COMPARATIVE ANALYSES FOR THE CENTRAL ASIAN CONTRIBUTION TO ANATOLIAN GENE POOL WITH REFERENCE TO BALKANS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY
CEREN CANER BERKMAN

IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
BIOLOGICAL SCIENCES

SEPTEMBER 2006

Approval of the Graduate School of Natural and Applied Sciences

Prof. Dr. Canan Özgen
Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Doctor of Philosophy.

Prof. Dr. Semra Kocabıyık
Head of Department

This is to certify that we have read this thesis and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Doctor of Philosophy.

Prof. Dr. İnci Togan
Supervisor

Examining Committee Members

Prof. Dr. Zeki Kaya (METU, Biological Sciences) _____

Prof. Dr. İnci Togan (METU, Biological Sciences) _____

Prof. Dr. Işık Bökesoy (Ankara Univ, Faculty of Medicine) _____

Prof. Dr. G. Wilhelm Weber (METU, Applied Mathematics) _____

Ass. Prof. Dr. Elif Erson (METU, Biological Sciences) _____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Ceren, Caner Berkman

Signature :

ABSTRACT

COMPARATIVE ANALYSES FOR THE CENTRAL ASIAN CONTRIBUTION TO ANATOLIAN GENE POOL WITH REFERENCE TO BALKANS

Ceren Caner Berkman

Ph.D. Department of Biological Sciences
Supervisor: Prof. Dr. İnci Togan

September, 2006, 151 pages.

Around 1000 ya, Turkic language started to be introduced to Turkey and Azerbaijan (Region of language replacement, RLR) in parallel with the migrations of Turkic speaking nomadic groups from Central Asia. The Central Asian contribution to the RLR was analyzed with four admixture methods considering different evolutionary forces. Furthermore, the association between the Central Asian contribution and the language replacement episode was estimated by comparatively analyzing the Central Asian contribution to RLR and to their non-Turkic speaking neighbors.

In the present study, analyses revealed that Chikhi *et al.*'s (2001) method represents the closest estimates to the true Central Asian contributions. Based on this method, it was observed that there were lower male (13%) than female (22%) contributions from Central Asia to Anatolia, with wide ranges of confidence intervals. Lower contribution, with respect to males, is to be explained by homogenization between the males of the Balkans and those of Anatolia. In Azerbaijan this contribution was 18% in females and 32% in males.

Moreover, results pointed out that the Central Asian contribution in RLR can not be totally attributed to the language replacement episode because similar, or even higher, Central Asian contributions in northern and southern non-Turkic speaking neighbors were observed. The presence of a 20% or more admixture proportion in the RLR, and the presence of even higher contributions around the region, suggested that language might not be replaced inaccordance with "elite dominance model".

Keywords: Anatolia, Admixture, Genetic drift, Central Asia, Language replacement

ÖZ

BALKANLARA GÖRE ANADOLU GEN HAVUZUNA ORTA ASYANIN KATKISININ KARŞILAŞTIRMALI OLARAK ÇALIŞILMASI

Ceren Caner Berkman

Doktora, Biyolojik Bilimler
Tez Yöneticisi: Prof. Dr. İnci Togan

Eylül, 2006, 151 sayfa

Yaklaşık 1000 yıl önce Türkçe konuşan göçebe toplulukların Orta Asya'dan Türkiye ve Azerbeycan'a göç etmelerine paralel olarak bu bölgelerde Türk dili yayılmağa başlamış ve sonunda Türkçe bundan önce bölgede bulunan dillerin yerini almıştır.. Sunulan çalışmada, Orta Asya'nın Türkiye ve Azerbaycan'ın gen havuzlarına katkısının büyüklüğü farklı evrimsel faktörleri göz önünde bulunduran dört karışma analizi yöntemi ile incelenmiştir. Ayrıca, Orta Asya katkısının dil değişimi ile olan ilişkisini ayrıştırmak için çalışmada Türkiye ve Azerbaycan'la birlikte onların Türkçe konuşmayan komşularında bulunan Orta Asya katkıları da karşılatırmalı olarak incelenmiştir.

Çalışma sonuçları, incelenen popülasyonlar için genetik sürüklenmeyi göz önünde bulunduran Chikhi *ve ark.*'nın (2001) modeline göre elde edilen karışma oranlarının gerçeğe en yakın değerleri temsil ettiği yönündedir. Bu yönteme göre, Orta Asya'nın Anadolu'ya katkısının dişi (%22) ve erkeklerde (%13) farklılık gösterdiği görülmüştür. Anadolu'da erkeklere olan katkının dişilerinkinden az olması Balkanlar ve Anadolu arasında erkek ağırlıklı göçler sonucunda ortaya çıkan homojenleşme ile açıklanabileceği düşünülmüşdür Azarbeycan'da katkının dişilerde % 18, erkeklerde % 32 olduğu gözlenmiştir.

Çalışma ayrıca Orta Asya katkısının Türkiye ve Azerbaycan'ın kuzey ve güneydeki komşularında benzer düzeyde yada daha yüksek olduğunu göstermiştir. Bu nedenle, Anadolu ve Azerbeycan için elde edilen Orta Asya katkısının tümünün dil değişimi olayı ile ilişkilendirilmemesi gerektiği sonucuna varılmıştır. Türkçe konuşan Anadolu ve Azerbeycan'da %20 ya da daha çok Orta Asya katkısı belirlenmesi ve buna ek olarak Türkçe konuşmayan komşu bölgelerde de aynı ya da daha fazla katkı görülmesi dil değişiminin seçkin küçük bir grup tarafından değiştirildiği modeline uymıyabileceğini düşündürmektedir.

Anahtar Kelimeler: Anadolu, Karışma, Genetik sürüklenme, Orta Asya, Dil değişimi

To My Family

ACKNOWLEDGEMENTS

I would deeply like to express my appreciation to my supervisor Prof. Dr. İnci Togan for her guidance, attention, advises and encouragements for the completion of this thesis.

I am thankful to Can Ozan Tan for his helpful ideas and support while I am dealing with the admixture analysis.

I am sincerely thankful to my dear friend Dr. Sibel Turkşen Selçuk for her encouragement and comments.

I would also like to thank Çiğdem Gökçek who was always with me with her beautiful heart and never-ending support.

Finally, but most entirely, I would like to thank to my husband Ali Emre Berkman for his endless support and attention throughout this thesis.

TABLES OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

º	: Degrees
BCE	: Before Common Era
bp	: Base pair
CE	: Common Era
CRS	: Cambridge Reference Sequence
DNA	: Deoxyribonucleic acid
E	: East
GenBank	: NIH genetic sequence database
HVRI	: Hypervariable region I
Mb	: Mega base pairs
mtDNA	: Mitochondrial DNA
N	: North
nDNA	: Nuclear DNA
N_p	: Sample Size of Population
N_R	: Sample Size of Region
LGM	: Last Glacial Maximum
PCA	: Principle Component Analysis
PC	: Principle component
RNA	: Ribonucleic acid
SINEs	: Short interspersed elements
SNP	: Single nucleotide polymorphism
STR	: Short Tandem Repeat
ya	: Years ago

CHAPTER I

INTRODUCTION

1.1 Short History of Anatolia

Anatolia, the Asian part of Turkey, is at the junction between the Balkans, Near East and Caucasus. Because of its geographical location, Anatolia has acted as a bridge for numerous movements of modern human beings since very early times. In the study, the terms "Anatolia" and "Turkey" were used interchangeably.

Literature about the origin of our species accepts that the modern humans originated in Africa (see e.g., Lahr and Foley, 1998; Ingman *et al*., 2000; Underhill *et al*., 2001) and started to migrate out of Africa 50,000 years ago (ya) (Underhill *et al*., 2001). Modern humans reached the Near East and Anatolia around 40,000 ya from which they expanded west, north, and east (Underhill *et al*., 2001; Cavalli-Sforza and Feldman, 2003). In Central Asia, populations started to expand around 30.000 ya, reaching Europe, the Near East, and Northern Pakistan (Underhill *et al*., 2001). It is believed that, modern humans migrated to Europe first through Asia, followed by a second migration, through Anatolia (25,000-20,000 ya) (Underhill *et al*., 2001; Semino *et al*., 2000).

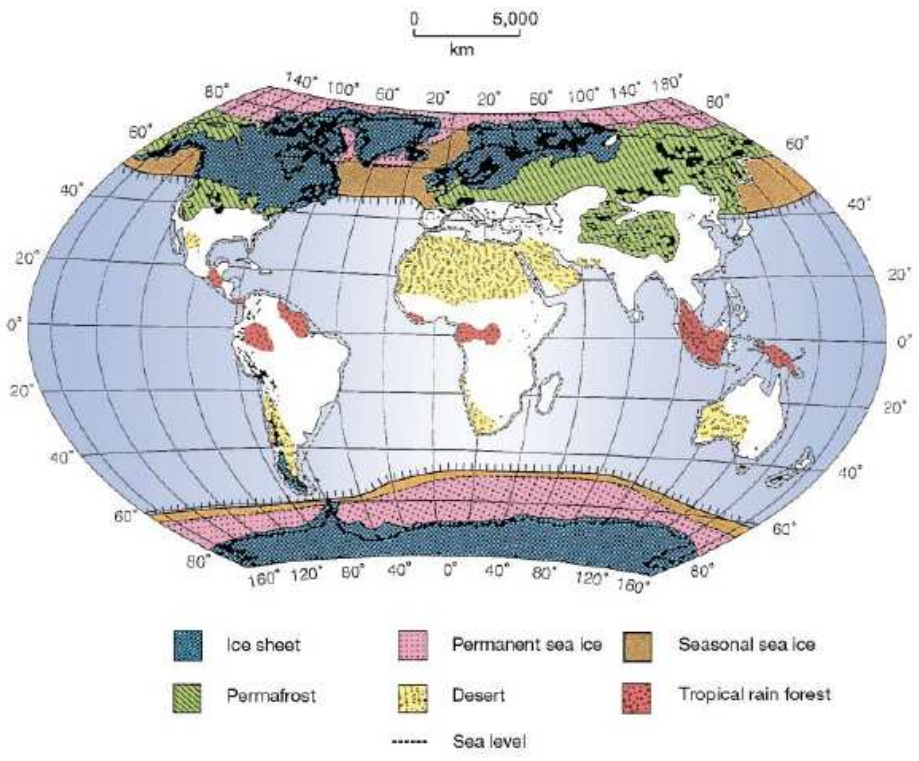


Figure 1.1: The maximum extent of ice sheets and permafrost areas around 20.000 ya. (Hewitt, 2000)

Climatic oscillations had an influence on the distribution of species (Hewitt, 2000; Jobling *et al*., 2004). Climatic conditions and changes in the distribution of plants and animals influenced the distribution of modern humans in turn. As can be seen in Figure 1.1, Northern Eurasia was covered by either an ice sheet or with permafrost around 20,000 ya. During this time, an ice sheet and permafrost together pushed the favorable area for the humans below 47º N in Europe (Hewitt, 2004). Therefore, at the Last Glacial Maximum, LGM, (18.000 – 16.000 ya), significant population contractions took place (Underhill *et al.,* 2001).

Together with Iberia, Anatolia became one of the refuges that modern humans could live during such harsh periods (Cinnioğlu *et al*., 2004). With the end of LGM, modern humans began to repopulate the areas that had previously been covered with ice sheets and permafrost, by moving north towards Europe and northwest into the Eurasian steppes.

The earliest communities to rely on farming emerged in the area known as the Fertile Crescent. As shown in Figure 1.2 the Fertile Crescent covers the area from the Zagros Mountains of Iraq, to the Southeastern regions of Turkey, Western Syria, Lebanon, and Israel (Cavalli-Sforza *et al*., 1994).

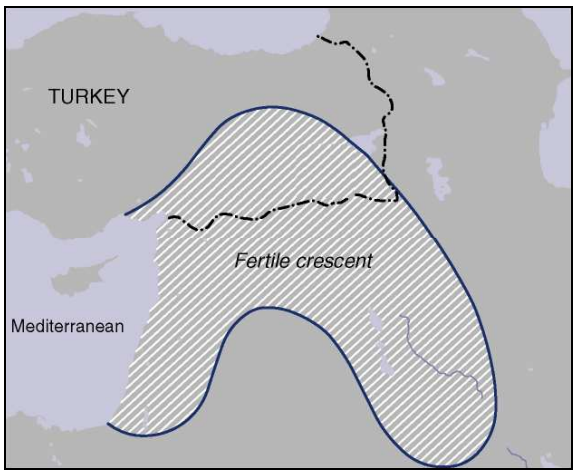


Figure 1.2: Fertile Crescent (Adapted from Jobling *et al*., 2004)

Çatal Höyük is one of the oldest settlement areas in Turkey. Excavations on this site have revealed the presence of developed agricultural communities living on Çatal Höyük from about 8500 to 7500 BCE (Akurgal, 2003).

In fact, the deep history of Anatolia belonging to the hunter-gatherer populations (Paleolithic age) and the farming populations (Neolithic age) can be seen at the 400 Paleolithic and 300 Neolithic sites listed in the Database of Archeological Sites in Turkey. The Paleolithic and Neolithic sites in Turkey are given in Figure 1.3.

Figure 1.3: Paleolithic and Neolithic sites in Turkey based on TAY Geographic Information System (http://taygis.tayproject.org/TAYGIS_ENG/TAYGISeng.html, retrieved July, 2006)

The Neolithic farmers of the Fertile Crescent started to grow in population size nearly 10,000 ya and spread into Europe and the Caucasus (Semino *et al*., 2000; Underhill *et al*., 2001; Cinnioğlu *et al*., 2004). Anatolia was an important reservoir for the farming industry as the farming culture spread through it towards Europe. Radiocarbon chronology of the spread of farming from Anatolia to Europe is given in Figure 1.4.

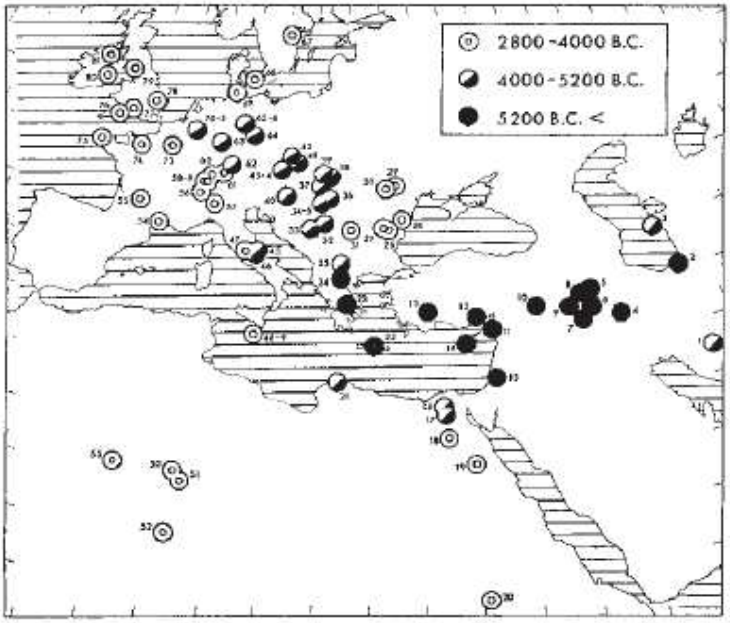


Figure 1.4: Radiocarbon dates for the earliest sites of farming settlements (Renfrew, 2000).

After the shift to sedentary life, Anatolia was populated by various civilizations, such as the Hattians, Hurries, Hittites, Phrygians, Lydians, Urartians, Persians, Meds, Romans, Sassanids, Byzantines, Seljuk Turks, and Ottomans (Akurgal, 2003).

The Hatti (25^{th} - 21^{th} Century BCE) and Hurrie (23^{th} – 21^{th} Century BCE) were the first states founded by the people living in Anatolia (Akurgal, 2003) whose languages show structural similarities with the Altaic language family (İnalcık, 1997). The Altaic language family includes the Turkic language family (Ruhlen, 1991). Hittites, the first Indo-European speaking population in Turkey (Renfrew, 1987), controlled most of Anatolia around 14^{th} Century BCE (Akurgal, 2003).

Origin of their migration (Caucasus or Balkans) is still not known (Umar, 1999; Akurgal, 2003). Starting from 13th Century BCE, several migrations took place from the Balkans to Anatolia such as the migrations of the Phrygians and Ionian Greeks (Akurgal, 2003). Together with Lydians and Medians, they (the Phrygians and the Ionian Greeks) became part of Achemenid Persia and then were controlled by Alexander's Empire. Control of the Indo-European speaking populations continued during the presence of Rome and Byzantium Empires (Tambets *et al.*, 2000).

The harsh climatic conditions of Eurasian steppes were not suitable for farming, thus making it necessary to rely primarily on pastoral, nomadic lifestyles (Manz, 1994). Domestication of horses and the use of wheeled vehicles (chariots) increased the mobility of the inhabitants (Calafell *et al.*, 2000) and allowed the development of more pronounced pastoral nomadism around 900 ya (Christian, 2001). Migrations of Cimmerians and Scythians from the Northern Black Sea region to Anatolia and Mesopotamia through the Caucasus were examples of these migrations (Christian, 2001).

Starting between the 5th - 7th Century CE, Central Asia was controlled by Turkic speaking nomadic groups (Roux, 1997). In the 6th Century CE, a nomadic force arose in Mongolia out of the union of Turkic speaking tribes, namely Göktürks (T'u-kü-e) (Roux, 1997). They were the first Turkic tribes to use the word "Türk" as a political name (Manz, 1994) and they controlled Central Asia until the rule of the Mongolian Empire (13th Century CE) (Manz, 1994). After the split of the Göktürk Empire, a group of Turkic tribes migrated west. They were called Oghuz. However, it was known that there were also unions of Turkic tribes called Oghuzs prior to the Göktürks, such as the Dokuz-Oghuz union that controlled the south and southwest region of Lake Baikal (Roux, 1997).

Around the 9th – 11th century CE Turkic speaking Pechenegs, Uz and Kipchaks, who occupied the region around Northern Black Sea, migrated to Eastern Europe and the Balkans (Roux, 1997; Salman, 2004).

Turkic tribes were not the only Asian tribes that entered Europe, the Near East and Anatolia. Around the 5th Century CE, the Huns, migrated west from Central Asia to the steppes of Eastern Europe, destabilizing the Germanic tribes and causing them to invade the Western Roman Empire in search of safer lands to settle. Furthermore, around the 13th Century CE, Genghis Khan brought Mongolian tribes together and started to extend the borders of the Khanate. Mongol troops eventually reached Eastern Europe, Southwest Asia and Near East (Rossabi, 1994).

In Anatolia, the well-known influence of Turkic speaking groups occurred around the 11th Century CE. As indicated before, beginning in the time of the Hittites, and lasting for centuries, Indo-European language was spoken in Anatolia (İnalcık, 1997; Akurgal, 2003). Turkic language was introduced recently (around 1000 years ago) with the invasion of Turkic speaking nomadic groups (Oghuz Turks) (Vryonis, 1971; Lewis, 1995). Forced by the Kipchaks, Oghuz Turks migrated mainly from their homeland, the area between the Caspian and Aral Seas (Vryonis, 1971; Lewis, 1995). One group traveled North of the Black Sea, through the Tuna River and entered to Balkans only to be destroyed by the European populations (Roux, 1997). The Seljuks (Kınık tribe of Oghuz Turks), who migrated from South of Caspian Sea, invaded and imposed their language onto the people of Turkey and Azerbaijan (Roux, 1997). Migrations of Turkic tribes did not cease after the arrival of Seljuks, instead they continued for more than two centuries (Vryonis, 1971; Roux, 1997). Oghuz Turks who entered Turkey and Azerbaijan were the founders of the Seljuk Dynasty and several other dynasties such as the White Sheep, Black Sheep and Ottomans.

