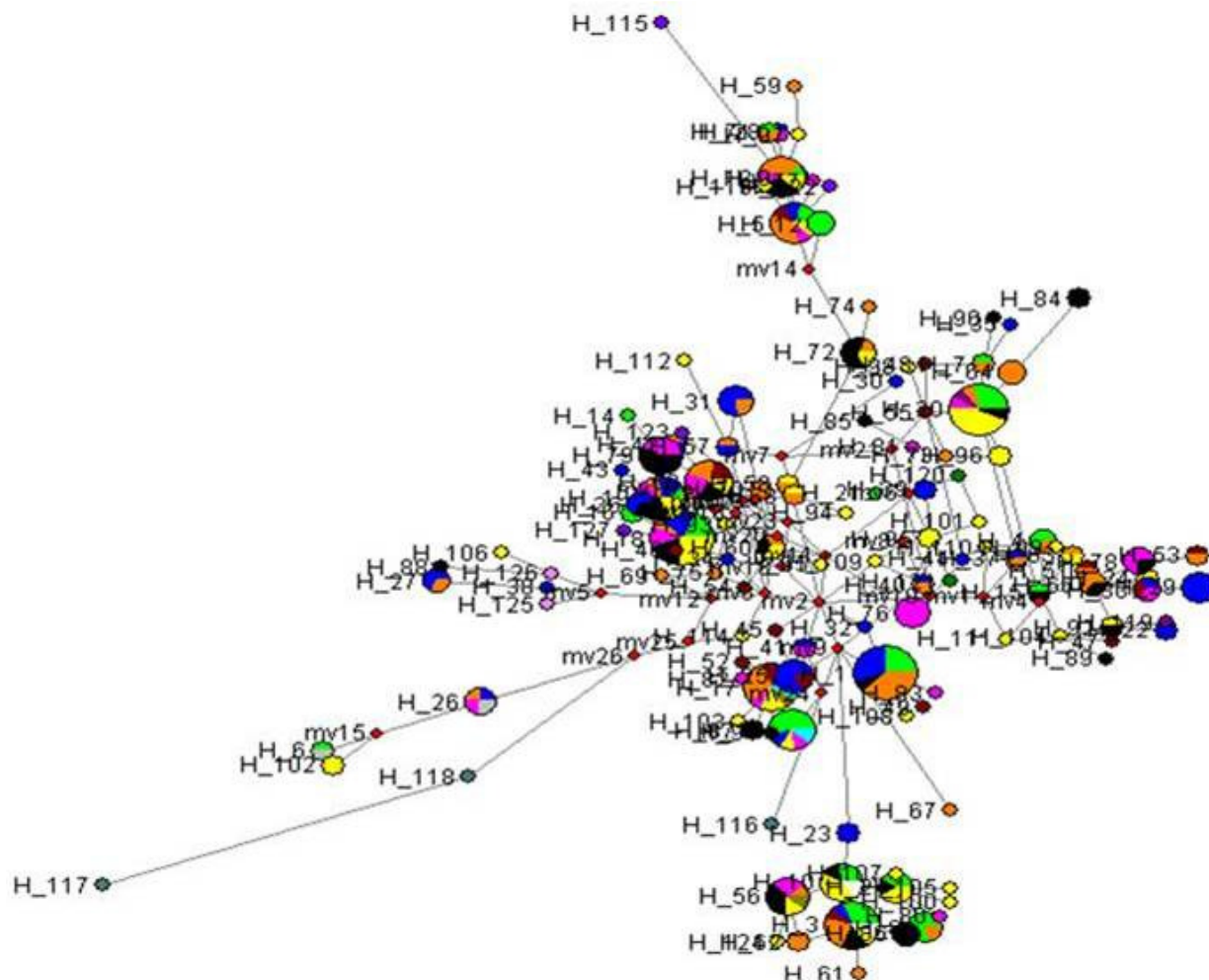


Figure 3.6. Median network of Turkish horse haplotypes including 22 GenBank sequences.

Haplotypes were visualized with pie-charts and the size of charts is proportional to the number of individuals showing a particular haplotype. One single mutation separates nodes connected by a straight segment.

Each breed has a unique color as follows: Green=AP,

Blue=CNK, Pink=MLK, Orange=HH, Black= EAST ANATOLIAN, Yellow=CKR,

Brown=ERZ, Purple=A, Dark Green=B, Turquoise=C, White=D, Light Pink=E, Dark

Blue=F

4. CONCLUSIONS AND RECOMMENDATIONS

As mentioned before, few numbers of mutations had occurred in horse mtDNA control region after horse domestication. Because sequences of ancient and Przewalski's horses were clustered together with the modern Turkish native horse breeds, it could be concluded that mtDNA diversification of Turkish native horse breeds is thought to have occurred prior to domestication.

Clustering of geographically divergent TNHB haplotypes in the same group and distribution of haplotypes of a same particular breed in different groups are consistent with the multiple origins scenario, which points out to the recruitment of wild founders of domesticated horses to be the reason of observed lack of demographic separation.

Back mutations and high rate of migration among Turkish native horse breeds have remarkably shaped the patterns that are related to breeds. In addition, because of active horse trading habits, and lack of effective horse breeding strategies, there is no clear separation among Turkish native horse breeds.

Because Anatolia had been a passage way of main trade roads and horses had been employed widely in transportation, all seven globally recognized haplogroups have been observed among TNHB. However, some haplotypes namely HAP_49, HAP_83, HAP_1, HAP_32 and HAP_108 were grouped in a in an independent branch of neighbor joining tree. These findings may be accepted to be a evidence of the eighth haplogroup.

In the future, cost effective and sustainable breed conservation programs and breeding strategies should be undertaken for the TNHB. Conservation program and breeding studies should employ evaluation of genetic diversity and relationships among the livestock breeds (Behl *et al.*, 2006). Turkish horse subpopulations would be classified into cold blood horse group, which belongs to a breed known for having a stable, calm temperament. Sampling of other cold blood horse lineages geographically originated close to Anatolia and reanalyzing whole data set with the addition of those could support and clarify some details about the evolutionary history of Turkish horse livestock breeds.

mtDNA is maternally inherited and it provides information only on maternal lineages. Moreover, as it does not recombine and it is inherited as a single locus, different genetic markers should also be used for setting up proper conservation and management strategies. mtDNA diversity alone does not provide enough information. In addition, genetic characterization is necessary for the management of genetic resources and preservation of the breed unity (Bjornstad and Roed, 2002). Since the microsatellite DNA markers have highly polymorphic nature and are abundant throughout most eukaryotic nuclear genomes, and they have been widely used in the analysis of genetic diversity among breeds of various livestock species including horses (Bjornstad *et al.*, 2000; Canon *et al.*, 2000; Kelly *et al.*, 2002; Tozaki *et al.*, 2003; Achmann *et al.*, 2004; Aberle *et al.*, 2004; Solis *et al.*, 2005), and in the characterization of horse populations and their evolutionary relationships (Tozaki *et al.*, 2003; Aberle *et al.*, 2004; Achmann *et al.*, 2004; Glowatzki-Mullis *et al.*, 2005; Iamartino *et al.*, 2005; Solis *et al.*, 2005; Szontagh *et al.*, 2005; Vega-Pla *et al.*, 2005, 2006; Zabeck *et al.*, 2005; Lee and Cho, 2006; Moodley *et al.*, 2006). A subsequent study with effective microsatellite markers to observe genetic variability within and between Turkish horse subpopulations can be undertaken as a follow-up study. In addition, the Y chromosome, except for its pseudo autosomal region which is transmitted paternally as a linkage group, is used to investigate male-specific histories of populations. A study based on Y-chromosomal sequence variation to investigate the genetic relationship among the TNHB paternal lines would also contribute to our understanding of the evolutionary processes within the TNHB.

REFERENCES

- Aberle, K.S., Hamann, H., Drogemuller, C., Distl, O., 2007. Phylogenetic relationships of German heavy draught horse breeds inferred from mitochondrial DNA D-loop variation. *Journal of Animal Breeding Genetics*, 124, 94-100.
- Achmann, R., Curik, P., Dovc, P., Kavar, T., Bodo, I., Habe, F., Marti, E., Solkner, J., Brem, G., 2004. Microsatellite diversity, population subdivision and gene flow in Lipizzan horse. *Animal Genetics*, 35, 385-392.
- Anonym, 2006 <http://www.tuik.gov.tr/VeriBilgi.do>
- Arpacık, R., 1994. Horse breeding, Şahin Press, Ankara.
- Bandelt, H.J., Quintana-Murci, L., Salas, A., Macaulay, V., 2002. The fingerprint of phantom mutations in mtDNA data. *American Journal of Human Genetics*, 71, 1150-1160.
- Bar, Y.O., Meadow, R.H., 1995. The origins of agriculture in the Near East, Price TD, Gebauer (Eds). *School of American Research Press*, Santa Fe, 39-94.
- Batu, S., 1938. Turk horses and horse breeding knowledge. Ankara Yüksek Ziraat Enstitüsü, Lecture Book, Number: 3, Ankara.
- Behl, R., Behl, J., Gupta, N., Gupta, S.C., Ahlawat, S.P.S., Ragnekar, M., Ahmed, Z., 2006. Genetic characterization of Zanskari breed of horse. *Journal of Genetics*, 85/3 199-203.
- Bjornstad, G., Gunby, E., Roed, K.H., 2000. Genetic structure of Norwegian horse breeds. *Animal Breeding Genetics*, 117, 307-317.

Bjornstad, G., Roed, K.H., 2002. Evaluation of factors affecting individual assignment precision using microsatellite data from horse breeds and simulated breed crosses. *Animal Genetics*, 33, 264–270.

Bowling, A.T., Del Valle, A., Bowling, M., 1998. Verification of horse maternal lineage based on derived mitochondrial DNA sequence. *Journal of Animal Breeding and Genetics*, 115, 351-357.

Bowling, A.T., 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.

Bradley, D.G., MacHugh, D.E., Cunningham, P., Loftus, R.T., 1996. Mitochondrial diversity and the origins of African and European cattle. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 5131-5136.

Bruford, M.W., Bradley, D.G., Luikart, G., 2003. DNA markers reveal the complexity of livestock domestication. *Nature Reviews Genetics*, 4, 900-910.

Brown, D.R., Anthony, D.W., 1998. Bit wear, horseback riding, and the Botai site in Kazakhstan. *Journal of Archaeological Science*, 25, 331–347.

Cai, D.W., Han, L., Xie, C.Z., Li, S.N., Zhou, H., Zhu, H., 2007. Mitochondrial DNA analysis of Bronze Age horses recovered from Chifeng region 'Inner Mongolia' China. *Progress in Natural Science*, 17, 544–550.

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

Charlesworth, D., Charlesworth, B., 1987. Inbreeding Depression and its Evolutionary Consequences. *Annual Review of Ecology and Systematics*, 18, 237-268.

Clutton-Brock, J.A., 1999. *Natural History of Domesticated Mammals*. Cambridge Univ. Press, Cambridge, London.

Cothran, E.G., Juras, R., Macijauskiene, V., 2005. Mitochondrial DNA D-loop sequence variation among 5 maternal lines of the Zemaitukai horse breed. *Genetics and Molecular Biology*, 284, 677-681.