1.2. Studies on Genetic Contribution of Central Asia to Anatolia

The episode of language replacement from Indo-European to Turkic language in Anatolia around 1000 ya (11th Century CE) might have been accompanied by a genetic contribution of the invaders to the existing Anatolian gene pool.

If the relatively few newcomers, who introduced the language, did not contribute much to the recipients' gene pool, the process would be described by the term "elite dominance" (Renfrew, 1987). If the newcomers did not have any genetic effect, the case is described by the term "pure-elite dominance" (Benedetto *et al.*, 2001). Furthermore, if the invading group is primarily male, then admixture estimates may have a sex-biased effect in favor of males (Benedetto *et al.*, 2001; Nasidze *et al.*, 2003).

Correspondance analysis based on protein markers (Brega *et al.*, 1998), phylogenetic analysis of mtDNA (Calafell *et al.*,1996; Comas *et al.*, 1996) and comparison of Y-chromosome haplogroup frequencies (Wells *et al.*, 2001) all indicate the relative genetic proximity of the Anatolian population to that of the European populations. Hence, these results pointed out that Central Asian populations had little genetic effect on the current day Turkish gene pool, thus supported the idea that the Turkic language was imposed in accordance with the model described by elite dominance. Rolf *et al.* (1999) analyzed mtDNA and Y-chromosome microsatellites with the median-joining phylogenetic network method and concluded that there might be a 10% east Asian genetic input in the Turkish gene pool. A more recent study, the study by Cinnioğlu *et al.* (2004) revealed that based on Y-chromosome markers, Anatolians shared most of the Y-chromosome haplogroups with those of Europe and the Near East, whereas there were few shared haplogroups with Central Asia and Africa. Furthermore, Cinnioğlu *et al.* (2004) estimated that the effect of recent migration of Turkic speaking nomadic groups might be lower than 9 %. Thus, supported the idea that language replacement was accompanied by low genetic input, whereas based on admixture analysis, Benedetto *et al.* (2001) determined 30% contribution from Central Asia to Anatolia for both males and females.

1.3. Admixture Analysis Methods

Contribution by migrations to the gene pool of populations can be partitioned using admixture analysis. In the simple admixture model shown in Figure 1.5, populations, over time, can be isolated from each other and thus evolve independently. The so-called parental populations can come into contact in several different ways:

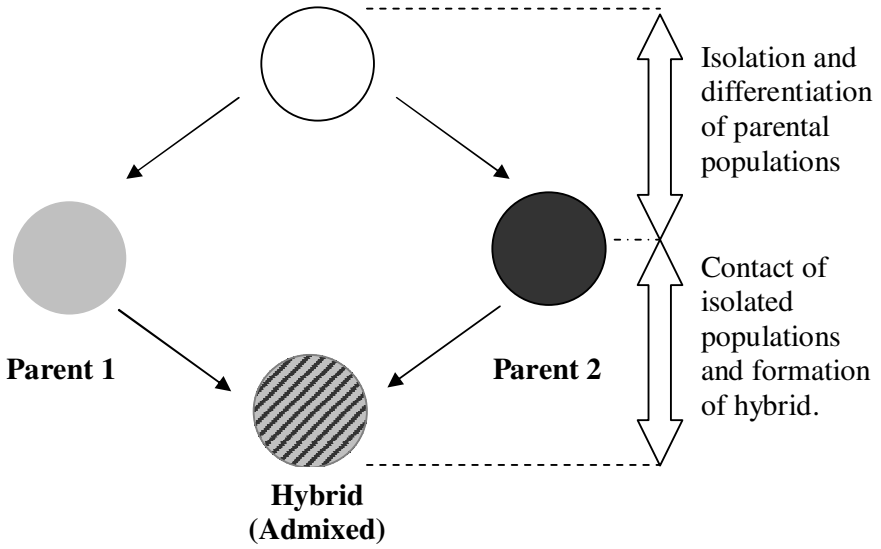


Figure 1.5: Schematic representation of genetic admixture

(1) For example, parental populations may produce a hybrid population by coming into contact through range expansion (Jobling *et al*., 2004). (2) Groups of individuals from both of the parental populations may migrate to a new area and form a new hybrid population there. (3) A group of individuals from one parental population may migrate into the territory of the other parental population and change the genetic make up of the second parental population (Choisy *et al*., 2004).

In general, when isolated populations, which are assumed to be the parental populations in the admixture model (Figure 1.5), come into contact, a genetic admixture occurs and a new hybrid (admixed) population is formed (Bertorelle and Excoffier, 1998; Chikhi *et al*., 2001; Dupanloup and Bertorelle, 2001).

One of the aims of admixture analysis is the determination of the proportional contribution of each parental population (admixture estimate) in the hybrid population. An important step in admixture analysis is the correct determination of parental populations. Methods could generate admixture estimates even if the parental populations were completely misidentified (Jobling *et al*., 2004). Therefore, while determining the parental populations, it is often required to find support from various disciplines such as physical and social anthropology, archeology, demography, and linguistics. Furthermore, the reliability of admixture proportions depends on the degree of differentiation of the parental populations (Bertorelle and Excoffier, 1998; Jobling *et al*., 2004).

Inferences about the past population processes, such as admixture, can be made by analyzing and interpreting either the current pattern of genetic variation or ancient DNA. However, since the data in terms of many different genetic markers and populations are available, the current patterns of genetic variations are being used to infer admixture proportions more frequently.

For the interpretation of the past population processes from current pattern of genetic variation, interaction of the various evolutionary forces such as migration, mutation and genetic drift must be considered. As it was indicated in Wang (2003), the admixture estimation procedure could be influenced by several factors:

1. As is evident for all genetic analysis, in admixture analysis, parental and hybrid populations are being represented by a small number of samples in comparison to the sizes of the real populations. Therefore, estimation errors can come from sampling (effect of sampling).

2. Since admixture events occurred in the past, genetic drift might influence the allele frequencies in parental and hybrid populations during the period between admixture and sampling events (effect of drift).

3. Allele frequencies can also be changed by the accumulation of mutations that have occurred since the admixture event, thus resulting in differentiation of parental and hybrid populations from each other (effect of mutation).

Many statistical methods (ex: Roberts and Hiorns' 1965; Long's 1991; Chakraborty *et al.*'s 1992; Bertorelle and Excoffier's 1998; Chikhi *et al.*'s 2001) have been developed to estimate admixture proportions from genetic data (Jobling *et al*., 2004). Methods differ based on the incorporation of the effect of sampling, genetic drift, and mutation. For example, the Robert and Hiorns' (1965) method ignores all of these factors (Jobling *et al.*, 2004), whereas the method of Chakraborty *et al*.'s (1992) incorporates the effect of sampling and drift only in the hybrid population. From the coalescent-based methods, the method of Bertorelle and Excoffier (1998) include the effect of sampling and mutations while the Chikhi *et al*.'s (2001) considers the effects of drift on hybrid and parental populations and also includes the effects of sampling.

1.4. Sex-Biased Admixture

Contribution of different sexes on the genetic structure of a hybrid population can vary if the males and females from parental populations contribute unequally (Jobling *et al*., 2004). Composition of the migrating group might result in unequal contribution of the parental populations in the genetic make up of hybrids.

For example, in male mediated migrations such as the military attacks or migrations of traders, only the paternal portion of the admixed population might be influenced. Furthermore, in some cases although both sexes arrive at the new region in similar numbers, one sex might have a greater chance to incorporate their genetic make up into that of the invaded population. Thus, directional mating, depending on the social characteristics of the parental and hybrid populations, might also cause unequal contribution of the males and females although they have migrated in equal numbers (Jobling *et al*., 2004). Therefore, while analyzing the evolutionary history of the admixed populations, it is necessary to study the evolutionary histories of maternal and paternal contributions separately. Comparative analyses of molecular markers having different inheritance patterns might be useful for determining the sex-based contributions of the parental populations.

1.5 Molecular Markers

Mitochondrial DNA (mtDNA) is inherited maternally and is used to follow the maternal lineage. The Y-chromosome shows the paternal inheritance pattern. Especially the non-recombining regions of the Y-chromosome are used to follow paternal lineages, whereas the autosomal markers such as the Alu insertions and autosomal microsatellites are inherited bi-parentally. They can give information about joint contribution of the two sexes (Jobling *et al*., 2004).

Since autosomal markers give information about the bi-parental inheritance, if a hybrid population sex-biased admixture is operating correctly, it is expected to observe the admixture estimates obtained from autosomal markers in between those obtained from mtDNA and Y-chromosome analysis. When there is a male mediated admixture, the admixture estimates obtained from different molecular markers will be in the following order: Y-chromosome > autosomal DNA > mtDNA. In contrary, in the female mediated admixture the event order will be reversed.

In human populations, about 85% autosomal genetic variation was found within continents and 10% was found between continents (Barbujani *et al*., 1997; Jorde *et al.*, 2000; Romualdi *et al*., 2002). Geographical variation increases by the use of mtDNA and Y-chromosome markers due to their smaller effective population sizes (Jobling *et al*., 2004).

Furthermore, as it is evident for all genetic analyses (Goldstein and Chikhi, 2002), admixture analyses based on single-locus lacks power (Chikhi *et al*., 2001; Dupanloup *et al*., 2004). However, analyzing mtDNA, Y-chromosome and autosomal markers, and combining the information coming from these different sources, increases the reliability of the analysis.

1.5.1 Mitochondrial DNA

Mitochondrial DNA (mtDNA) is a circular, double stranded DNA present in the mitochondria. Because of its characteristics, such as presence of high mutation rate, absence of recombination and its maternal inheritance pattern, mtDNA, especially its first hypervariable region, has been frequently used in evolutionary studies.

The control region (D-loop) of the mtDNA includes Hypervariable Region I (HVRI) which comprises the region between the nucleotide positions 16.024 -16.383 according to the Cambridge Reference Sequence (Anderson *et al*., 1981).

The mutational rate of coding and non-coding regions of the mtDNA differs. For example, general mutation rate for human mtDNA is about 3.4 x 10^{-7} (Ingman *et al*., 2000) whereas it is about 3.6 x10^{-6} for the HVRI (Richards *et al*., 2000).

1.5.2 Y-chromosome

The Y-chromosome is the male specific chromosome, which passes from father to son. More importantly, unlike other chromosomes in the human genome, except a region of three Mb, the Y-chromosome does not undergo meiotic recombination. Therefore, haplotypes usually pass unchanged from generation to generation, and preserve a simpler record of their history. A unique phylogeny of males can therefore easily be constructed (Jobling and Tyler-Smith, 2003). Hence, non-recombining property of Y-chromosome, like mtDNA, is important to determine the evolutionary history of organisms (Jobling and Tyler-Smith, 2003; Jobling *et al*., 2004).

1.5.3 Alu insertion Polymorphisms

Alu elements are the most abundant short interspersed elements (SINEs), which are approximately 300 bp in length and are found only in primates. They are ancestrally derived from the 7SL RNA gene (Ullu and Tschudi, 1984) and spread in the genome by retro-position (Shen *et al*., 1991). During the evolution of primates, the accumulation of Alu elements in the human genome resulted in groups of elements that are specific to humans. Studies on the Alu elements in humans that make up the 10% of the total genome (Batzer and Deininger, 2002) indicate that they are not distributed uniformly throughout the human genome (Deiniger *et al*., 1992).

Most of the Alu repeats have been integrated into the human genome recently. For this reason, they are generally dimorphic for the presence and absence of insertion and this makes them a useful source of genomic polymorphism (Batzer *et al*., 1991; Batzer and Deininger, 1991; Roy-Engel *et al*., 2001). The current rate of Alu insertion is estimated as one Alu insertion in every 200 births.

1.5.4 Autosomal Microsatellites

Microsatellite, also called short tandem repeat (STR), polymorphisms are composed of repeated sequences of two to five base pairs in length (such as ATATAT..). In microsatellites, new repeats occur due to DNA slippage during the DNA replication. The number of repeats in a microsatellite locus may vary between the individuals. They are highly polymorphic and densely distributed across the genome. They are mainly present in the non-coding regions of the genome. Based on these properties microsatellites have the potential to provide information about short-term evolutionary histories of the populations (Jorde *et al*., 1998; Zhivotovsky *et al*., 2003) such as population structures and differences, genetic drift, genetic bottlenecks and even the date of a last common ancestor by using relatively few loci (Bowcock *et al*., 1994).

1.6. Databases

The data obtained from molecular studies (ex: mtDNA and nDNA sequences, SNPs, Alu-insertion polymorphisms, STRs) are being collected in databases such as the National Center for Biotechnology Information, NCBI, (Benson *et al*., 2003)) , European Molecular Biology Laboratory, EMBL, (Stoesser *et al*., 2003) and DNA DataBank of Japan, DDBJ, (Miyazaki *et al*., 2003)).

These three databanks have formed the "International Nucleotide Sequence Database Collaboration" since 1982. They automatically update each other every 24 hours and share almost identical sets of sequences (Higgs and Attwood, 2005). Parallel to the improvement in the molecular genetic techniques, the amount of data accumulated in databases has also increased. Figure 1.6 shows the rapid, almost exponential, growth of the DNA sequence database (GenBank) of NCBI.

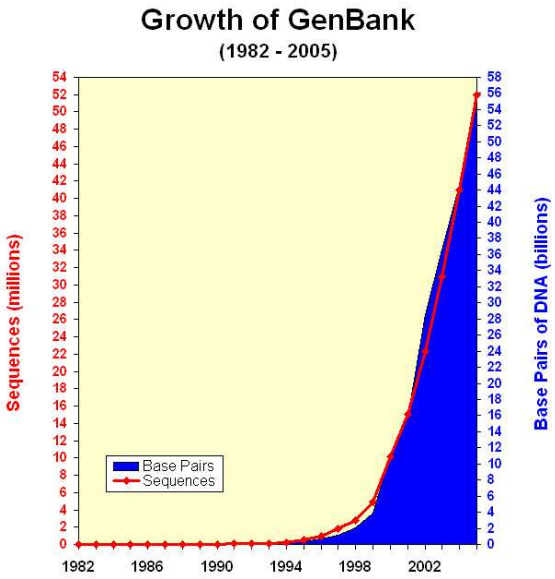


Figure 1.6: The growth of GenBank of NCBI between 1982 and 2005. (http://www.ncbi.nlm.nih.gov/Genbank/genbankstats.html, retrieved, August, 2006)

In addition to these three large databases, there are also databases for specific loci, molecular markers, and organisms. For example, HvrBase (Handt *et al.*, 1998) is a database that includes DNA sequence information for mtDNA HVRI and HVRII regions only for Apes, Neanderthals and modern humans. YHRD (Roewer *et al.*, 2001), database for human Y-chromosome microsatellites and Allele Frequency Database, ALFRED, (Rajeevan *et al.*, 2005) are other examples of such databases.

It is possible to use the raw data present in these databases to solve various biological questions.

1.7. Admixture Analysis Used for Estimation of the Central Asian Contribution to Anatolia

Benedetto *et al.* (2001) conducted the only study to use the admixture methods (3) to address the genetic consequences of recent migrations of Turkic speaking groups. In this study, they assume that the gene pools of the Kazakh, Kirghiz and Uighur populations are representing the parents of nomadic Turks whereas the Balkans (Bulgaria, Italy, Crete, Greece and Sicily) are used as the representatives of Turkish population before the invasion of Seljuk or, more general, the Oghuz Turks.

By combining the data analzed in the study from Turkey, along with other data collected from literature, Benedetto *et al*. (2001) used 146 mtDNA HVRI sequences, an average of 80 individuals for five Y-STR (DYS19, DYS390, DYS391, DYS392, DYS393) loci, and 590 individuals for autosomal microsatellite locus (TH01) from Turkey in an admixture analysis.

They tested the language replacement associated genetic effect in Anatolia with the help of three models shown in Figure 1.7. In the case of "pure elite dominance" model, they assumed that the gene flow from Central Asia into Anatolia was with very limited genetic contribution. The second model, which was named "instantaneous admixture", is also a type of elite dominance, in which the migrants were mainly composed of males (sex-biased admixture) and admixture was in a short time period; consequently resulting in a greater effect on Y-chromosome contribution. On the other hand, the third model, "continuous immigration", assumes that the language and the genetic make-up changed over time with a continuous gene flow. This time they expected to observe equally large admixture estimates for mtDNA and Y-chromosome analysis.

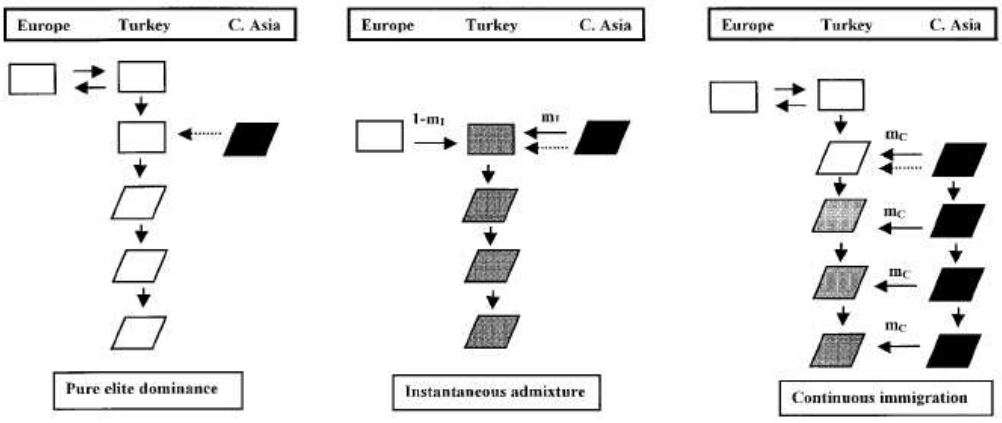


Figure 1.7: Schematic representation of the three models tested against DNA data in the study of Benedetto *et al*. (2001).

Rectangles: Indo-European speaking populations; **Lozenges:** Turkic-speaking populations.

Dashed arrows: linguistic transformations, **Horizontal solid arrows:** gene flow, **Vertical solid arrows:** inheritance, from older (top) to younger (bottom) generations.

Different shades of gray: the proportion of alleles of Central Asia in the Turkish allele pool

Based on the results obtained from mtDNA sequences, Y-chromosome, and autosomal microsatellite, they concluded that the male and female contributions from Central Asia to Anatolia were similar and around 30%. They attributed these admixture estimates mainly to the migrations of Oghuz Turks. The estimation indicated a huge Central Asian contribution had been integrated into the Turkish gene pool in one migration. Therefore, they concluded that after the language change the region became an important center, attracting Turkic speaking populations. Therefore, the language replacement, accompanied with a continuous gene flow at a rate of 1%, occurred for 40 generations.

1.8 Objectives of the Study

Objectives of the present study are:

1. To obtain accurate estimate(s) for the Central Asian contribution to the gene pool of Anatolian Turkish population with reference to Balkans

    a. By using the wealth of recently accumulated data on mtDNA, Y-chromosome and autosomal markers (Alu insertion polymorphisms and microsatellites) in many populations.
    b. By employing four admixture methods.

2. To ask if the calculated Central Asian contribution can be attributed solely to the language replacement episode by comparatively analyzing the Central Asian contribution to Turkey, Azerbaijan and to their eastern neighbors (Northern Caucasus, Armenia, Georgia, Syria, Iraq, Lebanon and Iran). For this purpose;

    a. Results obtained for Turkey and Azerbaijan were compared.
    b. Results obtained for Turkic speaking region (Turkey-Azerbaijan) were examined comparatively with the countries/regions speaking non-Turkic languages.

    Behind these comparative studies, the hypothesis is as follows: If Central Asian contribution was totally or mostly related with the language replacement episode, then contributions to Anatolia and Azerbaijan would be comparable with each other and they would be more than that of the non- Turkic speaking regions.

CHAPTER II

MATERIALS AND METHOD

2.1 Retrieved Data

All the data analyzed in the presented study was retrieved from databases and literature. In the study, Central Asia and the Balkans were accepted as the parental populations. Central Asia was composed of populations from, Kazakhstan, Kyrgyzstan, Uyghur, Altai, Uzbekistan, Turkmenistan, Tajikistan, together with the Khoremian Uzbek and Karakalpak populations, whereas Balkans were harboring populations from Greece, Bulgaria, Albania, Hungary and Romania. Admixed, hybrid populations were from Turkey, Azerbaijan, Armenia, Georgia, Northern Caucasus, Syria, Iraq, Lebanon and Iran. The regions for the collected data are indicated in Figure 2.1.

Data for the first hypervariable region of mitochondrial DNA (mtDNA HVRI) was collected from 2174 individuals from 26 populations associated with previously determined six regions. Data was retrieved mainly from NCBI (Benson *et al*., 2003) and HvrBase (Handt *et al*., 1998) databases between 2001 and 2005. The region 16.024 – 16.384 (with respect to Cambridge Reference Sequence, Anderson *et al*., 1981) mtDNA HVRI sequences were retrieved. Data sizes for each population, region and related reference are given in Table 2.1.

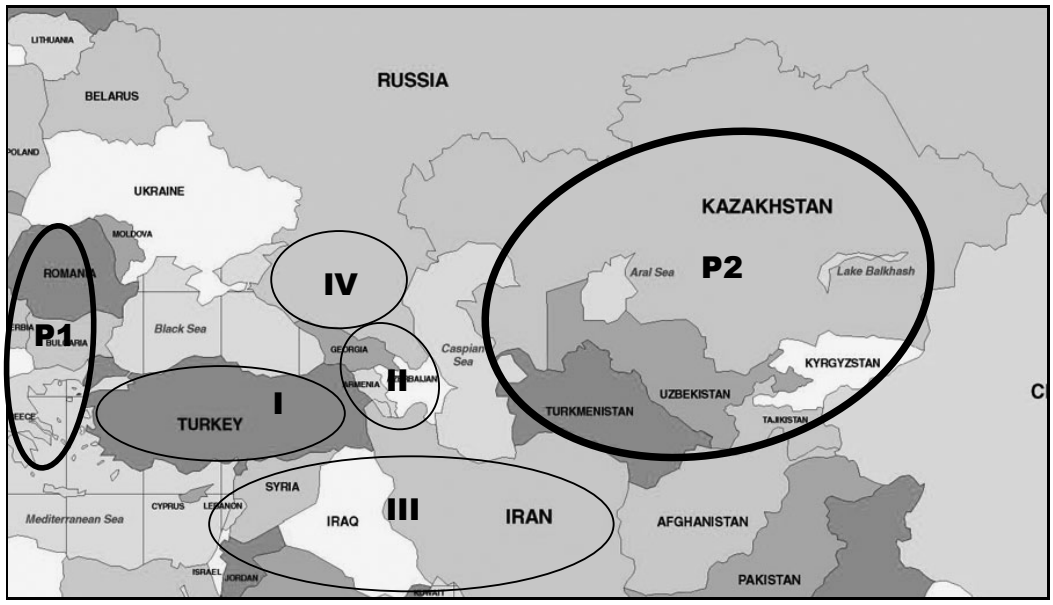


Figure 2.1: Map showing the regions that were used as the parental and hybrid populations in the presented study.

Parents: P1: Balkans (Greece, Bulgaria, Albania, Hungary and Romania), **P2:** Central Asia (Kazakhstan, Kyrgyzstan, Uyghur, Altai, Uzbekistan, Turkmenistan, Tajikistan, Khoremian Uzbek and Karakalpak) **Hybrids: I:** Turkey; **II:** Southern Caucasians (Armenia, Georgia, Azerbaijan); **III:** Near East (Syria, Iraq, Lebanon, and Iran); **IV:** Northern Caucasians (Ingushetia, Kabardino-Balkar, Abkhazia, Cherkessia, Chechnya, and Dagestan)

Table 2.1: List of employed populations and their sample sizes based on different molecular markers.

	Region	Population	mtDNA HVRI		Y-Chromosome haplogroups		Alu insertion polymorphisms		Autosomal Microsatellites	
			N_P	N_R	N_P	N_R	$2N_P$	$2N_R$	$2N_P$	$2N_R$
PARENTAL POPULATIONS	Balkans§	Greece	209	562	298	499	212	462	1495	2384
		Bulgaria	141		24		⁂		⁂	
		Albania	42		51		120		272	
		Romania	92		45		130		205	
		Hungary	78		81		⁂		412	
	Central Asia§	Uighur	117	453	134	1343	170	749	212	212
		Kazakh	105		112		155		⁂	
		Altai	17		51		203		⁂	
		Kirghiz	114		140		⁂		⁂	
		Tajik	20		190		129		⁂	
		Turkmen	20		68		⁂		⁂	
		Uzbek	20		648		92		⁂	
		Karakalpaks	20		⁂		⁂		⁂	
		Khoremian Uzbeks	20		⁂		⁂		⁂	
HYBRIDS	Turkey§	Turkey	290	290	813	813	474	474	3775	3775
	Northern Caucasus§	Ingushian	35	213	22	144	94	411	⁂	⁂
		Kabardinian	51		62		54		⁂	
		Abazian	23		14		38		⁂	
		Cherkessian	44		⁂		161		⁂	
		Chechenian	23		20		⁂		⁂	
		Darginian	37		26		64		⁂	
	Southern Caucasus	Georgia§	102	422	297	678	269	565	⁂	⁂
		Azerbaijan§	87		124		136		⁂	
		Armenia§	233		257		160		⁂	
	Near East	Syria§	118	234	111	407	137	137	⁂	⁂
		Iraq§	116		139		⁂		⁂	
		Lebanon§	⁂		104		⁂		⁂	
		Iran§	⁂		53		⁂		⁂	

⁂ no data was available § Populations which were used as parent or hybrid in admixture analysis, N_P: Sample size of populations N_R: Sample size of region. For Alu and autosomal microsatellites average numbers for the population sizes were given in the table.

Data retrieved from the following studies: **mtDNA** Shields *et al.* (1993), Comas *et al.* (1996), Comas *et al.* (1998), Calafell *et al.* (1996), Macaulay *et al.* (1999), Belledi *et al.* (2000), Comas *et al.* (2000), Richards *et al.* (2000), Lahermo *et al.* (2000), Yao *et al.* (2000), Benedetto *et al.* (2001), Vernesi *et al.* (2001), Kouvatsi *et al.* (2001), Nasidze and Stoneking. (2001), Comas *et al.* (2004a). **Y-Chromosome haplogroups:** Hammer *et al.* (1998), Karafet *et al.* (1999), Semino *et al.* (2000), Hammer *et al.* (2000), Rosser *et al.* (2000), Hammer *et al.* (2001), Karafet *et al.* (2001), Wells *et al.* (2001), Zerjal *et al.* (2002), Di Giacomo *et al.* (2003), Nasidze *et al.* (2003). Al Zahey *et.al.,*2003. Cinnioğlu *et al.* (2004), **Alu:** Nasidze *et al.* (2001), Antunez-de-Mayolo *et al.* (2002), Romualdi *et al.* (2002), Xiao *et al.* (2002), Khitrinskaya *et al.* (2003), Comas *et al.* (2004b), Mansoor *et al.* (2004), Dinç and Togan, 2005, Şekeryapan (2005). **Autosomal Microsatellites:** Iwasa *et al.* (1997); Takeshita 1997; Vural 1998; Brinkman *et al.* (1998); Szabo *et al.* (1998); Kondopoulou *et al.* (1999); Egyed 2000; Akbaşak et.al., (2001); Asicioğlu *et al.* (2002a); Asicioğlu (2002b); Çakır *et al.* (2002a); Çakır *et al.* (2002b); Filoğlu *et al.* (2002); Çerkezi *et al.* (2002); Sanchez-Diz 2002; Çakır 2003; Çetinkaya *et al.* (2003); Skitsa *et al.* (2003); Anghel *et al.* (2003); Çakır *et al.* (2004); Kubat *et al.* (2004); Ülküer 2004; Barbarii *et al.* (2004); Yavuz and Sarıkaya (2005); Zhu *et al.* (2005); Kovatsi *et al.* (2006).

To determine the male evolutionary history, Y-chromosome haplogroup data for 3884 individuals from 25 populations was retrieved from literature and databases between 2004 and 2005.

Furthermore, autosomal regions of the genome were analyzed by Alu insertion polymorphisms and autosomal microsatellites. Data for seven Alu-insertion polymorphisms (A25, B65, ACE, APO, PV92, TPA25 and FXIIIB) was retrieved from 18 populations by using the allele frequency database, ALFRED (Rajeevan *et al.,* 2005) and literature. Data for 12 autosomal microsatellites (TH01, VWA, TPOX, FGA, D13S317, D18S51, D2S11, D2S1338, D3S1358, D5S818, D7S820, and D8S1179) were also collected from the ALFRED database. In the analysis of autosomal microsatellites, because of the absence of data from Central Asia, only the Uighur population was used as a representative of this region.

2.2. Data Analysis

2.2.1. Multiple Sequence Alignment

To compare the DNA sequences, it is necessary to align the conserved and un-conserved sites across all of the sequences. In the presented study, retrieved sequences were aligned with ClustalW (Higgins *et al*., 1994), a multiple sequence alignment program, and the region of 275 base pair (between 16.090 and 16.365 of the Cambridge Reference Sequence, Anderson *et al*., 1981) was used in further analysis.

2.2.2. Measures of Molecular Diversity

Different measures of variation in DNA levels were calculated with the help of Arlequin 3.01 (Excoffier *et al*., 2005) and DISPAN (Ota, 1993) package programs. These are:

d. Number of different sequences (Haplotype Diversity)

A simple measure of DNA diversity is the number of different sequences in the sample. Different (polymorphic) sequences in a sample are called haplotypes, each haplotypes refers to a single or unique set of closely linked alleles (genes or DNA polymorphisms) inherited as a unit. The number of polymorphic sites and the associated haplotypes were determined with the Arlequin 3.01 package program (Excoffier *et al*., 2005).

e. Gene (Haplotype) Diversity

One of the ways of measuring the extent of variability in a population is to compute the gene diversity (mean expected heterozygosity). This statistic measures the probability that two haplotypes, drawn at random from a sample, are different from each other. Gene (haplotype) diversity (Nei, 1987) and its sampling variance are estimated as:

$$\hat{H} = \frac{n}{n-1}(1-\sum_{i=1}^{k} p_i^2)$$

$$V(\hat{H}) = \frac{2}{n(n-1)}\left(2(n-2)\left[\sum_{i=1}^{k} p_i^3 - \left(\sum_{i=1}^{k} p_i^2\right)^2\right] + \sum_{i=1}^{k} p_i^2 - \left(\sum_{i=1}^{k} p_i^2\right)^2\right)$$

Where n is the number of gene copies in the sample, k is the number of alleles (haplotypes) and p_i is the sample frequency of the i^{th} allele (haplotype).

The haplotype diversity was determined with the Arlequin 3.01 package program (Excoffier *et al.*, 2005). For Alu insertion polymorphisms and autosomal microsatellites, average heterozygosity values were calculated using the DISPAN package program (Ota, 1993).

f. Nucleotide Diversity

For DNA sequences, a measure of the diversity in a population is the average number of nucleotide differences per site between any two randomly chosen sequences. This measure is called the nucleotide diversity. It is the probability that two randomly chosen homologous nucleotides are different. The nucleotide diversity and the associated variance were determined with the Arlequin 3.01 package program (Excoffier *et al.*, 2005).

$$\hat{\pi}_n = \frac{\sum_{i=1}^{k}\sum_{j<i} p_i p_j \hat{d}_{ij}}{L}$$

$$V(\hat{\pi}_n) = \frac{n+1}{3(n-1)}\hat{\pi}_n + \frac{2(n^2+n+3)}{9n(n-1)}\hat{\pi}_n^{\,2}$$ Tajima, 1983

Where, d_{ij} is an estimate of the number of mutations that have occurred since the time divergence of haplotypes i and j. Furthermore, k is the number of haplotypes and p_i is the frequency of haplotype i.

2.2.3 Principle Component Analysis (PCA)

A way to generate a graphical representation of the relationship between populations is through the principle component analysis (PCA). PCA is a commonly used multivariate method to determine the relative positions of the populations on a two or three-dimensional space using frequency data. In the PCA, individual axes, known as principle components (PCs), are extracted sequentially.

The first PC explains the highest variation of all the data that can be accounted for by the compound axis; the second PC explains the next highest variation, and so on. Based on this, most of the variation was explained in the first two or three axes (Jobling *et al.*, 2004). The graphical representation of PCA of five populations shown in Figure 2.2 also indicates this.

In the proposed study, PCA was performed with the help of the NTSYS-pc2.1 package program (Rholf, 2000).

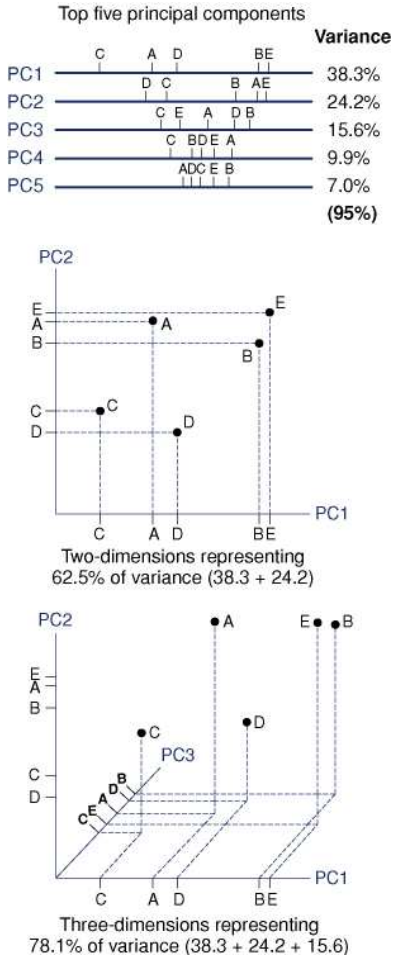


Figure 2.2: Graphical representation of PCA of five populations in two and three dimensions (Jobling *et al*., 2004).

2.2.4. Haplogroup Determination

A haplogroup is a cluster of similar haplotypes with variations on a common theme or "motif". These clusters are discrete groups of individuals who at some point in time shared a common ancestor.

Using the ancestral lineages of the haplotypes, i.e. haplogroups, may be more informative to determine historical events than using mtDNA in which high numbers of haplotypes with very low frequencies can be obtained.

However, since the mitochondrial phylogeny for Eurasia as a whole is not established yet, and since the sites which are most informative for identifying evolutionary relationships among sequences from the two continents is not exactly known, previously determined haplogroup motifs (Kolman *et al*., 1996; Starikovskaya *et al*., 1998; Macaulay *et al*., 1999; Richards *et al*., 2000; Benedetto *et al*., 2001) for Europe and Asia were tested. The data was classified in 33 groups based on HVRI motifs. The lists of motifs along with respective haplogroup motifs are given in Table 2.2.

Table 2.2: For the mtDNA Sequences, the list of motifs along with respective haplogroup motifs (based on Kolman *et al*., 1996; Starikovskaya *et al*., 1998; Macaulay *et al*., 1999; Richards *et al*., 2000; Benedetto *et al*., 2001).

	Haplogroup	Used haplogroup motifs and associated mutations
ASIA	M	16.223 C→T
	C	16.223 C→T/16.298 →C/16.327 C→T
	D	16.223 C→T / 16.362 T→C
	E	16.223C→T/16.227A→G/16.278C→T /16.362T→C
	A	16.223C →T/16.290C→T/16.319G→A /16.362T→C
	B	16.189 T→C / 16.217 T→C
	B5	16.140 T→C / 16.189 T→C
	F	16.304 T→C
BALKANS	CRS	
	V	16.298 T→C
	PRE-HV	16.126 C→T / 16. 362 T→C
	U1	16.189 T→C / 16.249 T→C
	U2	16.129 G→C
	U3	16.343 A→G
	U4	16.356 T→C
	U5	16.192C→T/16.256 C→T/16.270 C→T
	U7	16.318 A→T
	K	16.224 T → C / 16.311 T → C
	J1	16.126 T → C / 16.261 C→T
	J2	16.126 T → C / 16.193 C→T
	T	16.126 T→C/16.294 C→T/16.296 C→T
	T1	16.126 T→C / 16.163A →G / 16.186 C→T/16.189 T→C /16.294 C→T
	T2	16.126 T→C/16.294 C→T/16.304 T→C
	T3	16.126 T→C/16. 292C→T/16. 294 C→T
	T4	16.126 T→C/16.294C→T/16.324 T→C
	T5	16.126 T→C /16.153 G→A /16.294 C→T
	W	16.223 C→T / 16.292 C→T
	X	16.189 T→C/16.223 C→T/16.278 C→T
	I	16.129 G→A / 16.223 C→T
	R1	16.311 T→C
	L1	16.187C→T /16.189 T→C / 16.223 C→T/ 16.311 T→C
	L3a*	16.145G→A/16.176 C→G/16.223 C→T

For the Y-chromosome there is a detailed phylogeny (Y chromosome consortium, 2002). In the present study, the Y-chromosome haplogroup nomenclature was used according to the Y Chromosome Consortium (2002).

2.3 Admixture Analysis

In the presented study, the methods of Robert and Hiorns (1965), Chakraborty *et al.* (1992), Bertorelle and Excoffier (1998), implemented in ADMIX1.0 (Bertorelle and Excoffier, 1998) and the model of Chikhi *et al.* (2001) implemented in LEA package programs were used to determine the admixture proportions.

2.3.1. Robert and Hiorns' (1965) Method

The simplest equation to calculate the admixture proportion, μ, of parent 1 is as follows (Jobling *et al*., 2004).

$$\mu = \frac{(p_h - p_2)}{(p_1 - p_2)}$$

p_1, p_2 & p_h: allele frequencies of parental and hybrid populations

μ: proportional contribution of one of the parents

The Robert and Hiorns' (1965) method uses this relation but assumes that the estimates of admixture proportions from different alleles are related linearly (Jobling *et al.*, 2004). Based on this assumption, the method applies a least-square regression method and takes its gradient as the multi-locus estimate of μ (Jobling *et al.*, 2004).

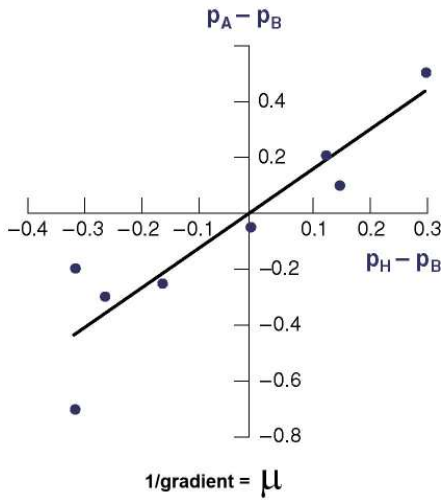


Figure 2.3: Least-Square method of Robert and Hiorns (1965). (Adapted from Jobling *et al*., 2004). Each dot represents μ_i obtained from i^{th} specific allele. Furthermore, μ is estimated by fitting a best line through the points.