Cozzi, M.C., Strillacci, M.G., Valiati, P., Bighignoli, B., Cancedda, M., Zanotti, M., 2004. Mitochondrial D-loop sequence variation among Italian horse breeds. *Genetic Selection Evolution*, 36, 663-672.

Di Bernardo, G., Galderisi, U., Del Gaudio, S., D'Aniello, A., Lanave, C., De Robertis, M.T., Cascino, A., Cipollaro, M., 2004. Genetic characterization of Pompeii and Herculaneum Equidae buried by Vesuvius in 79 AD. *Journal of Cellular Physiology*, 199, 200–205.

Excoffier, L.G., Laval, G., Schneider, S., 2005. Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, 1, 47-50.

Fu, Y.X., Li, W.H., 1993. Statistical tests of neutrality of mutations. *Genetics*, 133, 693-709.

Furman, A., Yenigün, O., 1998. *The Environmental Dimension: Introduction to Ecology and Environmental Problems*. Boğaziçi University Press, Istanbul.

Giuffra, E., Kijas, J.M., Amarger, V., Carlborg, Ø., Jeon, J. T., Andersson, L., 2000. The origin of the domestic pig: Independent domestication and subsequent introgression. *Genetics*, 154, 1785–1791.

Glowatzki-Mullis, M.L., Muntwyler, J., Pfister, W., Marti, E., Rieder, E., Poncet, P.A., Gaillard, C., 2005. Genetic diversity among horse populations with a special focus on the Franches-Montagnes breed. *Animal Genetics*, 37, 33–39.

Graur, D., Li, W. H., 2000. *Fundamentals of Molecular Evolution*; Second Edition. Sinauer Associates, Sunderland, M.A., 39-65.

Hall, T.A., 1999. "BioEdit: a user-friendly biological sequence alignment editor and analysis" <http://www.mbio.ncsu.edu/BioEdit/bioedit.html>

Helmer, D., Gourichon, L., Monchot, H., Peters, J., Saña Segui, M., 2005. Identifying early domestic cattle from prepottery Neolithic sites on the middle Euphrates using sexual dimorphism. In Vigne J.D., Peters J., Helmer D. (Eds), *The First Steps of Animal Domestication*, 86-95, Oxbow Books, Oxford.

Hill, E.W., Bradley, D.G., Al-Barody, M., Ertgrul, O., Splan, R.K., Zakharov, I., Cunningham, E.P., 2002. History and integrity of thoroughbred dam lines revealed in equine mtDNA variation. *Animal Genetics*, 33, 287–94.

Iamartino, D., Fidotti, M., Miraglia, N., Pilla, F., 2005. Genetic characterization of Petro young horses by micosatellite markers. *Conservation Genetics of Endangered Horse Breeds* Wageningen Academic Publishers, Wageningen.

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 prewalskii's wild horse and domestic horse. *Journal of Molecular Evolution*, 41, 190-198.

Ishida, N., Hasegawa, T., Takeda, K., Sakagami, M., Onishi, A., Inumaru, S., Komatsu, M., Mukoyama, H., 1994. Polymorphic sequence in the D-loop region of equine mitochondrial DNA. *Animal Genetics*, 25, 215–221.

Jansen, T., Forster, P., Levine, M.A., Oelke, H., Hurles, M., Renfrew, C., Weber, J., Olek K., 2002. Mitochondrial DNA and the origins of the domestic horse. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 10905–10910.

Jung, Y.H., Han, S.H., Shin, T., Oh, M.Y., 2002. Genetic Characterization of Horse Bone Excavated from the Kwakji Archaeological Site, Jeju, Korea. *Molecular Cells*, 14, 224–230.

Juras, R., Cothran, E.G., Klimas, R., 2003. Genetic analysis of three Lithuanian native horse breeds. *Acta Agriculturae Scandinavica, Section A* 53, 180–185.

Kavar, T., Habe, F., Brem, G., Dove P., 1999. Mitochondrial D-loop sequence variation among the 16 maternal lines of the Lipizzan horse breed. *Animal Genetics*, 30, 423–430.

Kavar, T., Brem, G., Habe, F., Sölkner, J., Dovc, P., 2002. History of Lipizzan horse maternal lines as revealed by mtDNA analysis. *Genetic Selection Evolution*, 34, 635–648.

Kavar, T., Dovc, P., 2008. Domestication of the horse: Genetic relationships between domestic and wild horses. *Livestock Science*, 116, 1–14.

Kelly, L., Postiglioni, A., DeAndres, D.F., Vega-Pla, J.L., Gagliardi, R., Biagetti, R., Franco J., 2002. Genetic characterization of Uruguayan Creole horses and analysis of relationships among horse breeds. *Research in Veterinary Sciences*, 72, 69–73.

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). *Animal Genetics*, 36, 203–209.

Kim, K. I., Yang, Y.-H., Lee, S. S., Park, C., Ma, R., Bouzat, J.L., Lewin, H.A., 1999. Phylogenetic relationships of Cheju horses to other horse breeds as determined by myDNA D-loop sequence polymorphism. *Animal Genetics*, 30, 102–108.

Leakey, R.E., Lewin, R., 1995. *The Sixth Extinction: Patterns of Life and the Future of Humankind*. Anchor Publishing. Port Moody.

Lee, S., Cho, G., 2006. Parentage testing of thoroughbred horse in Korea using microsatellite DNA typing. *Journal of Veterinary Science*, 7, 63–67.

Levine, M.A., 2005. Domestication and early history of the horse. In: Mills D.M., McDonnell S.M. (Eds.), *The Domestic Horse: The Origins, Development, and Management of Its Behaviour*, 5–22, Cambridge University Press, Cambridge.

Levine, M.A., 2006. mtDNA and horse domestication: the archaeologist's cut. In: Mashkour M. (Eds), *Equids in Time and Space. Proc. 9th International Conference of Archaeozoology*, 192–201, Oxbow Books, Oxford.

Lister, A. M., Kadwell, M., Kaagan, L. M., Jordan, W. C., Richards, M. B., Stanley, H.F., 1998. Ancient and modern DNA in a study of horse domestication. *Ancient Biomolecules*, 2, 267–280.

Loftus, R.T., MacHugh, D.E., Bradley, D.G., Sharp, P.M., Cunningham, P., 1994. Evidence for two independent domestications of cattle. *Proceedings of the National Academy of Sciences of the United States of America*, 91, 2757–2761.

Luís, C., Bastos-Silveira, C., Cothran, E.G., Oom, M.M., 2006. Iberian Origins of New World Horse Breeds. *Journal of Heredity*, 97, 107–113.

MacHugh, D.E., Bradley, D.G., 2001. Livestock genetic origins: goats buck the trend. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 5382–5384.

McGahern, A.M., Bower, M.A., Edwards, C.J., Brophy, P.O., Sulimova, G., Zakharov, I., Vizuete-Forster, M., Levine, M., Li S., Machugh, D.E., Hil, E.W. 2006. Evidence for biogeographic patterning of mitochondrial DNA sequences in Eastern horse populations. *Animal Genetics*, 37, 494–497.

Meadow, R.H. 1989. Osteological evidence for the process of animal domestication, (Eds) Clutton-Brock. Unwin Hyman, London, 80–90.

- Mirol, P.M., Garcia, P.P., Vega Pla, J.L., Dulout, F.N., 2002. Phylogenetic relationships of Argentinean Creole horses and other South American and Spanish breeds inferred from mitochondrial DNA sequences. *Animal Genetics*, 33, 356–363.
- Moodley, Y., Baumgarten, I., Harley, E.H., 2006. Horse microsatellites and their amenability to comparative equid genetics. *Animal Genetics*, 37, 258–261.
- Mousset, S., Derome, N., Veuille, M., 2004. A neutrality test based on the mismatch distribution. *Molecular Biology Evolution*, 21, 724–731.
- Nei, M., Li, W. H., 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, 76, 5269–5273.
- Nei, M., Tajima, F. 1981. DNA polymorphism detectable by restriction endonucleases. *Genetics*, 97, 145–163.
- Olsen, S.L., 2006. EarlyHorseDomestication on the Eurasian Steppe. In: Zeder M.A., Bradley D.G., Emshwiller E., Smith B.D. (Eds.), *Documenting Domestication. New Genetic and Archaeological Paradigms*, 245–272, University Presses of California, Princeton.
- Okumura, N., Ishiguro, N., Nakano, M., Sahara, M., 1996. Intra- and interbreed genetic variations of mitochondrial DNA major non-coding regions in Japanese native dog breeds *Canis familiaris*. *Animal Genetics*, 27, 397–405.
- Perez-Gutierrez, L.M., De la Pena, A., Arana, P., 2008. Genetic analysis of the Hispano-Breton heavy horse. *Animal Genetics*, 39, 506–514.
- Peters, J., von den Driesch, A., Helmer, D., Vigne, J.D., Peters J., 2005. The upper Euphrates-Tigris Basin, cradle of agro-pastoralism? In Vigne J.D., Peters, J., Helmer, D., (Eds), *The First Steps of Animal Domestication*, 96–124, Oxbow Books, Oxford.

Ramos-Onsins, S.E., Rozas, J., 2002. Statistical properties of new neutrality tests against population growth. *Molecular Biology Evolution*, 19, 2092–2100.

Rogers, A.R., Harpending, H.C., 1992. Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology Evolution*, 9, 552–569.

Rozas, J., Sanchez-De, I., Barrio, J.C., Messeguer, X., Rozas, R., 2003. DNASP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, 19, 2496–2497.

Salamini, F., Özkan, H., Brandolini, A., Schäfer-Pregl, R., Martin, W., 2002. Genetics and geography of wild cereal domestication in the near east. *Nature Review Genetics*, 3, 429–441.

Sambrook, J.; Fritsch, E.F., Maniatis, T., 1989. *Molecular Cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, New York.

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain terminating inhibitors. *Biochemistry*, 74, 5463–5467.

Solis, A., Jugo, B.M., Meriaux, J.C., Iriando, M., Mazon, L.I., Aguirre, A.I., Vicario, A., Estomba, A., 2005. Genetic diversity within and among four South European native horse breeds based on microsatellite DNA analysis: implications for conservation. *Journal of Heredity*, 96, 670–678.

Stiff, A.R., Capo, R.C., Gardiner, J.B., Olsen, S.L., Rosenmeier, M.F., 2006. Geochemical evidence of possible horse domestication at the copper age Botai settlement of Krasnyi yar, Kazakhstan. The Geological Society of America. 2006 Philadelphia Annual Meeting.

Szontagh, A., Ban, B., Bodo, I., Cothran, E.G., Hecker, W., Jozsa, Cs., Major, A., 2005. Genetic diversity of the Akhal-Teke horse breed in Turkmenistan based on microsatellite analysis. *Conservation Genetics of Endangered Horse Breeds*, Wageningen Academic Publishers, Wageningen.

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123, 585–595.

Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, 24, 1596–1599.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Researches*, 25, 4876–4882.

Tozaki, T., Takezaki, N., Hasegawa, T., Ishida, N., Kurusawa, M., Saitou, N., Mukoyama, H., 2003. Microsatellite variation in Japanese and Asian horses and their phylogenetic relationship using a European horse out group. *Journal of Heredity*, 94, 374–380.

Troy, C.S., MacHugh, D.E., Bailey, J.F., Magee, D.A., Loftus, R.T., Cunningham, P., Chamberlain A.T., Sykes B.C., Bradley D.G., 2001. Genetic evidence for Near-Eastern origins of European cattle. *Nature*, 4106832, 1088–1091.

Ueberman, H.P. 1979. Supplements to the Tuebingen Atlas of the Near Orient, series B, Problems of the Neolithization of the Mediterranean Region, Ludwig Reichert, Weisbaden, 130–145.

Uerpmann, H.P., 1990. Die Domestikation des Pferdes im Chalkolithikum West-und Mitteleuropas. Madrid. Mitt. 31, 109–153.

Vega-Pla, J.L., Calderon, J., Rodriguez-Gallardo, P.P., Alcaide, B., Sereno, F.T.P.S., Costa, M.R., Perez-Pineda, E., Martinez, A.M., Delgado, J.V., Rico, C., 2005. The Retuertas horse: the missing link_in the Iberoamerican horse breeds origin? *Conservation Genetics of Endangered Horse*, Wageningen Academic Publishers, Wageningen.

Vega-Pla, J.L., Calderon, J., Rodriguez-Gallardo, P.P., Martinez, A.M., Rico C., 2006. Saving feral horse populations: does it really matter? A case study of wild horses from Donnana National Park in southern Spain. *Animal Genetics*, 37, 571–578.

Vigne., J.D., Buitenhuis, H., Davis, S., 1999. The first steps in animal domestication into the east of the Euphrates: Cypres and Central Anatolia Paléorient, 49–62.

Vila C., Savolainen P., Maldonado J.E., Amorim I.R., Rice J.E., Honeycutt R.L., Crandall K.A., Lundeberg J., Wayne R.K., 1997. Multiple and ancient origins of the domestic dog. *Science*, 276, 1687–1689.

Vila, C., Leonard, J.A., Gotherstrom, A., Marklund, S., Sandberg, K., Liden, K., Wayne, R.K., Ellegren, H., 2001. Widespread origins of domestic horse lineages. *Science*, 291, 474–477.

Weinstock, J., Willerslev, E., Sher, A., Tong, W., Ho, S.Y., Rubenstein, D., Storer, J., Burns, J., Martin, L., Bravi, C., Prieto, A., Froese, D., Scott, E., Xulong, L., Cooper, A., 2005. Evolution, systematics, and phylogeography of pleistocene horses in the new world: a molecular perspective. *PLoS Biology*, 3, 241

Wolstenholme, D.R., 1992. Animal mitochondrial DNA: structure and evolution. *International Review of Cytology*, 141, 173–216.

Xu, X., Arnason, U., 1994. The complete mitochondrial DNA sequence of the horse, *Equus caballus*: extensive heteroplasmy of the control region. *Genetics*, 148, 357–362.

Yang, Y.H., Kim, K.I., Cothran, E.G., Flannery, A.R., 2002. Genetic diversity of Cheju Horses *Equus caballus* determined by using mitochondrial DNA D-loop polymorphism. *Biochemical Genetics*, 40, 175–186.

Zabeck, T., Nogag, A., Radko, A., Nogag, J., Slota, E., 2005. Genetic variation of Polish endangered Bilgoraj horses and two common horse breeds in microsatellite loci. *Journal of Applied Genetics*, 46, 299–305.

Zeder, M.A., 2006. A critical examination of markers of initial domestication in goats (*Capra hircus*). In Zeder M.A, Emshwiller E., Smith B.D., Bradley D.G. (Eds), Documenting Domestication, 181–208, Univ of California Press, Berkeley.

Zeder, M.A., 2008. Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact. PNAS, 33, 11597–11604.

**APPENDIX A: 319 BP LONG mtDNA SEQUENCES OF 114
DIFFERENT HAPLOTYPES**

>HAP*1

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
 CCCCCACATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
 TACATCGTGCATTAAATTGCTTGCCCCATGAATAATAAGCATGTACATAATATC
 ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
 AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
 CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP2

TAAACGACAACAATCCGCCCTCATGTGCTATGTCAGTATCAGATTATA-
 CCCCCATATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
 TACATCGTGCATTAAATTGTTACCCCATGAATAATAAGCATGTACATAATATC
 ATTTATCTTACATGAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
 AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
 CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP3

TAAACGACAACAATCCGCCCTCATGTGCTATGTCAGTATCAGATTATA-
 CCCCCATATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
 TACATCGTGCATTAAATTGTCCGCCCATGAATAATAAGCATGTACATAATATC
 ATTTATCTTACATGAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
 AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
 CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP4

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
 CCCCCACATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
 TACGTCGTGCATTGAATTGTTTACCCCATGAATAATAAGCATGTACATAATATC
 ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
 AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAGTCAC
 CAAGCCGCGGGAAATCAGCAACCCTCTCAACTGCGTGTCCCAATCCTCGCTC

*Haplotype

>HAP5

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAAACACTATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAGTTGTTTGCCCCATGAATAATAAGCATGTACATAATATT
ATTTATCTTACATAGGTACATTATATTATTAATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCCCAACTACGTGTCCCAATCCTCGCTC

>HAP6

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTA AATTGTTTGCCCCATGAATGGTAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTACATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTTCCAATCACGTGTCCCAATCCTCGCTC

>HAP7

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAGTTGTTTACCCCATGAATAATAAGCATGTACATAATATT
ATTTATCTTACATAAGTACATTATATTATTGGTCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTGCGTGTCCCAATCCTCGCTC

>HAP8

TAAACGACAACAATCCGCCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCCATATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTA AATTGTTACCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATGAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCCCAACTACGTGTCCCAATCCTCGCTC

>HAP9

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAAACTATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAATTGTTTGCCCCATGAATAATAAGCATGTACATAATATT
ATTTATCTTACATAGGTACATTATATTATTAATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCCCAACTACGTGTCCCAATCCTCGCTC

>HAP10

TAAACGACAACAATCCGCCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCCATATAAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAATTGTTTCGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATGAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP11

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAATTGTTTGCCCCATGAATAACAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTACATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP12

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAAACTATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAAGTTGTTTGCCCCATGAATAATAAGCATGTACATAATATT
ATTTATCTTACATAGGTACATTATATTATTAATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGACAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCCCAACTACGTGTCCCAATCCTCGCTC

>HAP13

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAAACACTATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAGTTGTTTGCCCCATGAATAATAAGCATGTACATAATATT
ATTTATCTTACATAGGTACATTATATTATTAATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCCCAACTACGTGTCCCAATCCTCGCTC

>HAP14

TAAACGACAATAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAAACACCATACCCACCTGACATGCAATATCTTATGAATGGCTTATG
TACGTCGTGCATTA AATTGTCTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAGGTACATTATATTATTAATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCCCAACTACGTGTCCCAATCCTCGCTC

>HAP15

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTA AATTGTTTACCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAGTCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTGCGTGTCCCAATCCTCGCTC

>HAP16

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTA AATTGTCTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAAGCCCATGTTCCACGAGCTTGATCAC
CAAGCCGCGGGAAATCAGCAACCCTCCCAACTACGTGTCCCAATCCTCGCTC

>HAP17

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAATTGTTTGCCCCATGAATAACAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTACATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP18

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAATTGTCTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCAGCCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCCCAACTACGTGTCCCAATCCTCGCTC

>HAP19

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAATTGTCTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAGGTACATTATATTATTAATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCCCAACTACGTGTCCCAATCCTCGCTC

>HAP20

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAAGTTGTTTACCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGGTCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTGCGTGTCCCAATCCTCGCTC

>HAP21

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAGTTGTTTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAATCCTCTCAACTGCGTGTCCCAATCCTCGCTC

>HAP22

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAATTGTTTACCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATGTTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATATTCCACGAGCTTAGTCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTGCGTGTCCCAATCCTCGCTC

>HAP23

TAAACGACAACAATCCGCCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCCATATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAATTGTTTCGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP24

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAGTTGTTTACCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTACATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATATTCCACGAGCTTAGTCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTGCGTGTCCCAATCCTCGCTC

>HAP25

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA--
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATGTA
CATCGTGCATTAAATTGTTTGCCCCATGAATAATAAGCATGTACATAATATCATT
TATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCAAAT
CATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCACCAAG
CCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP26

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTTCGTGCATTAAACTGTTTGCCCCATGAATGGTAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTTCCAATTACGTGTCCCAATCCTCGCTC

>HAP27

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAAATTACA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAATTGTTTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAGGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCCTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTTCCAACCTACGTGTCCCAATCCTCGCTC

>HAP28

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAAACAATATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTTCGTGCATTAAAGTTGTTTGCCCCATGAATAATAAGCATGTACATAATATT
ATTTATCTTACATAGGTACATTATATTATTAATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTTCCAACCTACGTGTCCCAATCCTCGCTC

>HAP29

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAGTTGTTTACCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATATTCCACGAGCTTAGTCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP30

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAAACACTATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAGTTGTTTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTAATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP31

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAGTTGTTTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAGGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCCCAACTACGTGTCCCAATCCTCGCTC

>HAP32

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAATTTGCTTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP33

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAATTGTTTACCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATATTCCACGAGCTTAGTCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP34

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAATTGTTTGCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAGCCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCCCAACTACGTGTCCCAATCCTCGCTC

>HAP35

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAAGTTGTTTACCCCATGAATAATAAGCATGTACATAATATT
ATTTATCTTACATAAGTACATTATATTATTGGTCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP36

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAATTGTCTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAGCCTATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCCCAACTACGTGTCCCAATCCTCGCTC

>HAP37

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAATTGTTTACCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGGTCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTGCGTGTCCCAATCCTCGCTC

>HAP38

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAAATTATA-
CCCCACATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAATTGTTTGCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCCTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTTCCAACCTACGTGTCCCAATCCTCGCTC

>HAP39

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAGCACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAATTGTTTACCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAGGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACATGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP40

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAGCACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAATTGTTTGCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACATGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP41

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAATTGTTTGCCCCATGAATAACAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP42

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAATTGTCTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAGGTACATTATATTATTAATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCCCAACTACGTGTCCCAATCCTCGCTC

>HAP43

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAATTGTCTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACAT~GGTACATTATATTATTAATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCCCAACTACGTGTCCCAATCCTCGCTC

>HAP44

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAATTGTTTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGGTCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTGCGTGTCCCAATCCTCGCTC

>HAP45

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA--
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATGT
ACGTCGTGCATTAAATTGTTTGCCCCATGAATAATAAGCATGTACATAATATCA
TTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCAA
ATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCACC
AAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP46

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAATTGTCTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAGGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCATAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCCCAACTACGTGTCCCAATCCTCGCTC

>HAP47

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAATTATTTACCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAGTCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTGCGTGTCCCAATCCTCGCTC

>HAP48

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAAGTTGTTTGCCCCATGAATAATAAGCATGTACATAATATT
ATTTATCTTACATAAGTACATTATATTATTGGTCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTGCGTGTCCCAATCCTCGCTC

>HAP49

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAATTGCTTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACCACGTGTCCCAATCCTCGCTC

>HAP50

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAATTGTCTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAGGTACATTATATTATTAATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCCCAACTACGTGTCCCAATCCTCGCTC

>HAP51

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAAGTTGTTTACCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAGTCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTGCGTGTCCCAATCCTCGCTC

>HAP52

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA--
CCCCACATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATGTA
CATCGTGCATTAAATTGTTTGCCCCATGAATAATAAGCATGTACATAATATCATT
TATCTTACATAAGTACATTATATTATTGATCGTACATACCCCATCCAAGTCAAAT
CATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCACCAAG
CCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP53

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATACC
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATGT
ACGTCGTGCATTAAATTGTTTGCCCCATGAATAATAAGCATGTACATAATATCA
TTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCAA
ATCATTTCCAGCCAACACGCATATCACAACCCATATTCCACGAGCTTAGTCACC
AAGCCGCGGGAAATCAGCAACCCTCTCAACTGCGTGTCCCAATCCTCGCTC

>HAP54

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAATTGTTTGCCCCATGAATAACAAGCATGTACATAATATC
ATTTATCTTACATAGGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCCCAACTACGTGTCCCAATCCTCGCTC

>HAP55

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAAGTTGTTTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGGTCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTGCGTGTCCCAATCCTCGCTC

>HAP56

TAAACGACAACAATCCGCCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCCATATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAATTGTTTCGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATGAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP57

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAAGTTGTTTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCCCAACTACGTGTCCCAATCCTCGCTC

>HAP58

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAAACACTATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAATTGTTTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAGGTACATTATATTATTAATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP59

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAAACACTATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAATTGTTTGCCCCATGAATAATAAGCATGTACATAATATT
ATTTATCTTACATAAGTACATTATATTATTAATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCCCAACTACGTGTCCCAATCCTCGCTC

>HAP60

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAATTGTTTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAGGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCCCAACTACGTGTCCCAATCCTCGCTC

>HAP61

TAAACGACAACAATCCGCCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCCATATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAATTGTCCGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATGAGTACATCATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP62

TAAACGACAACAATCCGCCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCCATATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAATTGTCCGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATGAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP63

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCCACATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAATTGTTTACCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATCATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAGTCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTGCGTGTCCCAATCCTCGCTC

>HAP64

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCCACATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAAGTTGTTTACCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAGGTACATTATATTATTGGTCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTGCGTGTCCCAATCCTCGCTC

>HAP65

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAGATTGTTTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAGGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCCCAACTACGTGTCCCAATCCTCGCTC

>HAP66

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAATTGTTTACCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATATTCCACGAGCTTAGTCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTGCGTGTCCCAATCCTCGCTC

>HAP67

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTACA-
CCCCACATAACGCCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAATTGTTTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATGAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCATAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP68

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGGTTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAATTGTCTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAGGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAGCCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP69

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAATTGTCTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTTCCAACCTACGTGTCCCAATCCTCGCTC

>HAP70

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAATTGTCTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAGGTACATTATATTATTAATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP71

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAAACTATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAAGTTGTTTGCCCCATGAATAATAAGCATGTACATAACATT
ATTTATCTTACATAGGTACATTATATTATTAATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCCCAACTACGTGTCCCAATCCTCGCTC

>HAP72

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAAACTATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAAGTTGTTTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAGGTACATTATATTATTAATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP73

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAGTTGTTTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGGTCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTGCGTGTCCCAATCCTCGCTC

>HAP74

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAAACACTATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAGTTGTTTACCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAGGTACATTATATTATTAATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP75

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAATTTGTCTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCCCAACTACGTGTCCCAATCCTCGCTC

>HAP76

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAATTTGTTTGCCCCATGAATAATAAGCATGTACATGATATC
ATTTATCTTACATAGGTACATTATATTATTGATCGTGCATACCCCATTCAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP77

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAAACACTATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAGTTGTTTGCCCCATGAATAATAAGCATGTACATAATATT
ATTTATCTTACATAGGTACATTATATTATTAATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACATGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCCCAACTACGTGTCCCAATCCTCGCTC

>HAP78

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTA AATTGTTTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATATTCCACGAGCTTAGTCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTGCGTGTCCCAATCCTCGCTC

>HAP79

TAAACGACAATAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTA AATTGTCTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAGGTACATTATATTATTAATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCCCAACTACGTGTCCCAATCCTCGCTC

>HAP80

TAAACGACAACAATCCGCCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCCATATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTA AATTGTTACCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATGAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCATAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP81

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAGCACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAGATTGTTTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAGGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACATGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP82

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA--
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATGTA
CATCGTGCATTAAATTATTTGCCCCATGAATAATAAGCATGTACATAATATCATT
TATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCAAAT
CATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCACCAAG
CCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP83

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAATTGCTTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAGTCCTCGCTC

>HAP84

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAAGTTGTTTACCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAGGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATATTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTGCGTGTCCCAATCCTCGCTC

>HAP85

TAAACGACAACAACCTCACCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAGCACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAGATTGTTTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAGGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACATGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP86

TAAACGACAACAATCCGCCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCCATATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAATTGTCCACCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATGAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP87

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAAGTTGTTTGCCCCATGAATAACAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTACATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP88

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAAATTATA--
CCCCACATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATGTA
CGTCGTGCATTAGATTGTTTGCCCCATGAATAATAAGCATGTACATAATATCATT
TATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCAAAT
CATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCCTAATCACCAAG
CCGCGGGAAATCAGCAACCCTTCCGACTACGTGTCCCAATCCTCGCTC

>HAP89

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAATTGTTTACCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAGTCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP90

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAAGTTGTTTACCCCATGAATAATAAGCATGTACATAATATT
ATTTATCTTACATAAGTACATTATATTATTGGTCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTGCGTGTCCCAATCCTCGCTC

>HAP91

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAATTGTTTGCCCCATGAATAACAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTACATTATTGATCGTGCATACCCCATCCAGGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP92

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAATTGTTTACCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATATTCCACGAGCTTAGTCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTGCGTGTCCCAATCCTCGCTC

>HAP93

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAATTGTCTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAGGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCCCAACTACGTGTCCCAATCCTCGCTC

>HAP94

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGGTTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAATTGTTTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAGGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCAGAGCCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP95

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAGCACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAATTGTTTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAGGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACATGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP96

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAAGTTGTTTACCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGGTCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTGCGTGTCCCAATCCTCGCTC

>HAP97

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAAACACTATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAGTTGTTTGCCCCATGAATAATAAGCATGTACATAATATT
ATTTATCTTACATAAGTACATTATATTATTAATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCCCAACTACGTGTCCCAATCCTCGCTC

>HAP98

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAGTTGTTTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATCATATTATTGGTCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTGCGTGTCCCAATCCTCGCTC

>HAP99

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAATTTGTTTACCCCATGAGTAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAGTCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTGCGTGTCCCAATCCTCGCTC

>HAP100

TAAACGACAACAATCCGCCCTCATGTGCTATGTCAGTATCAAATTATA-
CCCCCATATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAATTTGTTTACCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATGAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP101

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAGCACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAGTTGTTTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAGGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACATGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP102

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAACTGTTTGCCCCATGAATAGTAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTACATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTTCCAATCACGTGTCCCAATCCTCGCTC

>HAP103

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAAATTGTTTGCCCCATGAATAACAAGCATGTACATAGTATC
ATTTATCTTACATAAGTACATTACATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP104

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAATTGTTTACCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAGTCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTGCGTGTCCCAATCCTCGCTC

>HAP105

TAAACGACAACAATCCGCCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCCATATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAATTGTTACCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATGAGTACATCATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP106

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAAATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCGTGCATTAGATTGTTTGCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCATAACCCATGTTCCACGAGCCTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTTCCAACCTACGTGTCCCAATCCTCGCTC

>HAP107

TAAACGACAACAATCCGCCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCCATATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAATTGTTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATGAGTACATTATATTATTAATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP108

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATAACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAATTGCTTGCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAGGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP109

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAGCACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAATTGTTTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACATGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTACGTGTCCCAATCCTCGCTC

>HAP110

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCTGTCATTGAATTGTTTACCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTGCGTGTCCCAATCCTCGCTC

>HAP111

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAATTGTTTACCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCTCAACTGCGTGTCCCAATCCTCGCTC

>HAP112

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAAGTTGTTTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCCCAACTACGTGTCCCAATCCTCGCTC

>HAP113

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACTATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACATCGTGCATTAAGTTGTTTGCCCCATGAATGATAAGCATGTACATAATATT
ATTTATCTTACATAGGTACATTATATTATTAATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGCCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCCCAACTACGTGTCCCAATCCTCGCTC

>HAP114

TAAACGACAACAATTCACCCTCATGTGCTATGTCAGTATCAGATTATA-
CCCCACATAACACCATACCCACCTGACATGCAATATCTTATGAATGGCCTATG
TACGTCTGTGCATTAAATTGTTTGCCCCATGAATAATAAGCATGTACATAATATC
ATTTATCTTACATAAGTACATTATATTATTGATCGTGCATACCCCATCCAAGTCA
AATCATTTCCAGTCAACACGCATATCACAACCCATGTTCCACGAGCTTAATCAC
CAAGCCGCGGGAAATCAGCAACCCTCCCAATTACGTGTCCCAATCCTCGCTC

**APPENDIX B: LIST OF HAPLOTYPES INDICATING VARIABLE
NUCLEOTIDES IN A 478 BP FRAGMENT OF THE
MITOCHONDRIAL DNA D-LOOP**

	15490	15493	15494	15496	15521	15522	15526	15527i	15528	15534	15538	15540	15542	15577	15585	15595	15596	15597	15598	15600	15601	15602	15603	15604	15613	15615	15616	15617	15630	15631	15632	
HAP30	.	.	.	.	.	.	.	.	.	.	.	.	T	.	G	.	.	G	.	.	T	.	.	.	.	.	.	.	.	.	T	
HAP31	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	G	G	.	.	T	.	.	.	.	.	.	.	.	.	T	
HAP32	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
HAP33	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	T	.	.	A	.	.	.	.	.	.	T	
HAP34	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	
HAP35	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	G	.	.	T	.	.	A	.	.	.	.	.	.	T	
HAP36	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	T	
HAP37	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	T	.	.	A	.	.	.	.	.	.	T	
HAP38	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	
HAP39	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	T	.	.	A	.	.	.	.	.	.	T	
HAP40	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	
HAP41	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	T	.	.	.	.	.	.	C	.	.	T	
HAP42	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	T	
HAP43	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	T	
HAP44	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	
HAP45	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	
HAP46	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	T	
HAP47	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	T	.	.	A	.	.	.	.	.	.	.	T	
HAP48	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	G	.	T	.	.	.	.	.	.	.	.	.	.	T	
HAP49	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HAP50	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.	T	
HAP51	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	G	.	T	.	.	A	.	.	.	.	.	.	.	T	
HAP52	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	T	
HAP53	.	.	.	.	.	.	C	.	.	.	.	.	.	G	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	T	
HAP54	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	T	.	.	.	.	.	.	.	C	.	.	T	
HAP55	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	G	.	T	.	.	.	.	.	.	.	.	.	.	T	
HAP56	.	.	C	G	.	.	.	.	T	.	.	.	.	G	.	.	.	.	.	T	.	.	C	.	.	.	.	.	.	.	T	
HAP57	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	G	.	T	.	.	.	.	.	.	.	.	.	.	T	
HAP58	.	.	.	.	.	.	.	.	.	.	.	.	T	.	G	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	T	
HAP59	.	.	.	.	.	.	.	.	.	.	.	.	T	.	G	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	T	