2.3.2 Chakraborty *et al.*'s (1992) Method

The method of Chakraborty *et al.* (1992) is the extension of the method of weighted least-square admixture estimate of Long (1991). The method of Long (1991) again assumes that the allele frequencies of the hybrid population are linear combinations of the allele frequencies in the parental populations. However, in contrast to the previous one, Chakraborty *et al.*'s (1992) method introduces the effect of sampling errors in all populations but drift only in hybrid population. The formula for the admixture proportion, μ, of parent 1 is:

$$\mu = \frac{\sum_{j=1}^{r}\sum_{k=1}^{s_j+1}\left(p_{1jk} - p_{2jk}\right)\left(p_{hjk} - p_{2jk}\right)/E\left(p_{hjk}\right)}{\sum_{j=1}^{r}\sum_{k=1}^{s_j+1}\left(p_{1jk} - p_{2jk}\right)^2/E\left(p_{hjk}\right)}$$

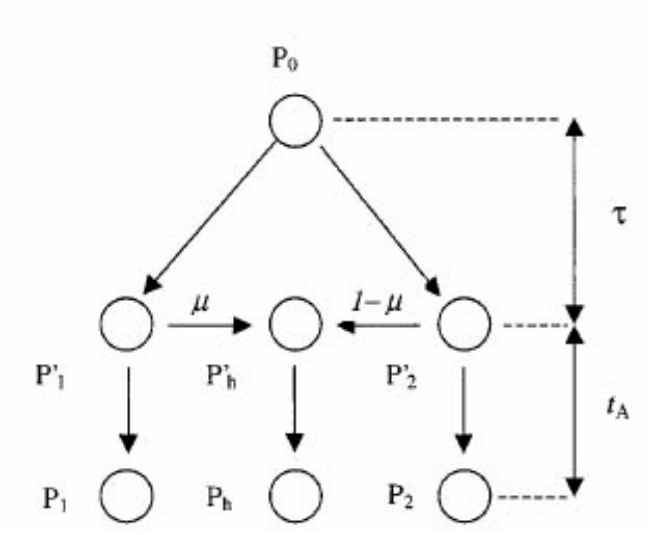
P_{ijk}: the frequency of k^{th} allele an the j^{th} loci in i^{th} parental population

μ: proportional contribution of one of the parents

$E(p_{hjk})$: expected allele frequency in hybrid

2.3.3. Bertorelle and Excoffier's (1998) Method

Bertorelle and Excoffier's (1998) method was used to determine the admixture proportions based on a coalescent approach. To determine the admixture proportions the method takes into account molecular information, i.e. the degree of dissimilarity in differences, as well as gene frequencies. Different data types (molecular markers) such as DNA sequences, restriction fragment length polymorphisms (RFLP) and microsatellites can be analyzed using t his method.



P_0: Hypothetical parental population

P_1', P_2' & P_h': Parental and Hybrid populations at the time of admixture

P_1, P_2 & P_h: Current day parental and hybrid populations

μ: proportional contribution of one of the parents

τ: time since admixture

t_A: time from the admixture event till today

Figure 2.4: Schematic representation of the Bertorelle and Excoffier's (1998) method.

This method, computes estimators of admixture coefficients based on the mean coalescent time of genes drawn either within or between admixed and parental populations. The estimated parameter is the admixture proportion of one of the parental populations (μ) and is estimated as:

$$\mu = \frac{c\hat{\bar{t}}_{h1} - d\hat{\bar{t}}_{h2} + d^2 + \hat{\bar{t}}_{12}(\hat{\bar{t}}_{h2} - \hat{\bar{t}}_{h1} + e)}{c^2 + d^2 + 2e\hat{\bar{t}}_{12}}$$

$$c = \hat{t}_A + \hat{\bar{t}}_{11}$$

$$d = \hat{t}_A + \hat{\bar{t}}_{22}$$

$$e = \hat{\bar{t}}_{12} - (c + d)$$

$$\bar{t}_{12} = t_A + \tau + \bar{t}_0$$

Since coalescent times between two genes are not directly available, mean coalescence times, $\bar{t}$'s, were estimated from genetic variability in this model.

Mean coalescence times for DNA sequences and RFLP data was estimated from the mean number of pairwise differences (π) based on the infinite site model in which it was assumed that each new mutation was occurring at a previously monomorphic site.

$$\hat{\bar{t}} = \pi / 2u$$

$$\mu = \frac{c\hat{\pi}_{h1} - d\hat{\pi}_{h2} + d^2 + \hat{\pi}_{12}(\hat{\pi}_{h2} - \hat{\pi}_{h1} + e)}{c^2 + d^2 + 2e\hat{\pi}_{12}}$$

$$c = \hat{t}'_A + \hat{\pi}_{11}$$

$$d = \hat{t}'_A + \hat{\pi}_{22}$$

$$e = \hat{\pi}_{12} - (c + d)$$

π_{11}, π_{22}: The mean number of pairwise differences within parental populations (P_1 and P_2 respectively).

π_{12} : The mean number of pairwise differences between parental populations.

π_{h1}, π_{h2} : The mean number of pairwise differences between hybrid and one of the parental populations (H & P_1 and H &P_2 respectively).

For the microsatellite data the mean coalescence times are estimated from the average squared differences in allele sizes ($\overline{S}$) based on the single-step stepwise mutation model in which it was assumed that each mutation could increase or decrease the allele size by a single repeat.

$$\hat{\hat{t}} = \overline{S} / 2u$$

$$\mu = \frac{c\hat{\hat{S}}_{h1} - d\hat{\hat{S}}_{h2} + d^2 + \hat{\hat{S}}_{12}(\hat{\hat{S}}_{h2} - \hat{\hat{S}}_{h1} + e)}{c^2 + d^2 + 2e\hat{\hat{S}}_{12}}$$

$$c = \hat{t}'_A + \hat{\hat{S}}_{11}$$

$$d = \hat{t}'_A + \hat{\hat{S}}_{22}$$

$$e = \hat{\hat{S}}_{12} - (c + d)$$

$\hat{\hat{S}}_{11}, \hat{\hat{S}}_{22}$: The average squared difference in allele size within parental populations (P_1 and P_2 respectively).

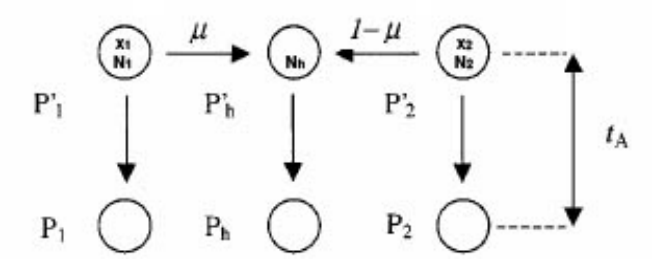
$\hat{\hat{S}}_{12}$: The average squared difference in allele size between parental populations.

$\hat{\hat{S}}_{h1}, \hat{\hat{S}}_{h2}$: The average squared difference in allele size between hybrid and one of the parental populations (H & P_1 and H &P_2 respectively).

Admix1.0 package program also calculates the standard deviations of the admixture estimates based on the bootstrap procedure (Bertorelle and Excoffier, 1998). In the present study, standard deviations are estimated by sampling with replacement 10,000 times.

2.3.4. Chikhi *et al.*'s (2001) Method

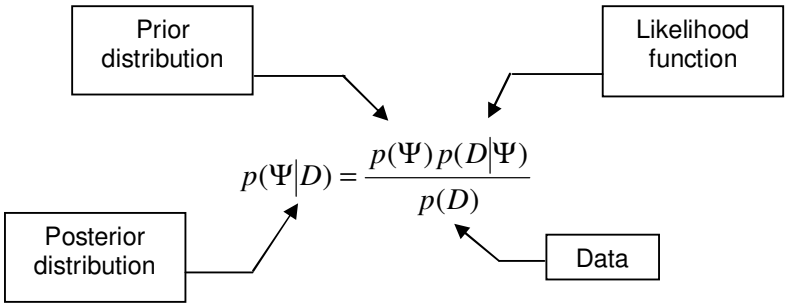
The method of Chikhi *et al.* (2001) was implemented in the LEA (Likelihood-based Estimation of Admixture) software. The model estimates the admixture proportion of one of the parents (μ) and the time since admixture (t_1, t_2, t_h) by applying a Bayesian (full-likelihood) and a coalescent based approach.



P_1', P_2' & P_h': Parental and hybrid populations at the time of admixture
P_1, P_2 & P_h: Current day parental and hybrid populations
N_1 & N_2: Sample sizes of the parental populations during admixture
x_1 & x_2 : allelic configurations of parental populations during admixture
μ: proportional contribution of one of the parents
t_A: time from the admixture event till today

Figure 2.5: Schematic representation of Chikhi *et al.*'s (2001) method.

In the Bayesian approach, inferences about a parameter (or a set of parameters), Ψ, are made by using the information provided through the observation of the data, D. This is shown by a probability density function:



The prior distribution, likelihood function, and posterior distribution are the three basic components in the Bayesian framework. The prior distribution describes analysts' beliefs, based on previous evidence, prior to the study. In Chikhi *et al.*'s (2001) method flat priors were used for μ , t_1, t_2, t_h and for x_1 and x_2 dirichlet distributions were used. By using these distributions as the priors, the model does not make any specific assumption about how genetically distant the parental populations are. In turn, this means that the model encompasses all possible histories of the parental populations. The likelihood function is a conditional distribution, which is defined as the distribution of one or more random variables when other random variables of a joint probability distribution are fixed at particular values. Based on a model of the underlying process, likelihood specifies the probability of the observed data given any particular values for the parameters (Beaumont and Rannala, 2004). The prior and likelihood functions combine all available information about the model parameters. The basic idea underlying the Bayesian approach is to calculate the posterior distribution by manipulating the joint distribution of the prior and likelihood functions in various ways to make inferences about the parameters given the data. The Bayesian approach is explained graphically in Figure 3.6.

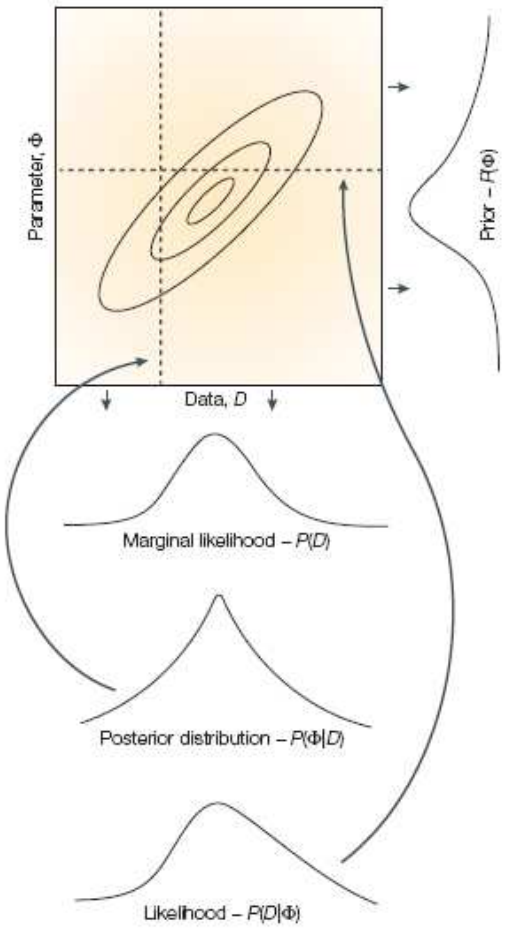


Figure 2.6: The basic features that underlay Bayesian inference (Beaumont and Rannala, 2004).

In the Chicki *et al.*'s (2001) method, the likelihood function is obtained using:

$$P(D|\mu,t_1,t_2,t_h,x_1,x_2) = p(a_1,a_2,a_h|\mu,t_1,t_2,t_h,x_1,x_2)$$
$$= \sum_{c_1c_2c_h}\sum_{f_1f_2f_h} ABC,$$

where;

$$A = p(a_1|f_1)p(a_2|f_2)p(a_h|f_h)$$
$$B = p(c_1|t_1,n_1)p(c_2|t_2,n_2)p(c_h|t_h,n_h)$$
$$c = p(f_1|x_1,c_1)p(f_2|x_2,c_2)p(f_h|\mu x_1 + (1-\mu)x_2,c_h)$$

a_1, a_2, a_h : sample frequencies in present day samples P_1, P_2, H.

f_1, f_2, f_h : founder frequency counts in P_1, P_2, H.

c_1, c_2, c_h: Number of coalescence in the genealogical history.

n_1, n_2, n_h: Sample size of P_1, P_2, H.

It was indicated that the number of allelic configurations among the founders, which is compatible with the data, could be very large and this might cause computational problems during the estimation of the likelihood function directly. Based on this, the formula was simplified as:

$$p(D|\Psi) = \int_{G,c} p(D|G)p(G|c)p(c|\Psi)dGdc$$

In this formula, *G* represents all possible genealogies consisting of a sequence of c coalescent events going backward from the time zero to time T and where the allelic frequency count among the lineages is recorded at each event.

In this method, to avoid the problem of analyzing all possible genealogies and allelic configurations, the Griffiths and Tavare (1994) algorithm was used. In this algorithm, Monte Carlo approach is applied to evaluate the likelihood at specific parameter values.

$$p(D|\Psi) = \int_{G,c} p(D|G)\frac{p(G|c)}{p^*(G|c)} p^*(G|c)p(c|\Psi)dGdc$$

$$p(D|\Psi) = \frac{1}{K}\sum_{1....K} p(D|G)\frac{p(G|c)}{p^*(G|c)}$$

$$\left.\begin{aligned} p(S_{k-1}|S_k) &= \frac{(n_{Ai}-1)}{(k-m)} \\ &= 0 \end{aligned}\right\} \begin{aligned} &\text{if } S_{k-1} = S_k - A_i \quad i = 1....m \\ &\text{Otherwise} \end{aligned}$$

where;

m= number of allelic types,

$A_i = i^{th}$ *allele,*

nA_i *= number of* A_i *alleles in the current state,*

$S_k - A_i$ *means that allelic configuration is identical to* S_k.

The waiting time is until the next coalescent is sampled from an exponential distribution. Under the coalescent model, the equivalent probability for each step in the chain is $(n_{Ai}-1)/(k-1)$. Thus, $p(G|c)/\,p^*(G|c)$ is obtained by multiplying each step by the ratio of these quantities, $(k-m)/(k-1)$.

In the model, the chain stops when the cumulative coalescence times become greater than the time of admixture event. The state at that time represents the allelic configuration among the founder lineages and is a random draw from the ancestral frequencies of the parental populations. To have an estimate of the likelihood of the sample, it is necessary to multiply the final probability by the probability of observing this founding state.

The convergences of the chains were tested using Gelman and Rubin Convergence Diagnostics (1992). Chains were run 100,000 steps for mtDNA, Alu insertion and autosomal microsatellites and 75,000 steps for the Y-chromosome.

The Griffiths and Tavare (1994) algorithm was used to calculate the likelihood $p(D|\mu,t_1,t_2,t_h,x_1,x_2)$ for specific values of μ,t_1,t_2,t_h,x_1x_2. However, to obtain information about these parameters (μ,t_1,t_2,t_h,x_1x_2), they should be sampled from the posterior distribution. To do this Markov Chain Monte Carlo (MCMC) method, using the Metrapolis-Hastings algorithm, was applied. In Monte Carlo simulations, samples X_i *(i= 1…..n)* of a random variable X are drawn from a distribution $\pi(.)$ and then used to evaluate functions of X. One method of doing this is by using Metrapolis-Hastings algorithm.

In Metrapolis-Hastings algorithm, X, is taken as the current state of the Markov chain in the parameter space defined by the model of interest. The algorithm first chooses a candidate for the next step of the chain, X_{t+1}, by using a proposal distribution $q(.|X_t)$. The chain then moves from state X_t to the candidate X_{t+1} with probability:

$$\alpha = \min\left(1, \frac{\pi(x_{t+1})q(x_t / x_{t+1})}{\pi(x_t)q(x_{t+1} / x_t)}\right)$$

The likelihood curves were constructed with R language.

2.4 Regression Analysis

To determine the relationship between the admixture estimates and geographical distances, regression analysis was used. The regression equations, statistical significance of the relationship, and regression graphs were constructed using the MINITAB13 package program (Minitab Inc., State College, PA, USA).

Two possible routes were assumed for the migrations from Central Asia. The first travels North of Caspian Sea, passing through Ural Mountains, and the other runs south of Caspian Sea, through Iran (Figure 2.7).

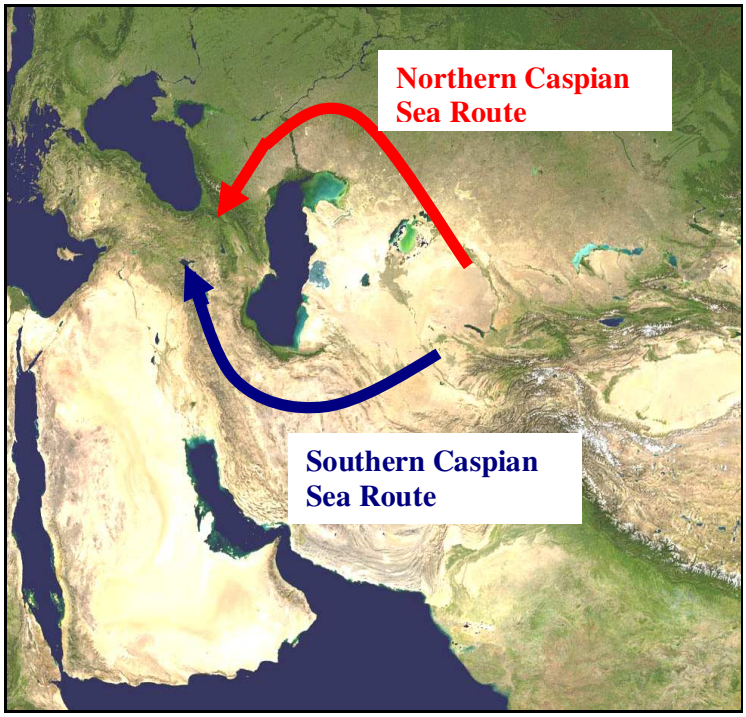


Figure 2.7: Possible migration routes from Central Asia.

The northern route was determined from the Barry center of Central Asia (45.1° N, 76.1°E) to Ural Mountains (56.51 °N, 60.34 °E), and from there to hybrids. In the same way, the southern route was determined from Central Asia to Iraq (33 °N, 44 °E) and from there to hybrids. The geographical coordinates of the hybrids are given in Table 2.3. For the estimations from regression lines the region that experienced the 'language replacement' was presented as the midpoint of the distance connecting the centers of Anatolia and Azerbaijan.

Table 2.3: Geographical coordinates of the hybrids

Hybrid	**Geographical Coordinates**
Language replacement region (Turkey and Azerbaijan)	39.65 ° N, 41.15 ° E
Armenia	40.00 ° N, 45.00 ° E
Georgia	42.00 ° N, 43.30 ° E
Northern Caucasus	43.50 ° N, 43.70 ° E
Syria	35.00 ° N, 38.00 ° E
Israel	31.30 ° N, 34.45 ° E
Iraq	33.00 ° N, 44.00 ° E

For each hybrid population, geographic distances were calculated from Central Asia based on great circle distances (d_{ij}). To calculate the distance, x_i and y_i are considered as the longitude and latitude of point *i*, the spherical distance between points *i* and *j* is calculated based on the formula:

$$\alpha = [\sin(y_i)]^2 + [\cos(y_i)]^2 \cos|x_i - x_j|$$

$$d_{ij} = R_E \tan^{-1}\left[\frac{\sqrt{1-\alpha^2}}{\alpha}\right]$$

where R_E is the radius of the Earth which is assumed to be 6379.34 km (Ramachandran *et al*., 2005 and references therein).