	15490	15493	15494	15496	15521	15522	15526	15527i	15528	15534	15538	15540	15542	15577	15585	15595	15596	15597	15598	15600	15601	15602	15603	15604	15613	15615	15616	15617	15630	15631	15632		
HAP60	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T		
HAP61	.	.	C	G	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	T	C	C	.	.	.	.	.	.	.	T		
HAP62	.	.	C	G	.	.	.	.	.	T	.	.	.	.	G	.	.	.	.	.	T	C	C	.	.	.	.	.	.	.	T		
HAP63	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	T	.	.	A	.	.	.	.	.	.	T		
HAP64	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	G	.	.	T	.	.	A	.	.	.	.	.	.	T		
HAP65	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	G	.	.	.	T	.	.	.	.	.	.	.	.	.	T		
HAP66	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	T	.	.	A	.	.	.	.	.	.	T		
HAP67	.	.	.	.	.	.	C	.	.	.	.	G	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T		
HAP68	.	.	.	.	.	G	.	.	.	.	.	.	.	.	G	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	T		
HAP69	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	T		
HAP70	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	T		
HAP71	.	.	.	.	.	.	.	.	.	.	.	.	T	.	G	.	.	G	.	.	T	.	.	.	.	.	.	.	.	.	T		
HAP72	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	G	.	.	T	.	.	.	.	.	.	.	.	.	T		
HAP73	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	T	.	.	.	.	.	.	.	.	.	T		
HAP74	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	G	.	.	T	.	.	A	.	.	.	.	.	.	T		
HAP75	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	T		
HAP76	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	T	.	.	.	.	.	.	.	G	.	T		
HAP77	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	G	.	.	T	.	.	.	.	.	.	.	.	.	T		
HAP78	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T		
HAP79	T	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	T		
HAP80	.	.	C	G	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	T	.	C	A	.	.	.	.	.	.	.	T	
HAP81	.	.	.	.	.	.	.	.	.	.	G	.	.	.	G	.	G	.	.	.	T	.	.	.	.	.	.	.	.	.	.	T	
HAP82	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	T	.	.	.	.	.	.	.	.	.	.	T	
HAP83	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T
HAP84	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	G	.	.	T	.	.	A	.	.	.	.	.	.	.	T	
HAP85	.	C	.	.	.	.	.	.	.	.	G	.	.	.	G	.	G	.	.	.	T	.	.	.	.	.	.	.	.	.	.	T	
HAP86	.	.	C	G	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	T	C	C	A	.	.	.	.	.	.	.	T	
HAP87	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	G	.	.	T	.	.	.	.	.	.	C	.	.	.	T	
HAP88	.	.	.	.	A	.	.	.	.	.	.	.	.	.	G	.	G	.	.	.	T	.	.	.	.	.	.	.	.	.	.	T	
HAP89	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	A	.	.	.	.	.	.	.	T	

	15490	15493	15494	15496	15521	15522	15526	15527i	15528	15534	15538	15540	15542	15577	15585	15595	15596	15597	15598	15600	15601	15602	15603	15604	15613	15615	15616	15617	15630	15631	15632*	
HAP90	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	G	.	.	T	.	.	A	.	.	.	.	.	.	T	
HAP91	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	T	C	.	.	.	.	.	C	.	.	T	
HAP92	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	T	
HAP93	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	T	
HAP94	.	.	.	.	G	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	
HAP95	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	
HAP96	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	T	.	.	A	.	.	.	.	.	.	T	
HAP97	.	.	.	.	.	.	.	.	.	.	.	T	.	.	G	.	.	G	.	.	T	.	.	.	.	.	.	.	.	.	T	
HAP98	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	G	.	.	T	.	.	.	.	.	.	.	.	.	T	
HAP99	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	T	.	.	A	G	.	.	.	.	.	T	
HAP100	.	.	C	G	A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	T	.	C	A	.	.	.	.	.	.	T	
HAP101	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	G	.	.	T	.	.	.	.	.	.	.	.	.	T	
HAP102	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	C	T	.	.	.	.	.	G	.	.	.	T	
HAP103	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	T	.	.	.	.	.	.	C	.	G	T	
HAP104	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	A	.	.	.	.	.	.	T	
HAP105	.	.	C	G	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	T	.	C	A	.	.	.	.	.	.	T	
HAP106	.	.	.	.	A	.	.	.	.	.	.	.	.	.	G	.	G	.	.	.	T	.	.	.	.	.	.	.	.	.	T	
HAP107	.	.	C	G	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	T	.	C	.	.	.	.	.	.	.	T	
HAP108	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T
HAP109	.	.	.	.	.	.	.	.	.	.	.	T	.	G	.	G	.	G	.	T	.	.	.	.	.	.	.	.	.	.	T	
HAP110	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	G	.	.	.	.	T	.	.	A	.	.	.	.	.	.	T	
HAP111	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	A	.	.	.	.	.	.	T	
HAP112	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	T	.	.	.	.	.	.	.	.	.	T	
HAP113	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	G	.	.	T	.	.	.	.	G	.	.	.	.	T	
HAP114	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	T	

*The genomic positions are continued through 15635 in the table at the page 81

	15635	15649	15650	15657	15659	15660	15666	15667	15672	15683	15686	15703	15709	15718	15720	15723	15726	15737	15739	15740	15766	15770	15771	15773	15775	15776	15777	15787
HAP1	C	A	A	T	T	A	G	A	G	C	A	T	C	C	A	C	G	T	A	A	C	C	T	A	C	T	A	A
HAP2	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HAP3	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
HAP4	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	G	.	.	.	.	.	G	.	
HAP5	T	.	G	.	.	A	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	
HAP6	.	.	.	.	C	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	T	C	.	T	C	.	
HAP7	T	.	.	.	.	.	G	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	
HAP8	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	
HAP9	T	.	G	.	.	A	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	
HAP10	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
HAP11	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
HAP12	T	.	G	.	.	A	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	
HAP13	T	.	G	.	.	A	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	
HAP14	.	.	G	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	
HAP15	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	G	.	.	.	.	.	G	.	
HAP16	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	G	.	.	.	C	.	.	.	.	
HAP17	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
HAP18	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	C	.	.	.	.	.	
HAP19	.	.	G	.	.	A	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	C	.	.	.	.	.	
HAP20	.	.	.	.	.	G	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	
HAP21	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	T	.	.	.	.	G	.	
HAP22	.	.	.	.	.	G	.	.	.	.	.	C	.	.	.	.	A	.	.	G	.	.	.	.	.	G	.	
HAP23	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
HAP24	.	.	.	.	C	.	.	.	.	.	.	C	.	.	.	.	A	.	.	G	.	.	.	.	.	G	.	
HAP25	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
HAP26	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	T	C	.	T	.	.	
HAP27	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	T	C	.	.	.	.	
HAP28	T	.	G	.	.	A	.	.	.	.	.	C	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	
HAP29	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	A	.	.	G	.	.	.	.	.	.	.	
HAP30	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	
HAP31	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	
HAP32	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	

	15635	15649	15650	15657	15659	15660	15666	15667	15672	15683	15686	15703	15709	15718	15720	15723	15726	15737	15739	15740	15766	15770	15771	15773	15775	15776	15777	15787
HAP33		.	.		.	.	.		.	.	.	C		.	.	.	A	.		G			.	.	.		.	
HAP34			.		.	.	.		.	.	.			.	G		.						C	.	.		.	
HAP35	T		.		.	.	.	G		.	.	C												.			.	
HAP36			.		.	.	.			.	.	.		.	G	T							C	.	.		.	
HAP37					.	.	.	G		.	.	C												.	.	G	.	
HAP38					.	.	.				.	.						C				T	C	.	.		.	
HAP39			G		.	.	.				.		T											.	.		.	
HAP40					.	.	.				.		T											.	.		.	
HAP41					.	.	.				.													.	.		.	
HAP42			G		.	.	A	.		.	.	C											C	.	.		.	
HAP43		-	G		.	.	A	.		.	.	.											C	.	.		.	
HAP44					.	.	.	G		.	.	C												.	.	G	.	
HAP45					.	.	.				.													.	.		.	
HAP46			G		.	.	.				.			T									C	.	.		.	
HAP47					.	.	.				.	C								G			.	.	.	G	.	
HAP48	T				.	.	.	G		.	.	C												.	.	G	.	
HAP49					.	.	.				.													.	.	C	.	
HAP50			G		.	.	A	.		.	.												C	.	.		.	
HAP51					.	.	.			.	.	C								G				.	.	G	.	
HAP52					.	.	.	.	A	.	.	.												.	.		.	
HAP53					.	.	.	.		.	.	C					A			G				.	.	G	.	
HAP54			G		.	.	.			.	.												C	.	.		.	
HAP55					.	.	.	G		.	.	C												.	.	G	.	
HAP56		G			.	.	.			.	.													.	.		.	
HAP57					.	.	.			.	.												C	.	.		.	
HAP58			G		.	.	A	.		.	.													.	.		.	
HAP59	T				.	.	A	.		.	.	C											C	.	.		.	
HAP60			G		.	.	.			.	.												C	.	.		.	
HAP61		G		C		.	.			.	.													.	.		.	
HAP62		G				.	.			.	.													.	.		.	
HAP63				C		.	.			.	.	C								G				.	.	G	.	
HAP64			G		.	.	.	G		.	.	C												.	.	G	.	

	15635	15649	15650	15657	15659	15660	15666	15667	15672	15683	15686	15703	15709	15718	15720	15723	15726	15737	15739	15740	15766	15770	15771	15773	15775	15776	15777	15787
HAP65	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.
HAP66	.	.	.	.	.	.	.	.	.	.	.	C	.	T	.	.	A	.	.	G	.	.	.	.	.	.	G	.
HAP67	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HAP68	.	.	G	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.
HAP69	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.
HAP70	.	.	G	.	.	A	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HAP71	T	.	G	.	.	A	A	.	.	.	.	C	C	C	A	.	.	.	.	.	.	.	C	.	.	.	.	.
HAP72	.	.	G	.	.	A	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HAP73	.	.	.	.	.	.	.	G	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.
HAP74	.	.	G	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HAP75	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.
HAP76	.	.	G	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HAP77	T	.	G	.	.	A	A	.	.	.	.	C	T	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.
HAP78	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	A	.	.	G	.	.	.	.	.	.	G	.
HAP79	.	.	G	.	.	A	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.
HAP80	.	G	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HAP81	.	.	G	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HAP82	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HAP83	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
HAP84	.	.	G	.	.	.	.	.	.	.	.	C	.	.	.	A	.	.	.	.	.	.	.	.	.	.	G	.
HAP85	.	.	G	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HAP86	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HAP87	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HAP88	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	T	C	G	.	.	.	.
HAP89	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.
HAP90	T	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.
HAP91	.	.	.	.	C	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HAP92	.	.	G	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.
HAP93	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.
HAP94	.	.	G	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.
HAP95	.	.	G	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HAP96	.	.	.	.	.	.	.	G	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.

	15635	15649	15650	15657	15659	15660	15666	15667	15672	15683	15686	15703	15709	15718	15720	15723	15726	15737	15739	15740	15766	15770	15771	15773	15775	15776	15777	15787
HAP97	T	.	.	.	.	.	A	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.
HAP98	.	.	.	C	.	.	.	G	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.
HAP99	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	G	.	.	.	.	.	.	G	.
HAP100	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HAP101	.	.	G	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HAP102	.	.	.	.	C	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	T	C	.	C	.	.
HAP103	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.
HAP104	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.
HAP105	.	G	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HAP106	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	C	.	.	.	T	C	.	.	.	.	.
HAP107	.	G	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HAP108	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HAP109	T	.	G	.	.	.	A	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.
HAP110	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.
HAP111	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.
HAP112	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.
HAP113	T	.	G	.	.	.	A	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.
HAP114	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	T	.	.	.

**APPENDIX C: MISMATCH DISTRIBUTION GRAPHS OF EACH
POPULATION BASED ON BOTH POPULATION EXPANSION AND
CONSTANT POPULATION SIZE**

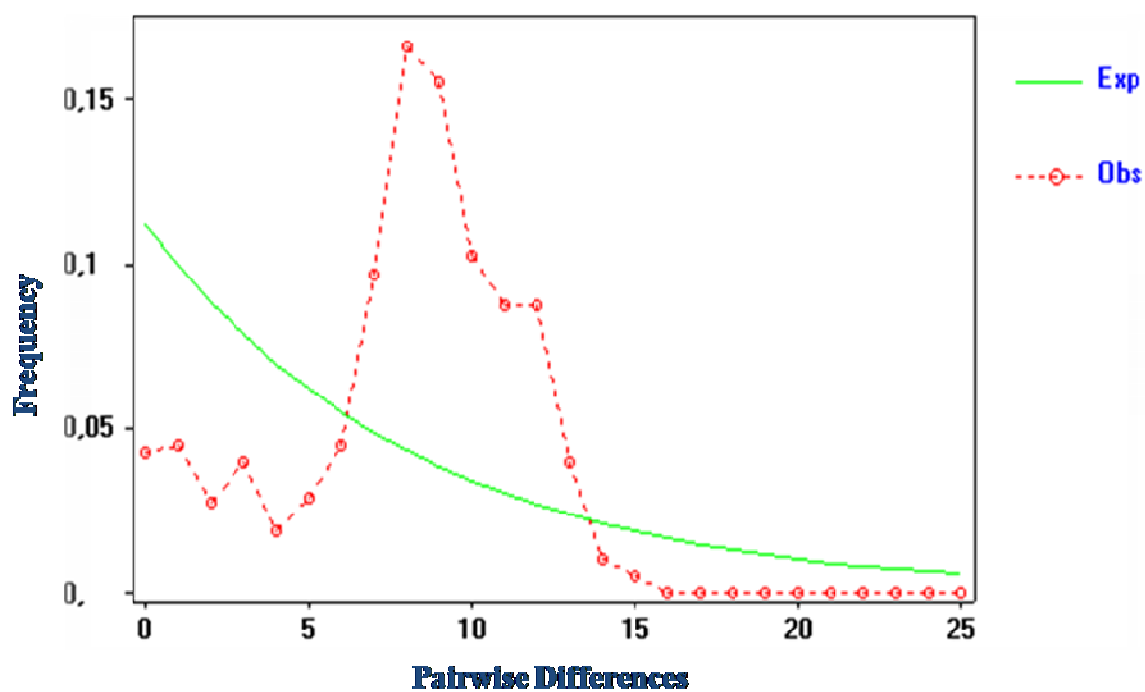


Figure C.1. The graph of AP based on the theory of a constant-sized population

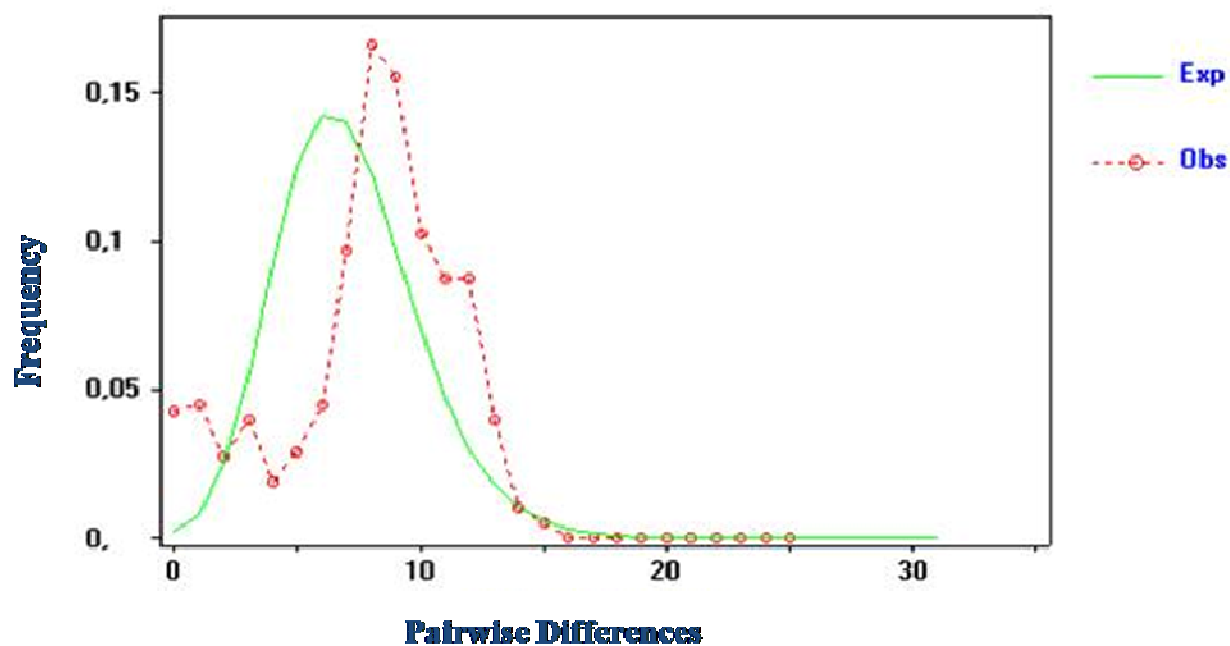


Figure C.2. The graph of AP based on the theory that a population expansion has occurred

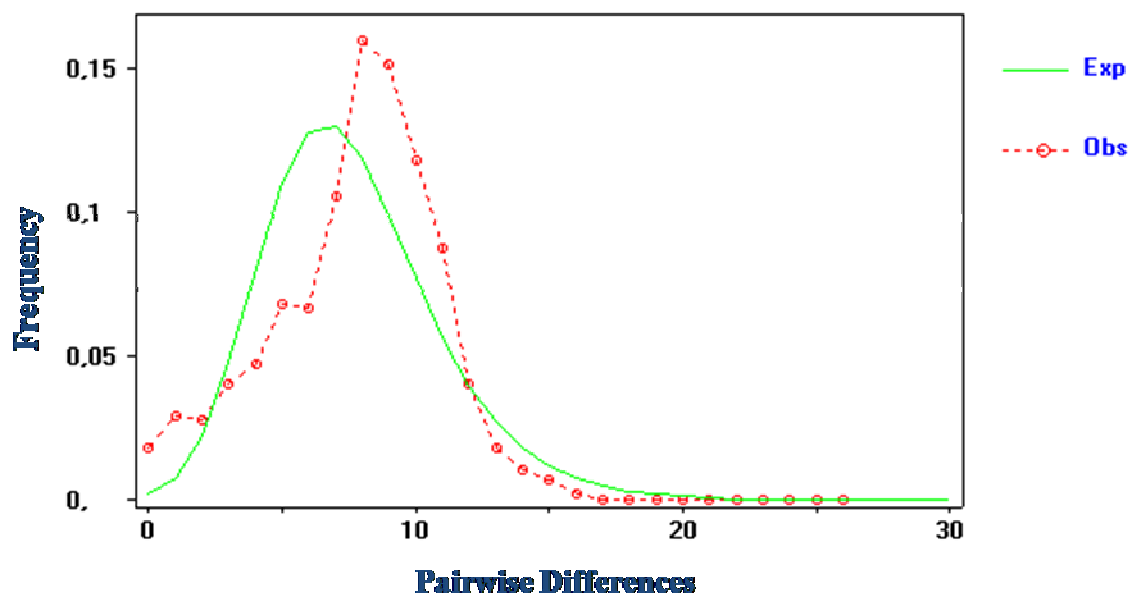


Figure C.3. The graph of CKR based on the theory that a population expansion has occurred