2.5. Verification of the Assumed Parents

In the present study, the Balkans and Central Asia are used as the predefined parental populations. The verification of the appropriateness of the composed parental populations was checked in two ways. First, parallel to the study of Dupanloup *et al.* (2004), the condition of using completely misidentified (random) parents in admixture analysis were simulated. For this, five simulation experiments (for the mtDNA data) were performed to form pseudo-samples with a sample size of at least 100. These pseudo-samples were in turn used as the parental populations in admixture analysis where Turkey was taken as the hybrid. The parental population combinations were also tested by excluding populations one by one from the parental population, and applying admixture analysis using Turkish population as the hybrid. In this way, the presence of an outlier population in the parents was tested.

2.6 Softwares Used in the Presented Study

The list of Statistical Softwares used in the presented study and their webpage addresses were as follows:

1. ClustalW: WWW Service at the European Bioinformatics Institute. http://www.ebi.ac.uk/clustalw, August, 2006.

2. Arlequin3.01: Department of Anthropology and Ecology, University of Geneva.
   http://lgb.unige.ch/arlequin, September, 2006.

3. DISPAN: Genetic distance and phylogenetic analysis. Pennsylvania State University.
http://iubio.bio.indiana.edu/soft/molbio/ibmpc, August, 2006.

4. NTSYSpc2.10q: Numerical Taxonomy System, Version 2.1. Exeter Software.
http://www.exetersoftware.com/cat/ntsyspc/ntsyspc.html, August, 2006.

5. ADMIX1.0: Inferring Admixture Proportion from Molecular Data. Department of Biology, University of Ferrara.
http://web.unife.it/progetti/genetica/Giorgio/giorgio_soft.html, July, 2006.

6. LEA: School of Animal and Microbial Sciences University of Reading.
http://www.rubic.rdg.ac.uk/~mab/software.html, August, 2006.

7. MINITAB13: Minitab Inc.
http://www.minitab.com/, September, 2006.

8. R 2.3.1: R-language.
http://www.r-project.org, August, 2006.

CHAPTER III

RESULTS

3.1. Mitochondrial DNA (mtDNA) Analysis

In the present study, 2174 mtDNA hypervariable region I (HVRI) sequences retrieved from databases were analyzed.

3.1.1. Multiple Sequence Alignment for mtDNA

Retrieved mtDNA HVRI sequences were aligned by employing CLUSTALW multiple sequence alignment software (Higgins *et al*., 1994) and the region of 275 base pairs (between 16.090 and 16.365 of the Cambridge Reference Sequence, Anderson *et al*., 1981) was used in further analysis.

3.1.2. Molecular Diversity Based on mtDNA HVRI

The molecular diversity of the mtDNA HVRI sequences were determined by using Arlequin 3.01 software (Excoffier *et al*., 2005).

Table 3.1: Populations used, together with their sample sizes, number of polymorphic sites, number of haplotypes, haplotype diversities, and nucleotide diversities for mtDNA HVRI sequence dataset.

Populations	Sample Size	Number of Polymorphic Sites	Number of Haplotypes	Haplotype Diversity	Nucleotide Diversity
Greece	209	82	114	0.976 ± 0.005	0.014 ± 0.008
Bulgaria	141	70	86	0.976 ± 0.007	0.015 ± 0.008
Albania	42	43	31	0.970 ± 0.018	0.018 ± 0.010
Romania	92	56	55	0.981 ± 0.005	0.015 ± 0.009
Hungary	78	62	63	0.988 ± 0.007	0.016 ± 0.009
Balkans	**562**	**121**	**236**	**0.964 ± 0.005**	**0.014 ± 0.008**
Kazakhstan	105	74	79	0.991 ± 0.004	0.023 ± 0.012
Kyrgyzstan	114	81	80	0.987 ± 0.005	0.022 ± 0.012
Altai	17	26	16	0.993 ± 0.023	0.020 ± 0.011
Uyghur	117	80	91	0.993 ± 0.003	0.021 ± 0.011
Tajikistan	20	41	19	0.995 ± 0.018	0.024 ± 0.013
Turkmenistan	20	35	16	0.963 ± 0.033	0.021 ± 0.012
Uzbekistan	20	37	19	0.995 ± 0.018	0.022 ± 0.012
Khoremian Uzbeks	20	35	17	0.984 ± 0.021	0.022 ± 0.012
Karakalpaks	20	43	19	0.995 ± 0.018	0.022 ± 0.012
Central Asia	**453**	**136**	**285**	**0.993 ± 0.001**	**0.022 ± 0.012**
Turkey	**290**	**113**	**198**	**0.986 ± 0.004**	**0.018 ± 0.010**
Abkhazia	23	32	19	0.980 ± 0.020	0.016 ± 0.009
Cherkessia	44	44	33	0.969 ± 0.017	0.016 ± 0.009
Chechnya	23	27	18	0.972 ± 0.022	0.015 ± 0.009
Dagestan	37	43	26	0.973 ± 0.015	0.017 ± 0.009
Ingushetia	35	27	23	0.950 ± 0.025	0.015 ± 0.008
Kabardino-Balkar	51	50	36	0.975 ± 0.011	0.016 ± 0.009
Northern Caucasus	**213**	**100**	**120**	**0.973 ± 0.007**	**0.016 ± 0.009**
Georgia	102	64	61	0.966 ± 0.011	0.017 ± 0.009
Azerbaijan	87	93	76	0.996 ± 0.003	0.021 ± 0.011
Armenia	233	112	152	0.987 ±0.004	0.019 ± 0.010
Southern Caucasus	**422**	**146**	**258**	**0.987 ± 0.003**	**0.019 ± 0.010**
Iraq	116	84	93	0.992 ± 0.004	0.020 ± 0.011
Syria	118	87	96	0.994 ± 0.003	0.019 ± 0.010
Near East	**234**	**104**	**189**	**0.996 ± 0.001**	**0.020 ± 0.011**
TOTAL	**2174**	**205**	**1033**	**0.989 ± 0.001**	**0.019 ± 0.010**

The number of polymorphic sites, the number of haplotypes and the nucleotide diversities for each population and region are given in Table 3.1. For the analyzed 2174 mtDNA HVRI sequences, 205 polymorphic sites defined 1033 haplotypes with a haplotype diversity of 0.989 ± 0.001. Frequencies of the haplotypes in each population are given in Appendix A.

The highest nucleotide diversities were observed for Central Asian group (0.022 ± 0.012). The next highest variant region was the Near East (0.020 ± 0.011) followed by the Southern Caucasus, Turkey, and Northern Caucasus. The lowest values were obtained for the Balkans (0.014 ± 0.008). The highest haplotype diversities were seen in the Near East (0.996 ± 0.001) and Central Asia (0.993 ± 0.001), whereas this value was just 0.964 ± 0.005 for Balkans.

For the Turkish sample, 290 sequences were analyzed. One hundred and nineteen polymorphic sites formed 198 haplotypes with 0.986 ± 0.004 haplotype and 0.018 ± 0.010 nucleotide diversities. The haplotype and nucleotide diversities of Turkish population were close to those values obtained for Southern Caucasus populations (0.987 ± 0.003; 0.019 ± 0.010 respectively).

3.1.3. mtDNA HVRI Haplogroups

For the collected mtDNA sequence data, haplogroups were determined based on the haplogroup motifs indicated in the studies of Kolman *et al.*, 1996; Starikovskaya *et al.*, 1998; Macaulay *et al.*, 1999; Richards *et al.*, 2000; Benedetto *et al.*, 2001. Mitochondrial DNA HVRI haplogroups and their observed numbers in parental and hybrid populations are given in Table 3.2.

Table 3.2: mtDNA HVRI haplogroups, and their observed numbers in parental and hybrid populations.

Haplogroups	Balkans	Central Asia	Turkey	Northern Caucasus	Georgia	Azerbaijan	Armenia	Iraq	Syria	Total
A	0	24	2	0	3	0	0	0	1	**30**
B	1	26	1	0	0	0	1	1	0	**30**
B5	0	8	0	0	0	0	0	0	0	**8**
C	0	42	4	8	1	3	0	0	0	**58**
D	3	81	8	9	1	0	0	0	1	**103**
E	0	21	0	1	0	1	0	0	0	**23**
M	2	16	1	3	0	0	0	1	1	**24**
F	0	7	2	0	0	1	0	0	0	**10**
CRS	66	22	31	31	17	4	23	8	10	**212**
I	12	7	8	4	2	2	3	0	0	**38**
W	12	8	7	5	0	2	7	0	2	**43**
X	11	3	12	11	4	2	5	2	0	**50**
K	21	6	17	8	8	3	14	6	6	**89**
V	15	7	2	4	2	5	0	0	2	**37**
R1	0	9	0	6	0	0	0	0	0	**15**
Pre-HV	7	2	2	0	0	0	1	5	4	**21**
J	29	6	11	10	2	2	15	6	6	**87**
J1	23	5	12	4	2	1	10	8	3	**68**
J2	2	4	7	2	0	1	6	1	1	**24**
T	4	7	10	11	9	4	7	3	5	**60**
T1	22	2	11	1	4	4	11	5	4	**64**
T2	6	6	2	3	0	1	5	1	3	**27**
T3	3	2	2	2	0	0	1	0	1	**11**
T4	1	0	2	1	0	2	1	0	0	**7**
T5	4	0	0	0	3	0	0	0	1	**8**
U1	5	11	11	11	1	0	10	1	5	**55**
U2	2	7	4	3	0	0	2	2	1	**21**
U3	11	0	17	13	2	1	10	6	8	**68**
U4	17	12	4	4	6	8	7	2	3	**63**
U5	48	7	3	15	3	4	7	1	2	**90**
U7	3	3	1	0	0	3	2	3	2	**17**
L1	1	0	1	0	0	0	0	3	2	**7**
L3	1	3	5	4	0	0	13	3	2	**31**
Total	**332**	**364**	**200**	**174**	**70**	**54**	**161**	**68**	**76**	**1499**

When the analyzed data set was examined, it was observed that 62 % of the Central Asian haplogroups were composed of A, B, B5, C, D, E, M and F haplogroups. On the other hand, this ratio was about 0.6 – 12 % in the rest of the regions. The haplogroups that had the highest frequencies in Central Asian populations were the haplogroups D and C (22.3 % and 11.5 %, respectively). The frequencies of these two haplogroups in Turkey and the Northern Caucasus were 6 % and 9.8 % respectively. In Southern Caucasus, these two haplogroups were observed as about 1.8 % and in the Balkans and Near East only haplogroup D was observed as 0.9 % and 0.5 % respectively.

Not all of the haplotypes could be grouped into specific haplogroups. The percentages of assigned sequences to specific haplogroups are given in Table 3.3. Only assigned sequences, 70% of the entire dataset, was used in the further analysis.

Table 3.3: Number of mtDNA sequences that could be assigned to specific haplogroups and hence used in the analysis.

Population	Number of Assigned Sequences to Haplogroups	Number of Available Sequences	Percent of Data Used in the Further Analysis
Balkans	332	562	59.1
Central Asia	364	453	80.4
Turkey	200	290	70.0
Northern Caucasus	174	213	81.7
Georgia	70	102	68.6
Azerbaijan	54	87	62.1
Armenia	161	233	69.1
Iraq	68	116	58.6
Syria	76	118	64.4
Total	**1499**	**2174**	**70.0**

3.1.4. Principal Component Analysis on mtDNA Haplogroups

To visualize the genetic relatedness of the examined populations, principle component analysis based on haplogroup frequencies was performed with the NTSYS-pc2.1 package program (Rholf, 2000). The constructed graphs in two dimensions are given in Figure 3.1.

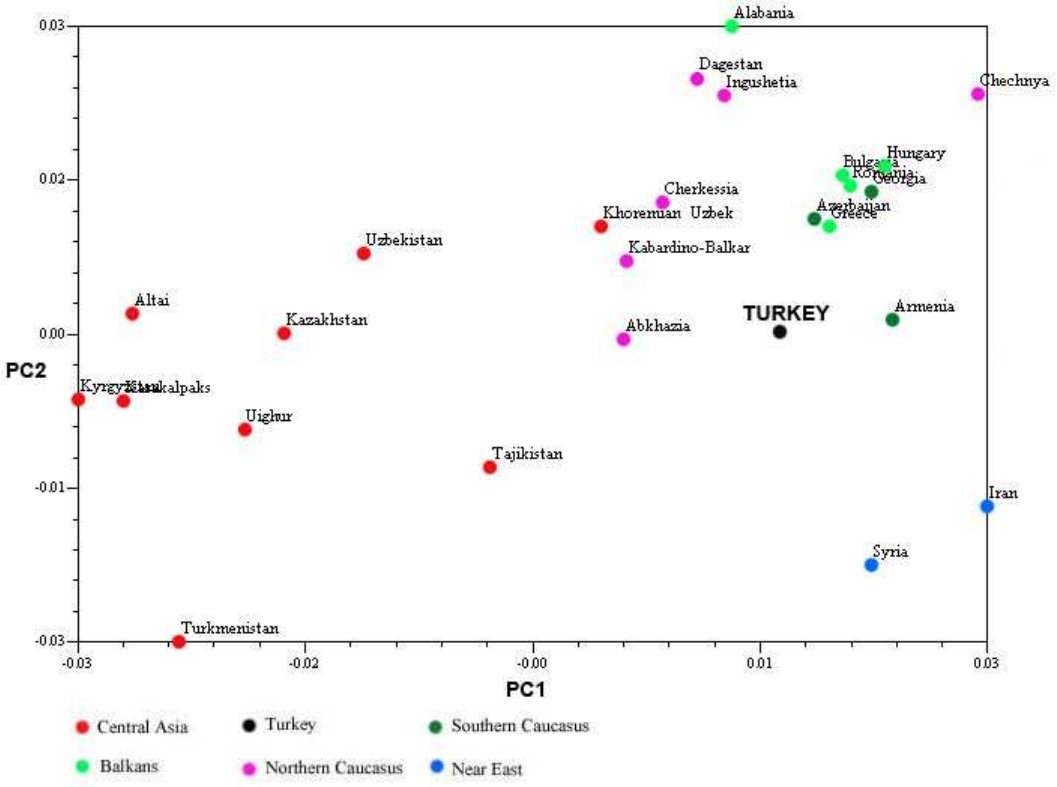


Figure 3.1: Two-dimensional plot of principle component analysis based on mtDNA haplogroup data. Colored dots indicate the populations from different regions.

The first principle component (PC1) covered 17.4 % of the overall variation exhibited by the haplogroups of the 275 bp mtDNA HVRI region. PC2 covered 12.9 %. Hence, the first two principal components together covered only the 30.3 % of the total variation in Figure 3.1. The PC3, which was not shown in the figure, covered further 8.2 % of the total variation. All together, the first three components explained the 38.5 % of the total variation. Compositions of the axes are given in Appendix B.

The weights of the variables (in Appendix B) indicated that on the first axis, Central Asian haplogroups C and D contributed most to the differentiation of the populations. They differentiated the Asian populations (Kazakhstan, Kyrgyzstan, Uyghur, Altai, Turkmenistan, Uzbekistan, Karakalpaks, and Tajikistan) from each other and from the others, whereas the second axis haplogroup CRS and Pre-HV differentiated Near Eastern populations (Syria and Iran) from Balkan, Turkish, Southern and Northern Caucasus population. In the third axis haplogroup V contributed the most. This third dimension was not shown but the haplogroup separated the Albania and Northern Caucasus (Ingushetia, Dagestan, Cherkessia, Abkhazia and Kabardino-Balkar) populations from rest of the populations analyzed.

At least on the first two axes, Balkan populations (Bulgarians, Greeks, Romanians and Hungarians) and two populations of Southern Caucasus (Azerbaijan and Georgia) could not be resolved from each other. Furthermore, in the two-dimensional plot, distinct intermediate positions of Northern Caucasus populations (Ingushetia, Dagestan, Cherkessia, Abkhazia and Kabardino-Balkar) were evident.

In addition, it could be seen that Turkey was relatively similar to some of the Northern Caucasian (Abkhazia and Kabardino-Balkar) and Southern Caucasian (Armenia) populations based on the mtDNA sequences.

3.2 Y-chromosome Analysis

3.2.1. Haplogroup frequencies for Y-chromosome

In the literature 19 main haplogroups for Y-chromosome biallelic markers were determined by several studies (see e.g., Semino *et al*., 2000; Underhill *et al*., 2001).

In the collected data set, 17 of these haplogroups were detected. Observed haplogroups, together with their observed numbers for each population and region are given in Table 3.4. When the whole data set was considered, the most frequently observed Y chromosome haplogroups were J (23.6 %) and R (23.7 %). The highest frequencies for haplogroup J were seen in the Near East (45 %), Turkey (33%) and Southern Caucasus (29.7%). Except D haplogroup, all haplogroups were observed in the Turkish population. In the collected data set, D haplogroup was presen in Asia with a low frequency (0.4 %). In Turkey, the highest frequencies were seen for J, R, E and G haplogroups (33%, 17.6%, 10.7% and 8.7% respectively). Haplogroups C and O were 18.2 % in Central Asia. These two haplogroups constitute only 1 % of the Turkish haplogroups. They were also seen in Iran (1.9 %), Northern Caucasus (1.4%), Lebanon (1%) and the Balkans (0.5 %).

Table 3.4: Y-chromosome haplogroups, and their observed numbers in different populations / regions.

Population	Balkans	Central Asia	Turkey	Northern Caucasus	Georgia	Azerbaijan	Armenia	Syria	Lebanon	Iran	Iraq	Total
A	0	0	2	0	0	0	0	0	0	0	0	**2**
C	1	161	7	2	0	0	0	0	1	0	0	**172**
D	0	15	0	0	0	0	0	0	0	0	0	**15**
E	99	28	87	5	6	8	13	13	31	10	17	**317**
F	3	102	13	38	46	17	24	9	23	22	1	**298**
G	12	5	68	29	43	13	11	3	1	0	3	**188**
H	0	14	3	0	0	0	0	1	0	0	0	**18**
I	49	36	29	22	4	3	7	1	1	0	1	**153**
J	85	159	268	21	93	43	65	55	35	12	81	**917**
K	3	65	35	14	5	11	14	7	1	1	12	**168**
L	2	45	23	0	1	0	3	0	2	2	0	**78**
N	1	10	45	0	0	1	3	8	0	0	0	**68**
O	0	83	1	0	0	0	0	0	0	1	0	**85**
P	65	128	38	8	21	5	35	1	0	3	0	**304**
R	126	482	143	5	43	22	52	13	9	2	24	**921**
Q	0	3	10	0	0	0	0	0	0	0	0	**13**
Y	52	7	41	0	35	1	30	0	0	0	0	**166**
Total	**498**	**1343**	**813**	**144**	**297**	**124**	**257**	**111**	**104**	**53**	**139**	**3883**

3.2.2 Molecular Diversity Based on Y-chromosome

The haplogroup diversity for the Y-chromosome haplogroups was determined using Arlequin 3.01 package program (Excoffier *et al*., 2005). The obtained results are given in Table 3.5.

Table 3.5: Populations used, together with their sample sizes, number of haplogroups, and haplogroup diversities for Y-chromosome dataset.