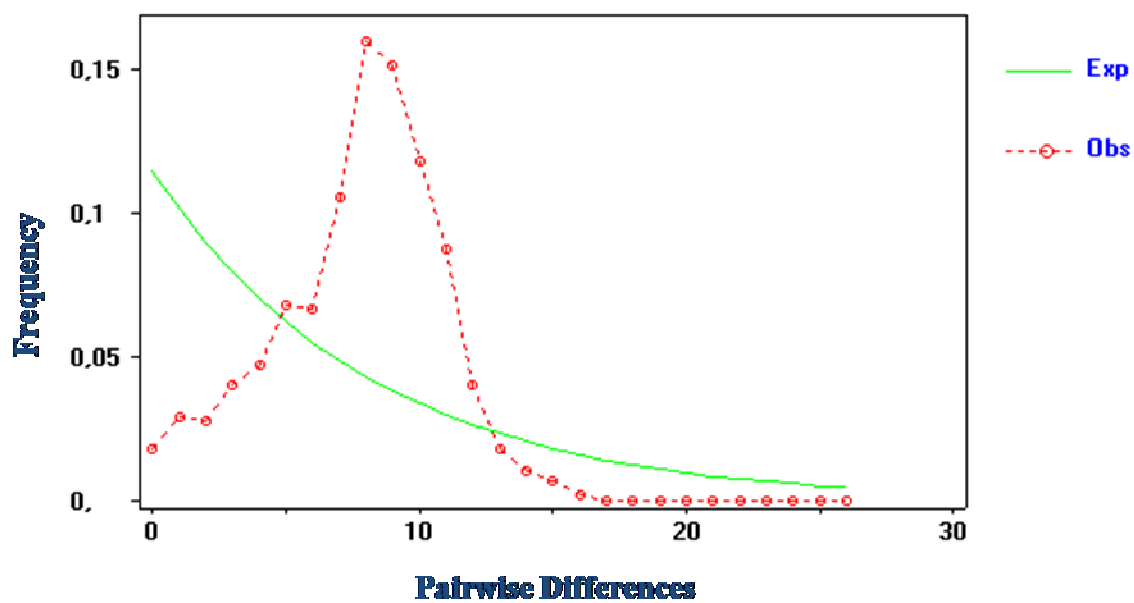


Figure C.4. The graph of CKR based on the theory of a constant-sized population

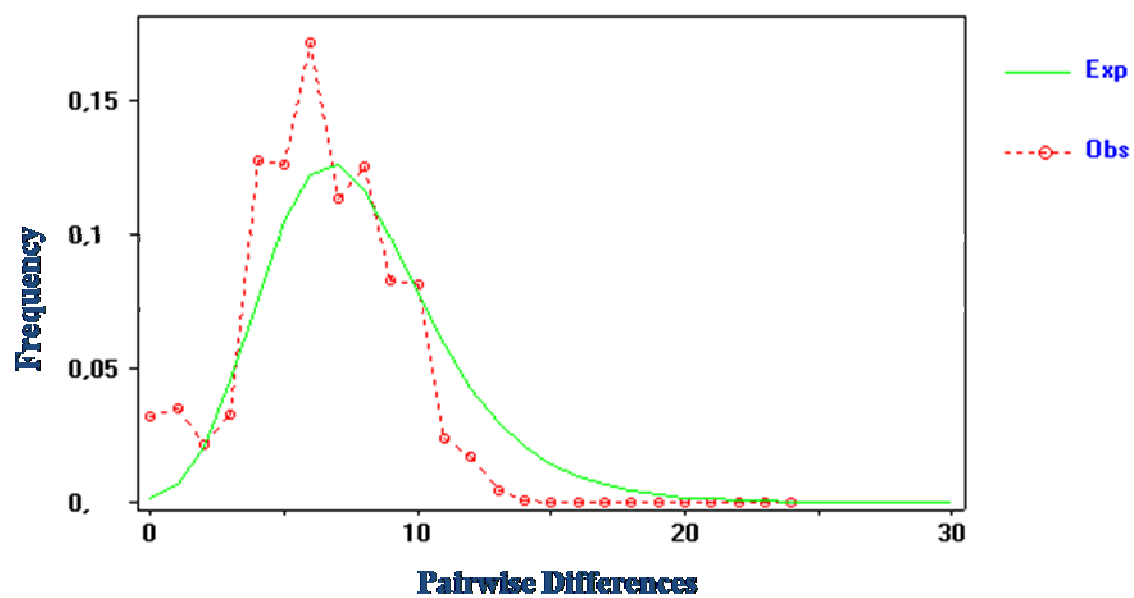


Figure C.5. The graph of CNK based on the theory that a population expansion has occurred

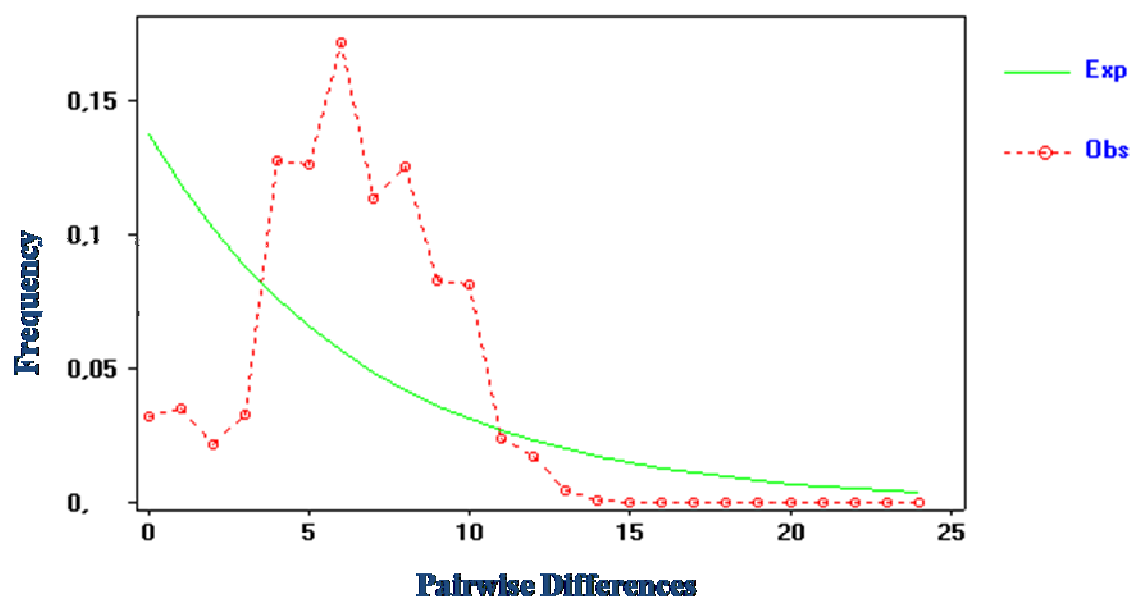


Figure C.6. The graph of CNK based on the theory of a constant-sized population

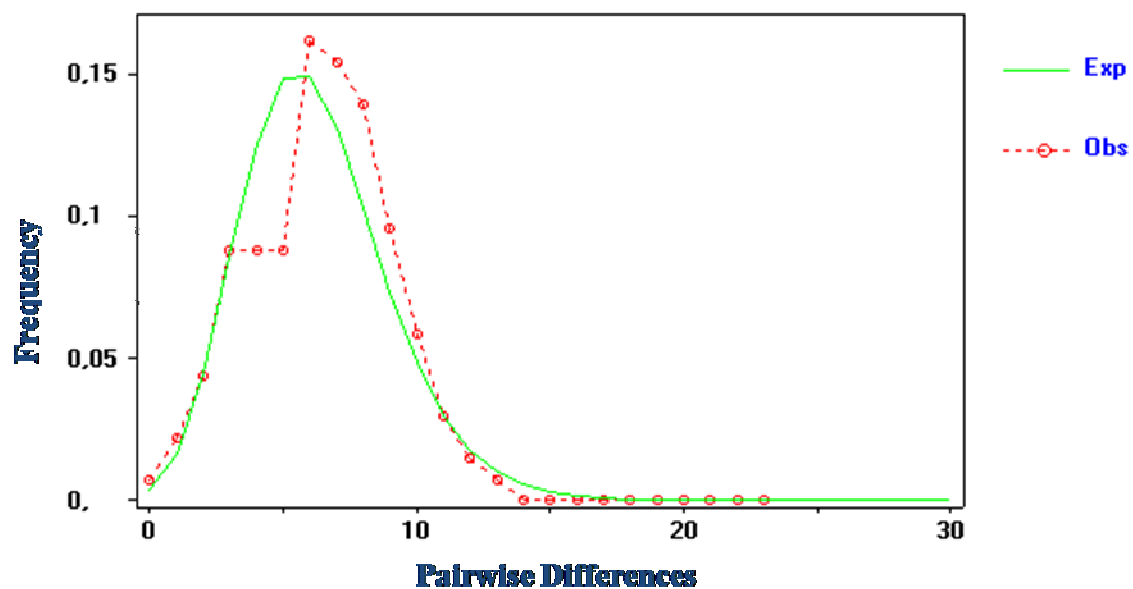


Figure C.7. The graph of HH based on the theory that a population expansion has occurred

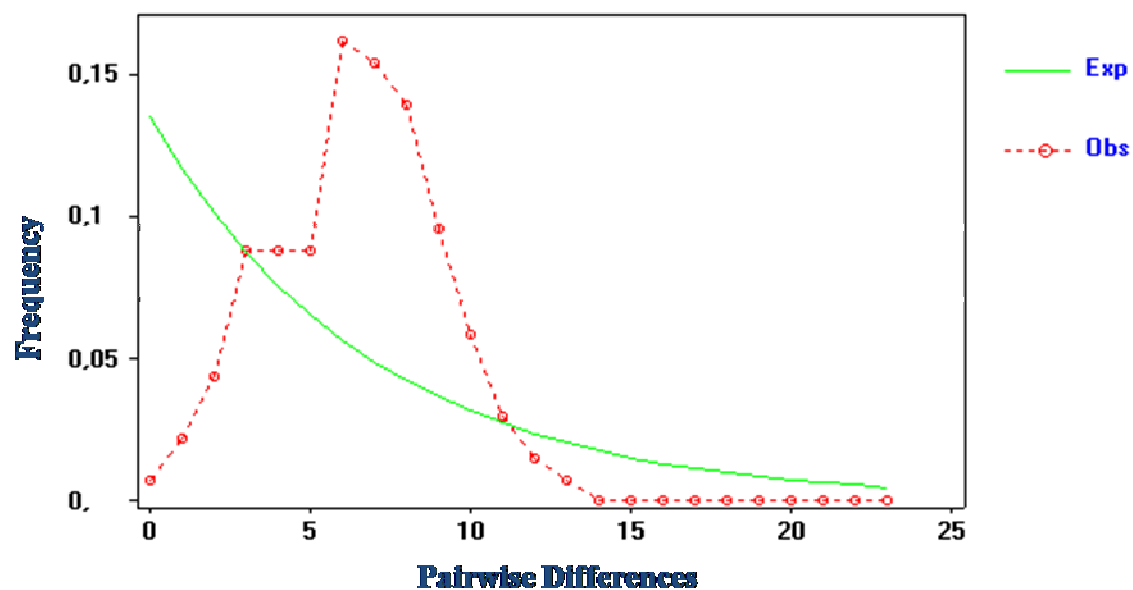


Figure C.8. The graph of HH based on the theory of a constant-sized population

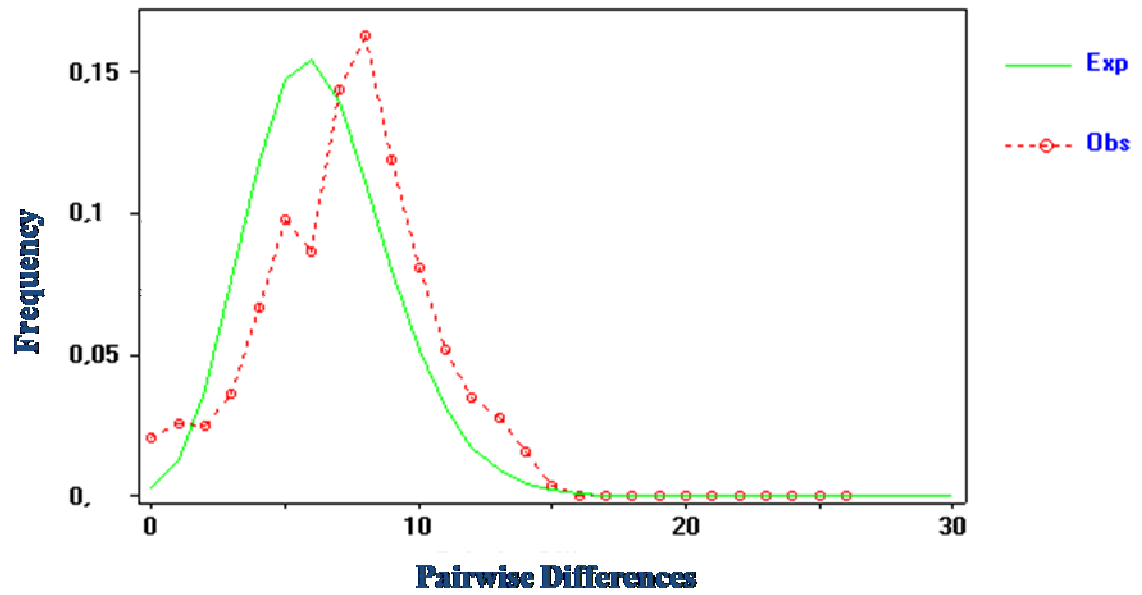


Figure C.9. The graph of ERZ based on the theory that a population expansion has occurred

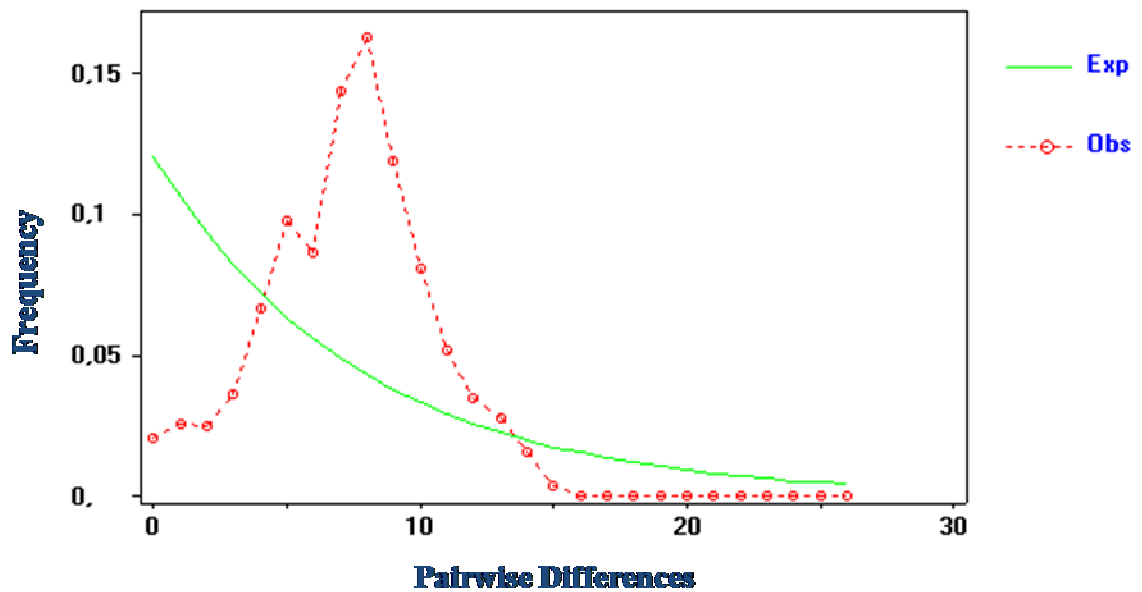


Figure C.10. The graph of ERZ based on the theory of a constant-sized population

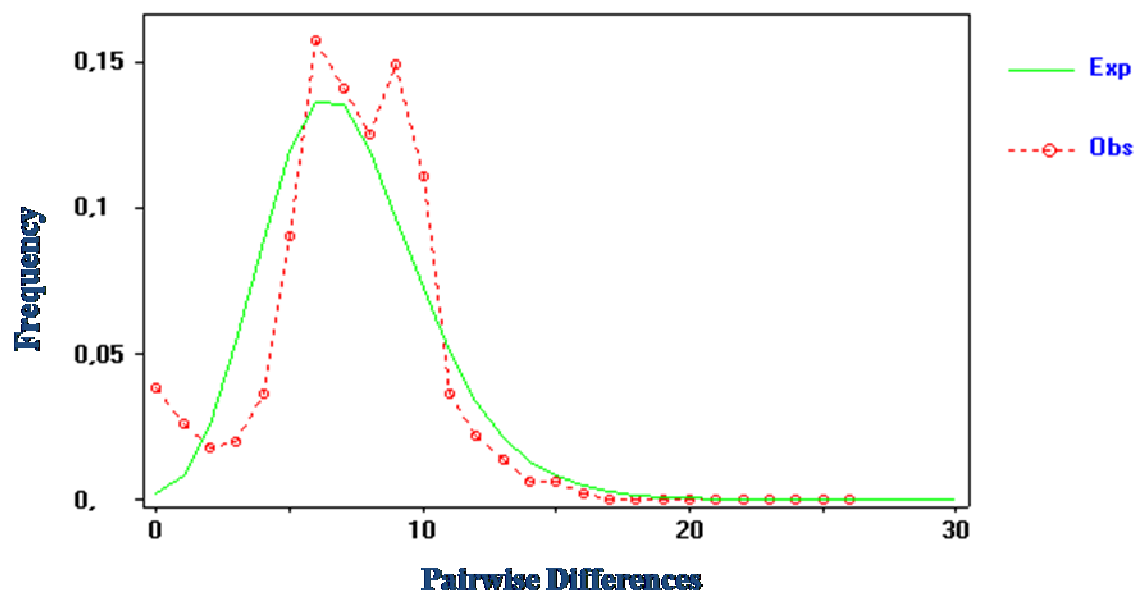


Figure C.11. The graph of MLK based on the theory that a population expansion has occurred

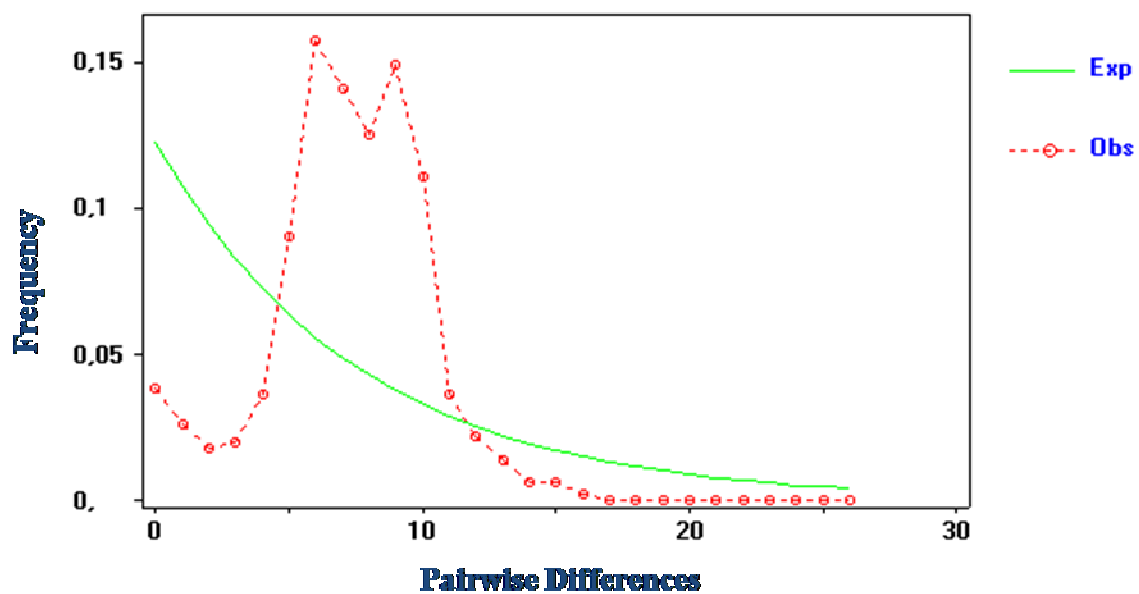


Figure C.12. The graph of MLK based on the theory of a constant-sized population

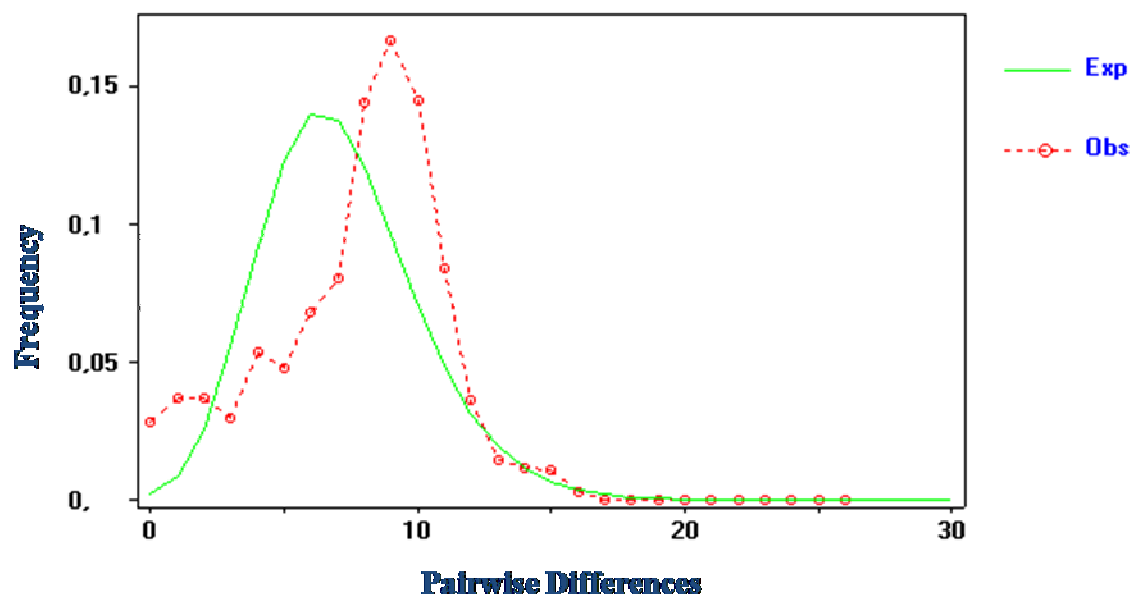


Figure C.13. The graph of EAST ANATOLIAN horses based on the theory that a population expansion has occurred

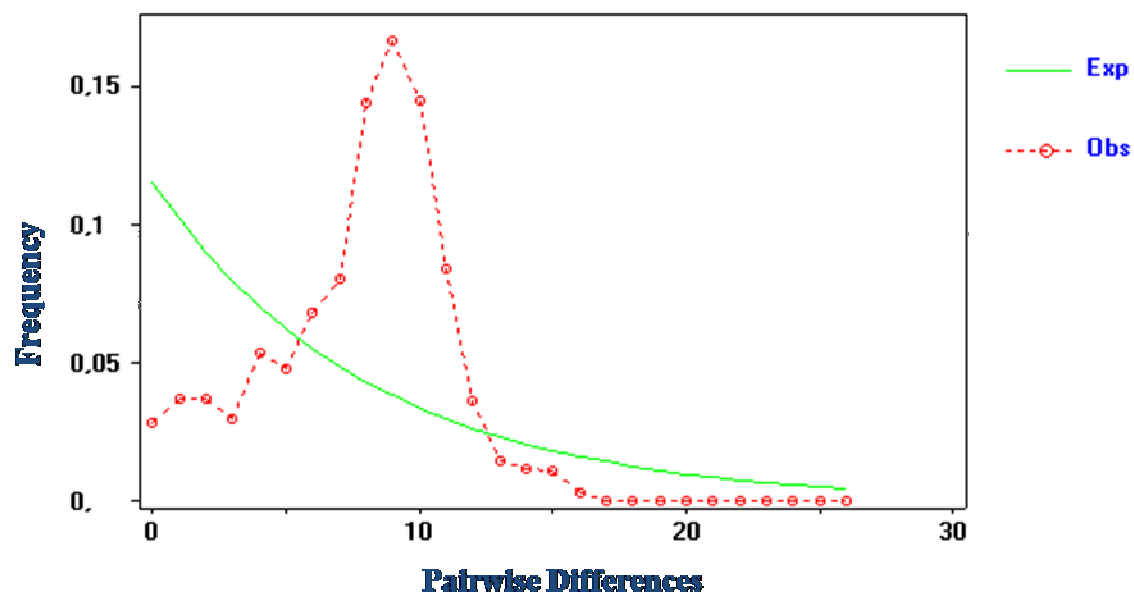


Figure C.14. The graph of EAST ANATOLIAN horses based on the theory of a constant-sized population