Region	Population	Sample Size	Number of Haplogroups	Haplogroup Diversity
Balkans	Greece	297	12	0.8330 ± 0.0072
	Hungary	81	8	0.6988 ± 0.0450
	Romania	45	6	0.8020 ± 0.0214
	Bulgaria	24	5	0.7717 ± 0.0593
	Albenia	51	6	0.7890 ± 0.0193
	Total	**498**	**12**	**0.8307 ± 0.0058**
Central Asia	Kazakhstan	105	10	0.7416 ± 0.0396
	Kyrgyzstan	140	11	0.6747 ± 0.0366
	Uyghur	141	13	0.8385 ± 0.0207
	Uzbekistan	648	15	0.8367 ± 0.0092
	Turkmenistan	68	8	0.8029 ± 0.0221
	Tajikistan	190	11	0.7276 ± 0.0276
	Altai	51	6	0.5898 ± 0.0664
	Total	**1343**	**16**	**0.8192 ± 0.0075**
Turkey	**Turkey**	**813**	**16**	**0.8308 ± 0.0085**
Northern Caucasians	Kabardino-Balkar	62	8	0.8355 ± 0.0220
	Ingushetia	22	6	0.7792 ± 0.0459
	Chechnya	20	7	0.8158 ± 0.0575
	Dagestan	26	6	0.6123 ± 0.0839
	Abkhazia	14	7	0.8462 ± 0.0614
	Total	**144**	**9**	**0.8359 ± 0.0129**
Southern Caucasians	Armenia	257	11	0.8491 ± 0.0099
	Azerbaijan	124	10	0.8106 ± 0.0214
	Georgia	297	10	0.8190 ± 0.0110
	Total	**678**	**11**	**0.8357 ± 0.0072**
Near East	Syria	111	10	0.7168 ± 0.0392
	Lebanon	104	9	0.7479 ± 0.0193
	Iran	53	8	0.7482 ± 0.0397
	Iraq	139	7	0.6120 ± 0.0387
	Total	**407**	**13**	**0.7369 ± 0.0211**

Haplogroup diversity for the Y-chromosome showed that the Southern (0.8357 ± 0.0072) and Northern Caucasus (0.8359 ± 0.0129) had the highest diversities in the analyzed dataset. The diversity estimates for the Balkans (0.8307 ± 0.0058) and Turkey (0.8308 ± 0.0085) were slightly lower than those for the populations of the Caucasus and they were very similar to each other. In contrast to mtDNA analysis, the lowest estimates were obtained for Central Asia (0.8192 ± 0.0075) and the Near East (0.7369 ± 0.0211).

3.2.3. Principal Component Analysis Based on Y-chromosome Haplogroups

Relative positions of the populations in two-dimensional space based on Y-chromosome haplogroup frequencies were obtained with the principle component analysis (Figure 3.2). The NTSYS-pc2.1 package program (Rholf, 2000) was used to construct the plot.

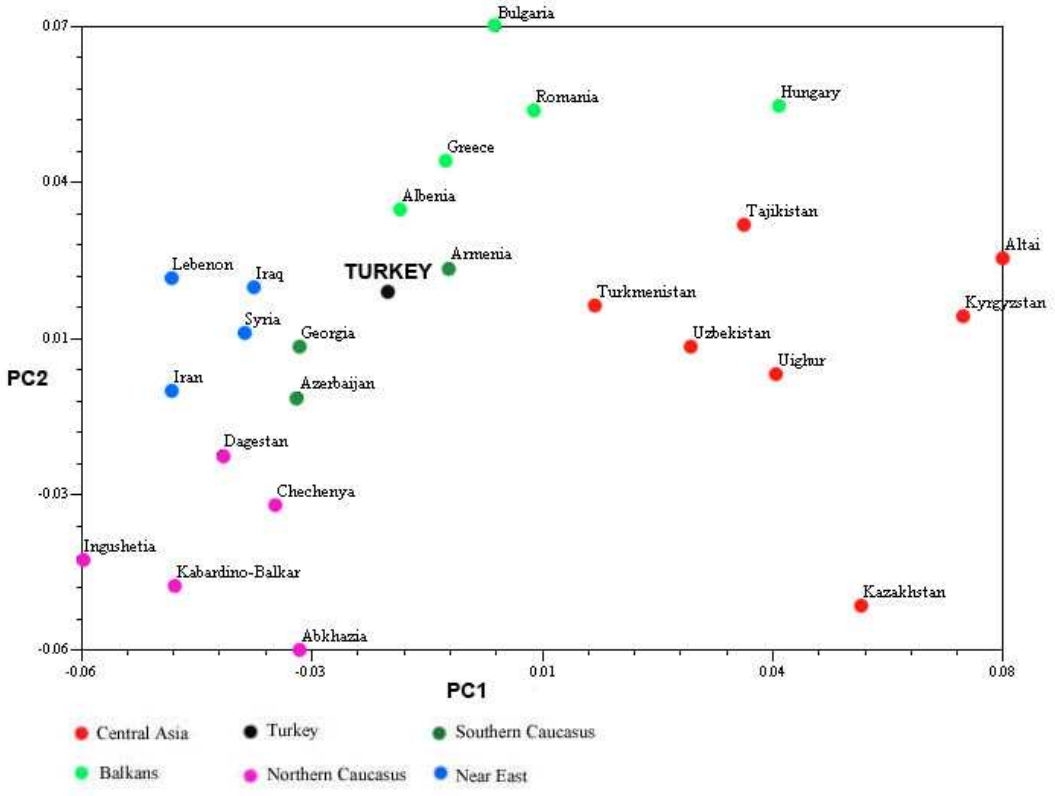


Figure 3.2: Two-dimensional plot of principle component analysis based on Y-chromosome haplogroup data. Colored dots indicate the populations from different regions.

The first principle component (PC1) covered 20.9 % of the overall variation. PC2 covered 16.0 % whereas PC3 covered 12.4 % of the total variation. Based on these, the first two components, shown in Figure 3.2, displayed 36.9 % of the total variation.

The weights of the haplogroups are given in Appendix C. The weights of the variables indicated that on the first axis R, D and O contributed most to the differentiation of the populations and differentiated the Asians from the Caucasians and Near Easterners. On the other hand, in the second axis haplogroup K differentiated populations of all groups but especially differentiates the Northern Caucasus from the Balkans. In the third axis (not shown) haplogroup A, contributes the most. It differentiates Near Easterners (Iraq, Syria, and Lebanon) from other populations.

It could be seen that differentiation of the populations based on the Y-chromosome haplogroups was better on the first two principle components compared to those based on mtDNA HVRI. Furthermore, Turkish the population was relatively close to the Southern Caucasian and Near Eastern populations.

3.3. Alu-insertion Polymorphism Analysis

3.3.1 Alu-insertion Frequencies and Molecular Diversity

Alu insertion frequencies for seven Alu insertion polymorphisms, together with the average heterozygosity values calculated with the DISPAN package program, (Ota, 1993) are given in Table 3.6. For each locus, the frequency of the allele with Alu insertion is given in the Table 3.6. When average heterozygosity (gene diversity) estimates for the regions were compared, it was observed that the highest diversity was seen in Central Asian populations with 0.394 ± 0.061.

Table 3.6: Alu insertion frequencies and average heterozygosities for populations/regions.

Population	A25	B65	ACE	APO	PV92	TPA25	FXIIIB	Heterozygosity
Albenia	0.075	0.667	0.467	1.000	0.200	0.558	0.600	0.345 ± 0.076
Romenia	0.031	0.569	0.469	0.915	0.200	0.577	0.408	0.362 ± 0.070
Greece	0.039	0.647	0.320	0.961	0.102	0.555	0.500	0.321±0.076
Balkans	**0.060**	**0.619**	**0.403**	**0.959**	**0.159**	**0.554**	**0.498**	**0.345 ± 0.071**
Kazakh	0.072	0.346	0.601	0.924	0.550	0.533	0.533	0.390 ± 0.065
Uighur	0.082	0.429	0.565	0.947	0.547	0.529	0.776	0.372 ± 0.067
Tajik	0.128	0.590	0.433	0.893	0.399	0.553	0.527	0.416 ± 0.053
Uzbek	0.100	0.585	0.598	0.870	0.467	0.511	0.554	0.416 ± 0.057
Kirghiz	0.087	0.546	0.630	0.904	0.539	0.470	0.808	0.375 ± 0.059
Central Asia	**0.083**	**0.476**	**0.572**	**0.909**	**0.535**	**0.517**	**0.663**	**0.394 ± 0.061**
Azerbaijan	0.000	0.694	0.218	0.942	0.382	0.514	0.102	0.294 ± 0.074
Armenia	0.060	0.452	0.476	0.873	0.013	0.428	0.340	0.332 ± 0.078
Georgia	0.089	0.726	0.348	0.933	0.248	0.493	0.611	0.358 ± 0.058
Southern Caucasians	**0.059**	**0.639**	**0.353**	**0.919**	**0.215**	**0.479**	**0.430**	**0.359 ± 0.063**
Kabardinian	0.111	0.433	0.267	0.932	0.145	0.290	0.139	0.312 ± 0.051
Cherkessian	0.048	0.675	0.302	0.935	0.274	0.446	0.253	0.347 ± 0.067
Darginian	0.029	0.317	0.162	0.870	0.162	0.368	0.150	0.292 ± 0.053
Ingushia	0.065	0.209	0.337	0.939	0.127	0.221	0.000	0.230 ± 0.061
Northern Caucasians	**0.055**	**0.489**	**0.265**	**0.925**	**0.180**	**0.351**	**0.167**	**0.310 ± 0.057**
Syria	0.000	0.307	0.400	0.926	0.176	0.507	0.283	0.324 ± 0.072
Turkey	0.064	0.594	0.384	0.965	0.197	0.434	0.509	0.352 ± 0.071

The heterozygosity values of Turkey (0.352 ± 0.071) and Southern Caucasus (0.359 ± 0.063) were very close to each other and slightly lower than that of Central Asian. The lowest values were obtained for the Northern Caucasus as 0.310 ± 0.057.

3.3.2. Principal Component Analysis Based on Alu-insertion Polymorphisms

Principle component analysis was also employed for the Alu-insertion polymorphism data. The PCA graph, obtained with NTSYS-pc2.1 package program (Rholf, 2000) for populations, is given in Figure 3.3.

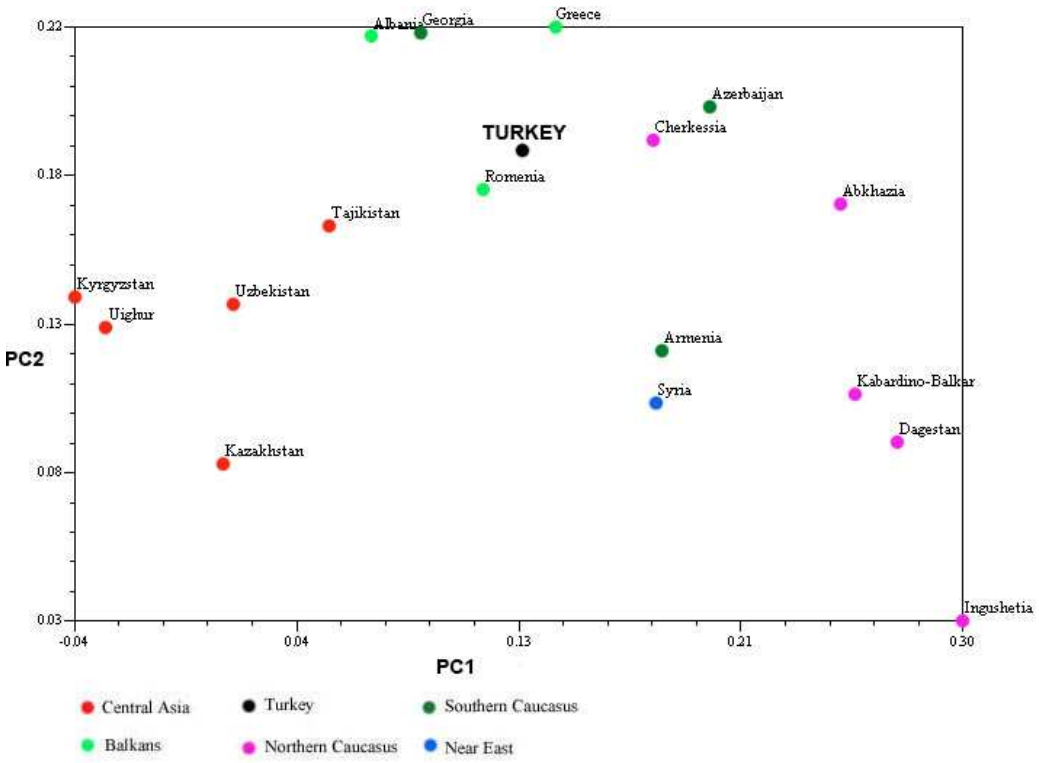


Figure 3.3 Two-dimensional plot of principle component analysis based on Alu insertion polymorphism dataset. Colored dots indicate the populations from different regions.

The first principle component (PC1) covered 46.0 % of the overall variation. PC2 covered 21.5 %. Hence, the variation displayed in Figure 3.3 was 67.5 %. In addition to this, PC3 covered 13.3 % (not shown) of the total variation and together, the first three components explained 80.8% of the total variation. The weighs of the variables (Appendix D) indicated that on PC1, FXIIIB and ACE, on PC2 B65 and on PC3 A25 contributed most to the differentiation of the populations. On the first axis, Central Asians (Kazakhstan, Kyrgyzstan, Uzbekistan, Tajikistan, and Uighur) were differentiated especially from the populations of Northern Caucasians (Cherkessia, Abkhazia, Kabardino-Balkar, Dagestan and Ingushetia). The second axis separated especially populations from Northern Caucasus but not the regions.

3.4. Autosomal Microsatellite Analysis

3.4.1 Allele Frequencies of Autosomal Microsatellites

Data for 12 autosomal microsatellites namely TH01, VWA, TPOX, FGA, D13S317, D18S51, D2S11, D2S1338, D3S1358, D5S818, D7S820, D8S1179 were collected from databases. Alleles and frequencies seen in populations for these autosomal microsatellites are given in Tables 3.7 – 3.9.

Table 3.7: Alleles for TH01, TPOX, D13S317, D8S1179, D5S818, D7S820 and their observed numbers in different populations\regions.

Loci	P*	Alleles															
		5	**6**	**7**	**8**	**9**	**9.1**	**9.3**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
	Balkans	3	678	355	423	512		549	40	71	8						
	Uighur	0	33	53	16	61		42	7	0	0						
TH01	Turkey	5	1476	927	651	1127		920	178	15	0						
	Balkans		7	6	1326	267			173	667	73	1					
	Uighur		0	0	128	11			6	55	12	0					
TPOX	Turkey		7	18	2135	423			318	1056	127	2					
	Balkans			3	269	190			107	641	601	168	71	0			
	Uighur			0	34	24			26	61	47	16	4	0			
D13S317	Turkey			2	676	455			346	1444	1547	430	137	3			
	Balkans			0	41	34			171	142	285	732	486	294	68	10	3
	Uighur			0	3	1			26	9	17	63	48	31	9	5	0
D8S1179	Turkey			2	68	44			269	215	386	1081	871	593	188	32	7
	Balkans			6	6	83			175	632	719	397	26	5	1		
	Uighur			0	1	11			24	76	68	30	1	0	1		
D5S818	Turkey			3	16	145			254	821	914	455	36	4	0		
	Balkans		0	36	343	235	1		572	500	299	54	10	0			
	Uighur		0	3	44	15	0		48	48	49	5	0	0			
D7S820	Turkey		1	79	685	379	0		976	961	637	111	13	2			

*Population/region

Table 3.8: Alleles for D2S11 and their observed numbers in different populations\regions.

Loci	P*	Alleles																					
		24.2	**25**	**26**	**27**	**28**	**28.2**	**29**	**29.2**	**30**	**30.2**	**31**	**31.2**	**32**	**32.2**	**33**	**33.2**	**34**	**34.2**	**35**	**35.2**	**36**	**38**
D2S11	Balkans	0	1	9	76	312	69	464	69	377	83	143	231	51	188	11	76	0	4	1	1	0	0
	Uighur	0	0	0	0	15	1	58	0	63	6	10	18	1	31	0	8	0	1	0	0	0	0
	Turkey	1	7	7	74	563	2	870	5	804	116	221	411	34	452	6	158	1	20	2	0	1	1

*Population/region

Table 3.9: Alleles for D18S51, VWA, D2S1338, D3S1358, FGA and their observed numbers r in different populations\regions

Loci		D18S51			VWA			D2S1338			D3S1358			FGA		
P*		Balkans	Uighur	Turkey	Balkans	Uighur	Turkey	Balkans	Uighur	Turkey	Balkans	Uighur	Turkey	Balkans	Uighur	Turkey
Alleles	**9**	0	0	7												
	10	13	0	25												
	10.2	0	0	1												
	11	43	3	60	5	0	14									
	12	261	17	461	1	0	1				3	0	3			
	13	336	29	525	12	0	120	0	0	0	5	0	8			
	13.2	0	0	13												
	14	368	51	727	335	27	549				319	11	256			
	14.2	0	0	13												
	15	287	36	527	350	14	698	1	0	3	850	81	999			
	16	288	27	497	656	42	1450	50	1	89	894	59	1015	1	0	0
	17	220	15	333	942	69	1730	223	21	435	765	42	832	0	0	1
	18	162	10	254	609	44	1140	79	37	250	609	18	591	36	5	34

Table 3.9 continued

Loci		D18S51			VWA			D2S1338			D3S1358			FGA		
P*		Balkans	Uighur	Turkey	Balkans	Uighur	Turkey	Balkans	Uighur	Turkey	Balkans	Uighur	Turkey	Balkans	Uighur	Turkey
Alleles	**18.2**													0	1	0
	19	90	11	163	268	14	442	88	35	270	57	1	49	199	4	270
	20	50	7	76	59	2	73	131	34	300	4	0	3	345	11	418
	20.2													3	1	3
	21	24	6	36	0	0	5	25	4	69				512	34	799
	21.2							0	0	0				14	1	20
	22	19	0	21	0	0	2	17	11	94				512	35	750
	23	4	0	4				124	27	262				25	1	7
	22.3													1	0	0
	23													403	44	777
	23.2													19	2	8
	24	0	0	3				104	22	170				341	46	718
	24.2													6	0	5
	25	1	0	0				78	16	130				212	18	410
	25.2													3	0	1
	26							17	4	22				80	8	166
	27							1	0	11				17	1	37
	28							0	0	3				5	0	9
	29													0	0	2
	30													0	0	1

3.4.2. Molecular Diversity for Autosomal Microsatellites

The diversity estimates and associated standard deviations were calculated with the help of the DISPAN package program (Ota, 1993). The results are given in Table 3.10.

Table 3.10: Average heterozygosity values for autosomal microsatellites.

Region	Population	Heterozygosity
Balkans	Albania	0.730 ± 0.068
	Romania	0.612 ±0.108
	Greece	0.730 ± 0.069
	Total	**0.803 ± 0.020**
Central Asia	**Uighur**	**0.786 ±0.024**
Turkey	**Turkey**	**0.803 ± 0.019**

In the autosomal microsatellites from Central Asia, only the Uighur population was used due to the absence of data. It was observed that the average heterozygosity values for Turkey (0.803 ± 0.019) and the Balkans (0.803 ± 0.020) were very close to each other, whereas small sample sized Uighur had lower estimate.

3.4.3. Principal Component Analysis Based on Autosomal Microsatellites

The relatedness of the populations was visualized in two dimensions. Principle component analysis was performed with the NTSYSpc21 package program (Rholf, 2000). The obtained PCA graphs were given in Figure 3.4.

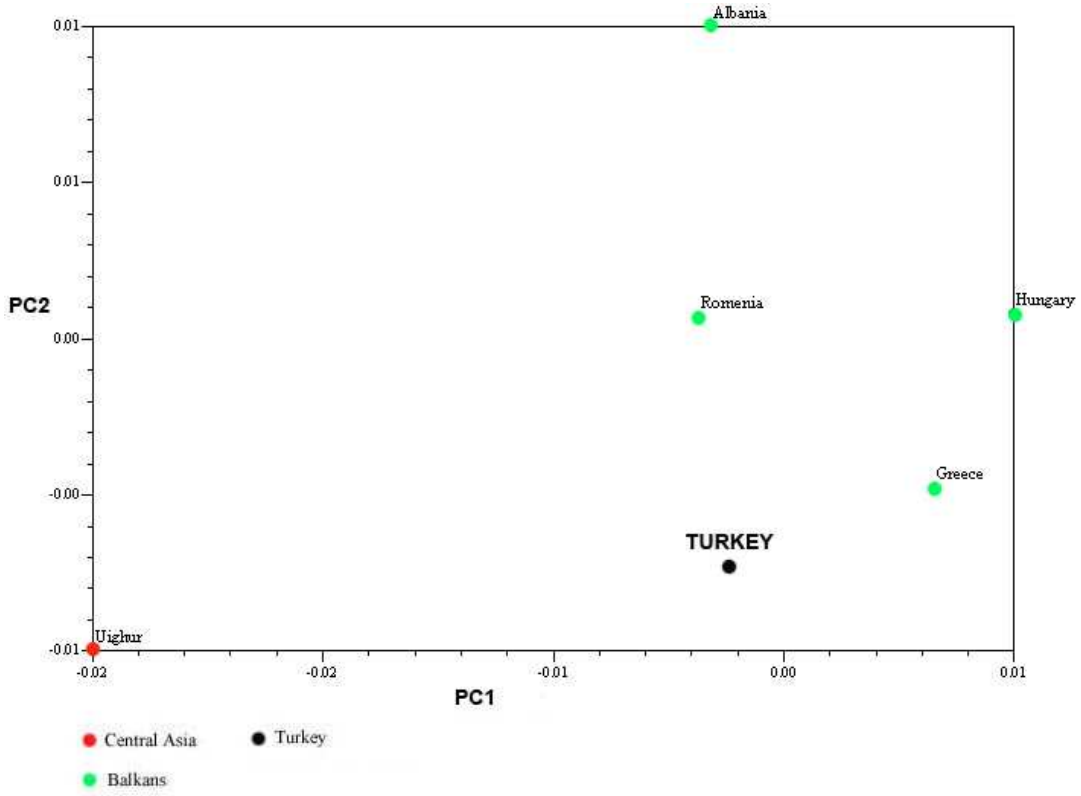


Figure 3.4: Two-dimensional plot of principle component analysis based on autosomal microsatellite dataset. Colored dots indicate the populations from different regions.

When PCA was applied to the data of autosomal microsatellites, it was observed that 31.7% of the total variation was explained by the first axis (PC1), while 24.7% was explained with PC2 and 22.3% was explained with PC3. In the Figure 3.4, the first two components that make up the 56.7% of the total variation are given. When the third component (not shown) was also considered, the amount of explained variation increased to 78.8 %.

The weights for the first three components of the alleles were given in Appendix E. It was observed that in the first component the allele 10 of the D13S and 12 of the TPOX, in the second principal component allele 24 of the D2S13 and in the third component allele 26 of D2S13 contributed to the differentiation of the populations from each other. PC1 differentiated Uighur from other populations.

3.5. Admixture Analysis

3.5.1 Admixture Estimates Obtained by Different Methods

Central Asian contribution to Turkey was calculated using Robert and Hiorns' (1965), Chakraborty *et al.*'s (1992), Bertorelle and Excoffier's (1998) and Chikhi *et al.*'s (2001) admixture methods. The estimated Central Asian admixture proportions ($\mu_{R\&H}$, μ_{CY}, $\mu_{B\&E}$, and μ_C respectively) were given in Table 3.11. The estimates for the proportional contribution of Central Asia in the Turkish gene pool were examined with the mean and median values of the posterior distribution from Chikhi *et al.*'s (2001) method. Due to the skewness of the distributions, the means were higher than the medians for all molecular markers tested for Turkish population.

From the Table 3.11, it could be seen that for a molecular marker, the admixture estimates were different with respect to each different admixture estimation method. For example, for the mitochondrial DNA the lowest estimate was $\mu_{R\&H}$ (19 %) whereas the highest one was 38 % using the method of Chakraborty *et al.* (1992). A similar range (22 % - 31 %) was obtained for autosomal microsatellite analysis with the Robert and Hiorns' (1965), and Bertorelle and Excoffier's (1998) methods respectively. A narrower range was obtained for the other autosomal marker (Alu insertions). Estimators for Alu insertion polymorphisms ranged from 9 % (μ_{CY}) - 15 % (μ_C, median).

Table 3.11: Central Asian admixture estimates, their 95% confidence interval (CI) for Turkey based on different methods.

	Robert & Hiorns (1965)		**Chakraborty *et al.* (1992)**		**Bertorelle & Excoffier (1998)**		**Chikhi *et al.* (2001)**		
Admixture estimatorsϒ	$\mu_{R\&H}$	**95% CI**†	μ_{CY}	**95% CI**†	$\mu_{B\&E}$	**95% CI**†	$\mu_{C\ Mean}$	$\mu_{C\ Median}$	**95% CI**†
mtDNA	0.19	0.09– 0.30	0.38	0.10 – 0.66	0.33	0.12 – 0.55	0.24	0.22	0.03– 0.56
Y-chromosome	0.14	-0.03–0.31	0.16	0.03 – 0.28	0.64	0.49 – 0.79	0.17	0.13	0.0 – 0.52
Alu	0,10	-0.02–0.23	0.09	-0.02– 0.20	0,10	-0.03– 0.24	0.19	0.15	0.01– 0.62
Autosomal STR*	0.22	0.17– 0.27	0.38	0.13 - 0.63	0.31	0.07 - 0.55	0.26	0.24	0.06- 0.42

† These intervals represent the values between the 0.025 and 0.975 quartiles for *P*1, respectively.
* Uighur population was used as the representative of Central Asian parental population.
ϒ $\mu_{R\&H}$: Robert and Hiorns' (1965); μ_{CY}: Chakraborty *et al.*'s (1992), $\mu_{B\&E}$: Bertorelle and Excoffier's (1998), μ_C: median of Chikhi *et al.*'s (2001) admixture estimator.

On the other hand, the most drastic difference between the estimates obtained through different admixture methods was obtained for the Y-chromosome analysis. The range was 13% (μ_C, median) - 64 % ($\mu_{B\&E}$). However, when the estimate obtained from the method of Bertorelle and Excoffier (1998) was excluded the range decreased to 13 % (μ_C, median) - 16 % (μ_{CY}). Furthermore, it was observed that in general, the estimates obtained from Robert and Hiorns' (1965) method were quite close to the ones obtained from Chikhi *et al.* (2001). Whereas these two were lower than the ones obtained from Bertorelle and Excoffier's (1998) and Chakraborty *et al.*'s (1992) admixture methods.

Table 3.12: Comparisons of admixture estimates (i) when only the Uighur population and (ii) when Kazakhstan, Kyrgyzstan, Uighur, Altai, Tajikistan, Turkmenistan, and Uzbekistan populations were representing the Central Asian parental population for mtDNA, Y-chromosome and Alu insertion polymorphisms.

	mtDNA			Y-chromosome			Alu		
	μ		% difference	μ		% difference	μ		% difference
	P1	P2		P1	P2		P1	P2	
Robert & Hiorns (1965)	0.190	0.231	- 4.1	0.136	0.169	+ 3.3	0,103	0.078	- 2.5
Chakraborty *et al.* (1992)	0.380	0.364	- 1.6	0.158	0.134	- 2.4	0.091	0.081	- 1.0
Bertorelle & Excoffier (1998)	0.334	0.310	- 2.4	0.642	0.728	+ 9.6	0.104	0.09	- 1.4
Chikhi *et al.* (2001)	0.219	0.257	+ 3.8	0.131	0.160	+ 2.9	0.151	0.186	+ 3.5

P1: Kazakh, Kirghiz, Altai, Uighur, Tajik, Turkmen, Uzbek; P2: Uighur
% difference: Absolute differences with respect to P1

In the analysis, because of the absence of autosomal microsatellite data, the Uighur population represented Central Asia. For the other markers, when the analysis was repeated by taking the Uighur population as the representative of Central Asian parent, the estimates did not changed more than 10%. Results were given in Table 3.12. Furthermore, if Bertorelle and Excoffier's (1998) Y-chromosome estimate was excluded the absolute difference between the estimates were lower than 4%.

It was also determined that when analysis was repeated with the Uighur population as the Central Asian parent, the admixture estimates (μ_C median values) increased in all molecular markers using the method of Chikhi *et al.* (2001).

Table 3.13: Central Asian Admixture Estimates in hybrid population for mtDNA, Y-chromosome and Alu insertion polymorphisms.

HYBRID	Admixture Estimators[ɾ]	mtDNA		Y-chromosome		Alu	
		μ	95% CI[†]	μ	95% CI[†]	μ	95% CI[†]
Turkey	$\mu_{R\&H}$	0.190	0.085 – 0.295	0.136	-0.034 – 0.306	0,103	-0.018 – 0.225
	μ_{CY}	0.380	0.099 – 0.662	0.158	0.034 – 0.282	0.091	-0.021 – 0.203
	$\mu_{B\&E}$	0.334	0.120 – 0.548	0.642	0.485 – 0.790	0,104	-0.028 – 0.236
	μ_{C}	0.219	0.034 – 0.555	0.131	0.006 – 0.520	0.151	0.009 – 0.616
Azerbaijan	$\mu_{R\&H}$	0.189	0.020 – 0.357	0.400	0.150 – 0.649	-0.115	-0.329 – 0.099
	μ_{CY}	0.621	0.309 – 0.934	0.869	0.627 – 1.111	-0.549	-0.988 - -0.110
	$\mu_{B\&E}$	0.337	-0.001 – 0.675	1.266	1.084 – 1.589	0.166	-0.082 – 0.414
	μ_{C}	0.177	0.011 – 0.614	0.317	0.015 – 0.867	0.348	0.017 – 0.941
Armenia	$\mu_{R\&H}$	0.035	-0.029 – 0.099	0.310	0.130 – 0.490	-0.179	-0.341 - -0.017
	μ_{CY}	0.408	0.115 – 0.701	0.683	0.469 – 0.897	-0.333	-0.675 - 0.009
	$\mu_{B\&E}$	0.067	-0.165 – 0.299	0.736	0.507 – 0.995	0.109	-0.099 – 0.317
	μ_{C}	0.055	0.002 – 0.306	0.277	0.016 – 0.774	0.232	0.009 – 0.847
Georgia	$\mu_{R\&H}$	0.131	0.011 – 0.251	0.284	0.073 – 0.495	0.145	0 003 – 0.287
	μ_{CY}	0.330	-0.032 – 0.692	0.853	0.627 – 1.079	0.166	0.024 – 0.308
	$\mu_{B\&E}$	0.231	-0.075 – 0.537	0.881	0.681 – 1.126	0.199	0.033 – 0.365
	μ_{C}	0.235	0.027 – 0.628	0.262	0.010 – 0.843	0.231	0.014 – 0.743
Northern Caucasus	$\mu_{R\&H}$	0.215	0.109 – 0.321	0.379	0.017 – 0.741	-0.218	-0.379 - -0.057
	μ_{CY}	0.278	0.089 – 0.466	0.155	-0.491 – 0.801	-0.252	-0.424 - -0.080
	$\mu_{B\&E}$	0,084	-0.118 – 0.286	1.364	1.178 – 1.629	0.161	-0.033 – 0.355
	μ_{C}	0.263	0.013 – 0.892	0.453	0.033 – 0.932	0.244	0.012 – 0.909
Syria	$\mu_{R\&H}$	0.132	0.010 – 0.253	0.087	-0.224 – 0.398	0.010	-0.163 – 0.183
	μ_{CY}	0.439	0.136 – 0.741	0.674	-0.002 – 1.350	0.075	-0.136 – 0.287
	$\mu_{B\&E}$	0.372	-0.038 – 0.782	1.027	0.731 – 1.354	0.200	-0.022 – 0.422
	μ_{C}	0.335	0.065 – 0.635	0.302	0.012 – 0.879	0.333	0.017 - 0.924
Iraq	$\mu_{R\&H}$	0.025	-0.068 – 0.117	0.107	-0.177 – 0.390	⁂	⁂
	μ_{CY}	0.523	0.054 – 0.993	0.450	-0.145 – 1.044	⁂	⁂
	$\mu_{B\&E}$	0.186	-0.174 – 0.547	1.005	0.720 – 1.290	⁂	⁂
	μ_{C}	0.130	-0.231 – 0.491	0.409	-0.110 – 0.927	⁂	⁂
Lebanon	$\mu_{R\&H}$	⁂	⁂	-0.186	-0.537 – 0.164	⁂	⁂
	μ_{CY}	⁂	⁂	0.077	-0.573 – 0.728	⁂	⁂
	$\mu_{B\&E}$	⁂	⁂	0.170	-0.310 – 0.649	⁂	⁂
	μ_{C}	⁂	⁂	0.354	-0.137 – 0.845	⁂	⁂
Iran	$\mu_{R\&H}$	⁂	⁂	0.253	-0.288 – 0.794	⁂	⁂
	μ_{CY}	⁂	⁂	-0.389	-1.557 – 0.778	⁂	⁂
	$\mu_{B\&E}$	⁂	⁂	0.755	0.202 – 1.309	⁂	⁂
	μ_{C}	⁂	⁂	0.569	0.042 – 1.097	⁂	⁂

⁂ no data was available,

† These intervals represent the values between the 0.025 and 0.975 quartiles for $p1$, respectively.

ɾ $\mu_{R\&H}$: Robert and Hiorns' (1965); μ_{CY}: Chakraborty *et al.*'s (1992), $\mu_{B\&E}$: Bertorelle and Excoffier's (1998), μ : median of Chikhi *et al.*'s (2001) admixture estimator.

Table 3.13 presents the Central Asian admixture estimates (P1) and associated 95% confidence intervals (CI) for each of the hybrid populations. As was the case for the Turkish population, admixture estimates differed from method to method. However, when hybrids from the Near East, Southern, and Northern Caucasus were also considered, it was observed that similarity between the estimates obtained from the methods of Robert and Hiorns' (1965) - Chikhi *et al.*'s (2001), and previously observed similarities (for the estimates of Turkish population) between Chakraborty *et al.*'s (1992) - Bertorelle and Excoffier's (1998) could not be seen.

In the model by Bertorelle and Excoffier (1998), male admixture estimates of Azerbaijan (127 %), Northern Caucasus (136 %) and Syria (103 %) did not fall in the range of 0 % - 100 %. Furthermore, Lebanon (-17 %) in Robert and Hiorns` (1965) and Iran (-39 %) in Chakraborty *et al.*'s (1992) gave negative admixture estimates for the Central Asian contribution. For Alu insertion polymorphisms, the methods of Robert and Hiorns' (1965) and Chakraborty *et al.*'s (1992) gave negative estimates for Northern Caucasus, Armenia and Syria.

Table 3.13 also presents that for mtDNA, Chikhi *et al.*'s (2001) estimates of Turkey and Azerbaijan were close to each other. On the other hand, estimates by the same method, for Y and Alu, as well as mtDNA except Iraq (13%) and Armenia (6%) for females, the Central Asian contribution to the hybrids (Armenia, Georgia, Northern Caucasus, Syria, Iran, Iraq, Lebanon) were similar or even higher than to those of Turkey and Azerbaijan.

3.5.2 Verification of the Assumed Parents

Admixture estimates exceeding 0 % - 100 % range might occur when parental populations were incorrect or when the method applied did not represent the evolutionary past of the populations. The correctness of the choice of parental populations was verified (roughly) in two ways.

Firstly, randomly selected populations were used as the parents in admixture analysis. To save time, simulations were performed on the mtDNA data in accordance with the method of Bertorelle and Excoffier (1998). In these simulation experiments, Turkey was taken as the hybrid. This process was repeated five times.

Table 3.14: Based on the method of Bertorelle and Excoffier (1998) admixture estimates and their standard deviations for the pseudo-parent contribution to Turkey and proportion of the estimator beyond the range for seven other hybrids; mean standard deviations in different simulations.

Simulation		**Turkey**		**% of the estimator beyond the range for seven hybrids**	**Mean SD**
		$\mu_{B\&E}$ for mtDNA	**SD for $\mu_{B\&E}$**		
Parents as in the study		**0.334**	**0.107**	**0 % [0.101 – 0.210]**	**0.151**
Randomly assumed parents	Simulation 1	0.340	0.378	50 % [-1.384 – 1.567]	0.693
	Simulation 2	0.546	0.343	25 % [-1.732 – 0.546]	0.706
	Simulation 3	0.695	0.401	62,5% [-1.526 – 4.989]	4.341
	Simulation 4	1.425	1.430	100 %[-0.879 – 5.817]	3.281
	Simulation 5	0.679	1.602	12,5 % [0.449 – 1.082]	3.891

Simulation 1 *P1:* Kazakh, Altai, Abazian, Cherkessian, Ingushian. *P2:* Uzbek, Tajik, Khoremian Uzbeks, Chechenian, Israel. **Simulation2** *P1:* Hungary, Armenia, Azerbaijan, Chechenian, Kabardinian. *P2* Romania, Kırghiz, Altai, Ingushian, Syria. **Simulation3** *P1:* Romania, Karakalpaks, Georgia, Cherkessian, Syria. *P2:* Albania, Tajik, Armenia, Abazian, Ingushian. **Simulation 4** *P1:* Greece, Bulgaria, Kazakh, Turkmen, Armenia. *P2:* Albania, Hungary, Karakalpaks, Azerbaijan, Chechenian. **Simulation 5** *P1:* Bulgaria, Turkmen, Tajik, Darginian, Cherkessian. *P2:* Albania, Kazakh, Uzbek, Azerbaijan, Ingushian.

Table 3.15: mtDNA admixture estimates of Asian contribution to check the appropriateness of the parental populations.

Excluded Population	$\mu_{R\&H}$	μ_{CY}	$\mu_{B\&E}$
-	**0.190**	**0.380**	**0.334**
Greece	0.236	0.404	0.338
Albania	0.188	0.327	0.295
Bulgaria	0.216	0.355	0.349
Romania	0.182	0.304	0.346
Hungary	0.177	0.308	0.296
Kazakh	0.209	0.327	0.290
Kirghiz	0.206	0.348	0.367
Uzbek	0.183	0.321	0.305
Uighur	0.172	0.317	0.336
Karakalpaks	0.193	0.340	0.317
Turkmen	0.189	0.329	0.312
Tajik	0.182	0.324	0.314
Khoremian Uzbeks	0.182	0.322	0.304
Altai	0.188	0.331	0.323
χ^2	**0.022 (ns)**	**0.024 (ns)**	**0.022 (ns)**

ns statistically not significant

The admixture estimates and the standard deviation for one of the pseudo-parents are given in Table 3.14. As it can be seen from the table, when randomly chosen pseudo-parents were used, or in other words, when parental populations are not identified correctly, the standard deviations become extremely high. Moreover, when the analysis repeated for the other hybrids, it was observed that the number of results exceeding the 0% - 100% range also increased for the mtDNA.

Table 3.16: Y-chromosome admixture estimates of Asian contribution to check the appropriateness of the parental populations.

Excluded Population	$\mu_{R\&H}$	μ_{CY}	$\mu_{B\&E}$
-	**0.136**	**0.158**	**0.642**
Greece	0.217	0.182	0.584
Albania	0.168	0.168	0.647
Bulgaria	0.115	0.153	0.613
Romania	0.164	0.160	0.637
Hungary	0.178	0.166	0.669
Kazakh	0.145	0.187	0.590
Kirghiz	0.175	0.168	0.670
Uzbek	0.043	0.128	0.569
Uighur	0.130	0.160	0.630
Turkmen	0.124	0.151	0.659
Tajik	0.145	0.139	0.672
Altai	0.163	0.158	0.662
χ^2	**0.159 (ns)**	**0.019 (ns)**	**0.024 (ns)**

ns statistically not significant

The possible presence of an outlier in the parental population combinations was also tested by excluding populations one by one from the parental population and applying admixture analysis using the Turkish population as the hybrid. The chi-square analysis indicated that the results obtained after excluding populations from parents were not significantly different from the actual (assumed) ones. Results are given in Tables 3.15 – 3.17. Homogeneity of the results indicated that there were no outliers in the assumed parents and hence no single population dominated the results.

Table 3.17: Alu insertion polymorphism admixture estimates of Asian contribution to check the appropriateness of the parental populations.

Excluded Population	$\mu_{R\&H}$	μ_{CY}	$\mu_{B\&E}$
-	**0.103**	**0.091**	**0.104**
Greece	0.005	-0.011	-0.025
Albania	0.149	0.126	0.110
Romania	0.124	0.124	0.151
Macedonia	0.133	0.124	0.155
Kazakh	0.104	0.090	0.106
Kirghiz	0.105	0.093	0.107
Uzbek	0.100	0.090	0.101
Uighur	0.109	0.090	0.103
Tajik	0.091	0.084	0.095
χ^2	**0.127 (ns)**	**0.151 (ns)**	**0.208 (ns)**

ns statistically not significant

3.5.3. Drift

The method of Chikhi *et al.* (2001) generates estimates of time ($t_{Ai} = T/N_i$). It is a measure of amount of genetic drift since admixture. Figure 3.5 displays the posterior distributions for the t_{Ai} for the two parental populations and their hybrids based on mtDNA HVRI analysis. When parental populations were considered, analysis for the females revealed two almost identical and narrow curves (Figure 3.5 A and B) which in turn indicated that the effects of genetic drift on parental populations (composite populations) were low since the time of admixture event.

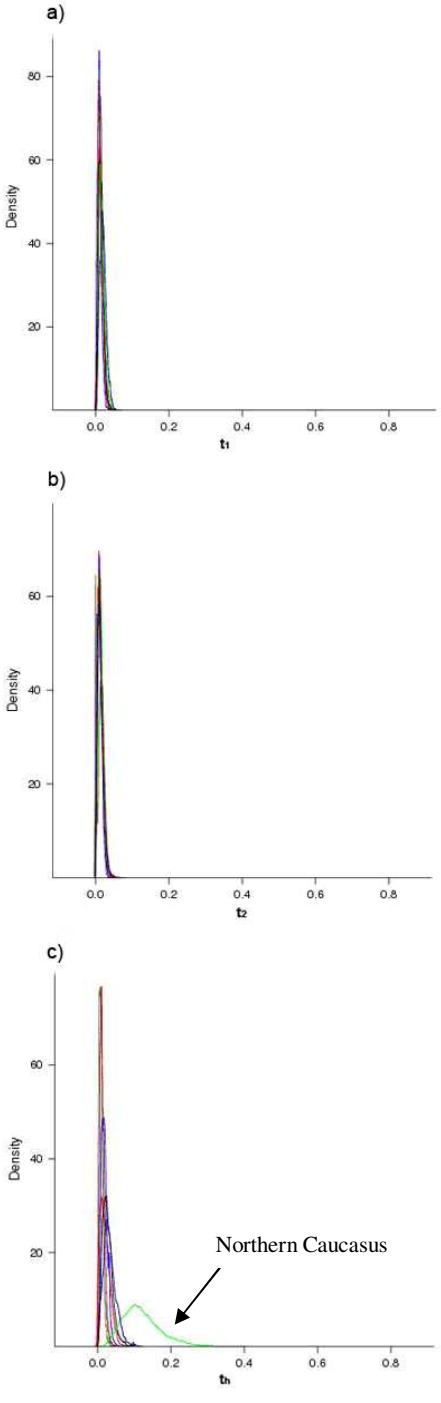


Figure 3.5: Female posterior distributions of T/N_i distribution (A) for Balkans (B) for Central Asia (C) for Hybrids. Turkey: Olive green; Azerbaijan: Red; Armenia: Blue; Georgia: Black; Northern Caucasians: Green; Syria: Dark red; Iraq: Dark Blue

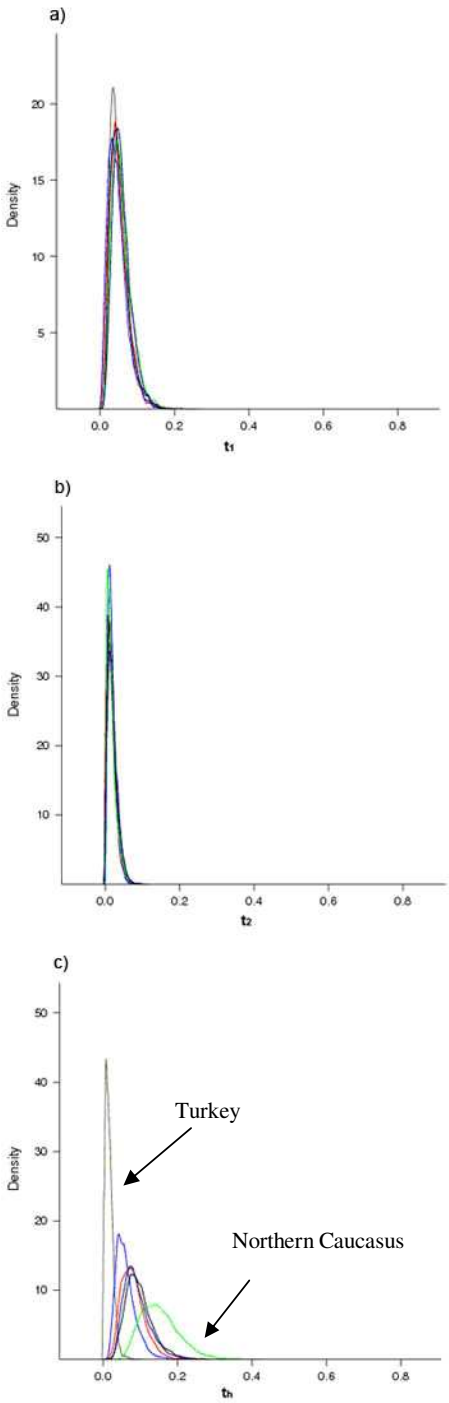


Figure 3.6: Male posterior distributions of T/N_i distribution for (A) for Balkans (B) for Central Asia (C) for Hybrids
Turkey: Olive green; Azerbaijan: Red; Armenia: Blue; Georgia: Black; Northern Caucasians: Green; Syria: Dark red; Iraq: Dark Blue.

As indicated before, mtDNA and Y-chromosome might experience different amounts of genetic drift due to a difference in their effective population sizes. The dissimilar effect of genetic drift on mtDNA and Y-chromosome was also evident in the drift estimations ($t_{Ai} = T/N_i$ distributions) obtained from the method of Chikhi *et al.* (2001) (Figures 3.5 and 3.6) For both of the sexes the distributions of parental populations were almost identical, but for males the distributions were slightly wider especially for the Balkans.

In contrast to the parental populations, when the distributions for the hybrids were considered, for all of the hybrids the distributions were relatively wide, even wider for the Y-chromosome (Figure 3.6c). These results indicated that random genetic drift was experienced by many of the employed populations (such as male Balkan parents and hybrids), but especially by male hybrids.

The narrowest curve was obtained for the Turkish population (Figure 3.5c and 3.6c), which in turn indicated that Turkey had a large population size, and was not greatly affected by drift.

Among the four admixture methods employed, the genetic drift experienced by the populations was taken into account only by Chikhi *et al.*'s (2001) method. Hence, under the light of Figures 3.5 and 3.6, the most plausible method to be used in our study seems to be the Chikhi *et al.*'s (2001).

3.5.4. Expected Compatibility of Different Estimations for a Population

If males and females do not contribute in equal amounts to the hybrids and if genetic drift was not significant, than the estimate obtained from Alu insertion polymorphisms is expected to be found between the range of mtDNA and Y-chromosome estimates.

In all of the methods it was observed that the estimates based on mtDNA and Y-chromosome were different. Only by the method of Chikhi *et al.* (2001) all Alu insertion polymorphism estimates were very close or in between the estimates obtained from mtDNA and Y-chromosome (Table 3.13). Therefore, in the present study the highest confidence among the estimates of the different methods is given to those obtained by Chikhi *et al.*'s (2001) method.

3.5.5. Central Asian Contribution to Hybrids with a Special Emphasis to Turkey

By using the Chikhi *et al.*'s (2001), it was observed that there were differences between the male and female contributions from Central Asia to hybrids.

Female contributions from Central Asia to Turkey, to Northern Caucasus and to populations from the Southern Caucasus were similar and relatively moderate but the highest contribution was to Syria. On the other hand, for the males, the Central Asian contribution was lowest for the Turkish population, whereas the highest contribution was to the Northern Caucasus and Iraq. Alu insertion polymorphisms also indicated that the Central Asian contribution to Turkey was lower than the contribution to the other neighboring populations.

3.6. Regression Analysis

To find out what background degree of Central Asian contribution was present in non-Turkish speaking neighbors of Turkey and Azerbaijan, furthermore, to see if this background was changing as a function of distance from Central Asia, regression analysis was carried out. In the analysis, the admixture estimates obtained by using Chikhi *et al.*'s (2001) method were employed.

For the regression lines, as well as the neighbors of Anatolia and Azerbaijan (Georgia, Armenia, Syria, Iran and Iraq), Lebanon, and the Northern Caucasus were used within the limits of available data. In the regression analysis Anatolia and Azerbaijan were not employed.

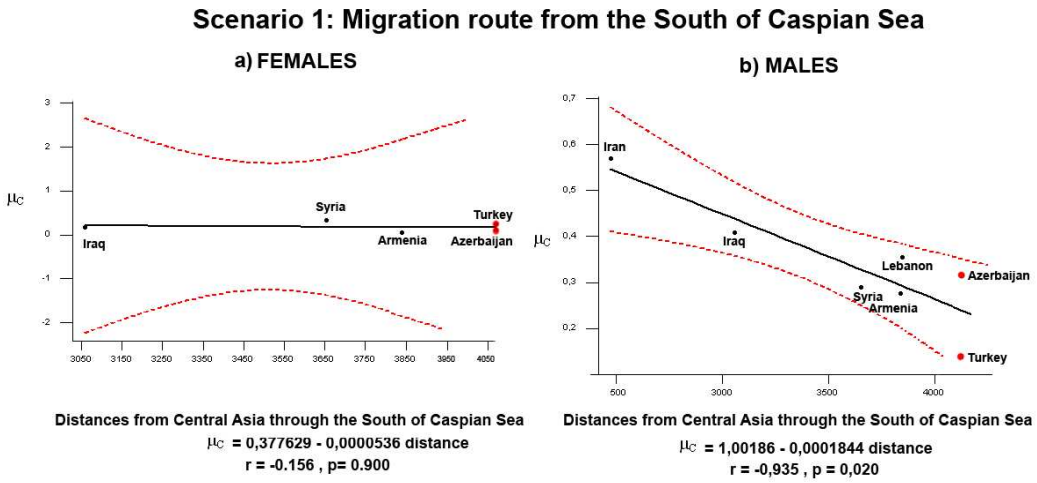


Figure 3.7: Linear regression analysis showing the relationship between the Central Asian contributions based on Chikhi *et al.*'s (2001) method to the hybrids as a function of the geographic distances in accordance with the Scenario 1. Regression Equation, Pearson correlation coefficient (r), associated significance (p). a) Females. b) Males.

Considering two possible routes (two scenarios) for migrations from Central Asia, two sets of regression lines for mtDNA and Y-Chromosome markers were drawn and displayed together with their statistics in Figures 3.7 – 3.8.

Table 3.18: For scenario 1, the expected admixture estimates for Turkey and Azerbaijan from the obtained regression equation and observed estimates based on Chikhi *et al.*'s (2001) method.

	Turkey	Azerbaijan	Turkey	Azerbaijan
Expected*	0.17		0.24	
Observed	0.22	0.18	0.13	0.32
Difference	+5%	+1%	-11%	+8%

* expected based on the regression line

For these lines, origin was the Barry center of Central Asia and the end of the routes were midpoints between Turkey and Azerbaijan, the countries up to or very near to the Turkic speaking region were included in the analysis and only the non-Turkic speaking hybrids were considered.

When scenario 1 was considered, results of the analysis of Figure 3.7.a and Table 3.18 indicated that female contributions from Central Asia was not changing significantly as a function of distance from Central Asia on the suggested route, although there seemed to be higher contributions for Turkey (5%) and Azerbaijan (1%). However, for the Y-chromosome, a significant linear relationship (Figure 3.7.b) was found. For the Azerbaijan males, there was 8% greater Central Asian contribution than that of the expected values (Table 3.18). Yet, for Turkey, males were 11% short of expected values (Table 3.18).

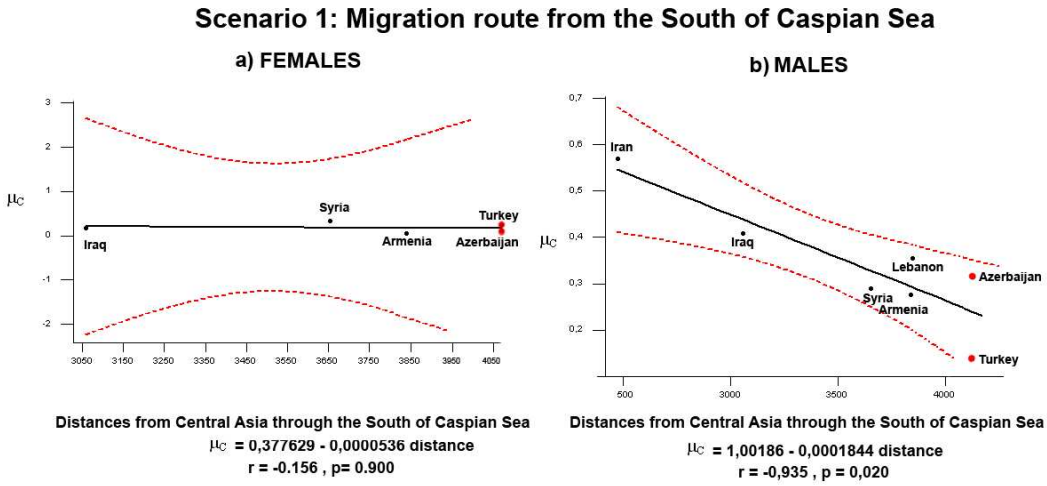


Figure 3.8: Linear regression analysis showing the relationship between the Central Asian contributions to the hybrids as a function of the geographic distances in accordance with the Scenario 2. Regression Equation, Pearson correlation coefficient (r), associated significance (p). The expected admixture estimates for Turkey and Azerbaijan from the obtained equation and observed estimates based on Chikhi *et al.*'s (2001) method. a) Females. b) Males.

Table 3.19: For scenario 2, the expected admixture estimates for Turkey and Azerbaijan from the obtained regression equation and observed estimates based on Chikhi *et al.*'s (2001) method.

	Turkey	Azerbaijan	Turkey	Azerbaijan
Expected*	-0.07		0.13	
Observed	0.22	0.18	0,13	0,32
Difference	+29%	+25%	0%	+19%

* expected based on the regression line

For the Scenario 2, none of the regression lines were significant. Yet, there were negative slopes for both of the sexes. For the females both in Azerbaijan and in Turkey, observed values were more than those observed around the region. Again, in Azerbaijan higher values than expected can be detected for males. However, for Turkey, males did not show presence of any excess contribution compared to other non-Turkic speaking neighbors.

3.7. Comparison of Admixture Estimates of the Region of Language Replacement with its Closest Neighbors

Figure 3.9 shows the Central Asian contribution to the region of language replacement (Azerbaijan and Turkey).

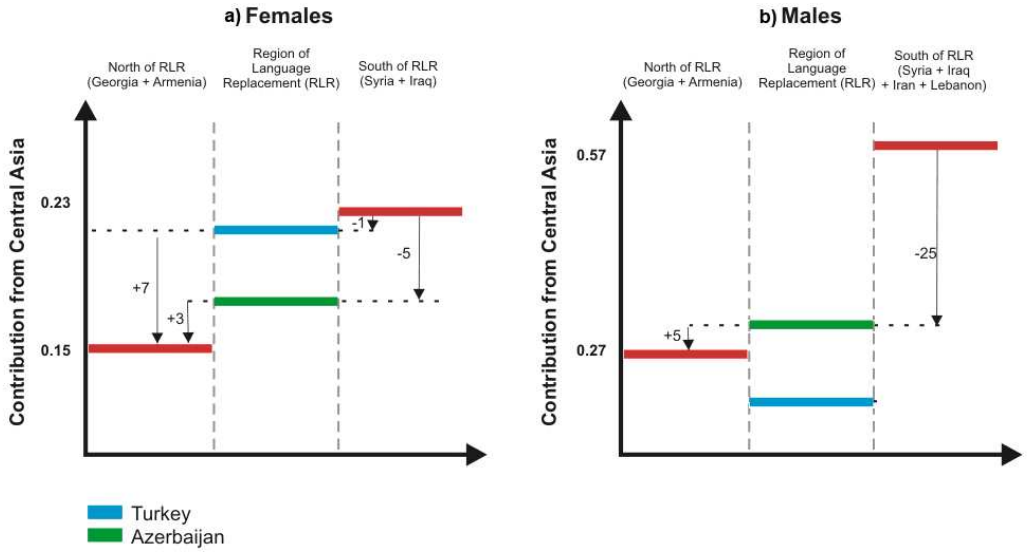


Figure 3.9: Comparison of the contributions from Central Asia to the region of language replacement together with its northern and southern neighbors. a) Females, b) Males.

It also shows the average central Asian contribution to the Northern and Southern parts of the region of language replacement comparatively.

Figure 3.9a shows that when the admixture estimates of Turkey and Azerbaijan (language replacement region) were compared with those of closest neighbors from the north (Armenia and Georgia), for females excess in both Turkey (~7%) and Azerbaijan (~3%) were observed. On the other hand, for males, only in Azerbaijan was an excess (~5%) observed (Figure 3.9b). When southern neighbors (Iraq, Iran, and Syria in males; Iraq and Syria in females) were considered, higher contribution from Central Asia was observed in this region than that of the region of language replacement.

CHAPTER IV

DISCUSSION

"Elite dominance" is the term proposed by Renfrew (1987) to describe the process when the language of few individuals (elites) is adopted by the rest of a population with little or no genetic contribution from the elites to the population (Renfrew, 1991). Indo-European language, spoken in Anatolia and Azerbaijan, was replaced with a Turkic language, starting with the arrival of Turks into the region. Conventionally, the date of their arrival was 1071. They were formally referred to as Seljuks, who belonged to the Oghuz living along the north bank of Syr Derya. However, there had been individual arrivals from the Central Asia since the days of Abbasid Caliphate and arrivals by early incursions under names such as Danishmends in the early eleventh Century (Cohen, 1968). Furthermore, migrations did not cease after the arrival of the Seljuks (Lewis, 1995). Perhaps, these continuous migrations were from linguistically related areas and were facilitated after the change of the language in the area (Benedetto *et al*., 2001).

The Oghuz Turks are one of the major branches of Turks. Their original homeland was the Ural-Altay region of Central Asia and they were also present in areas west of the Caspian Sea. They may have been living together with other Turks even centuries before their mass-migrations started from Central Asia in the 9th Century CE (Roux, 1997). Arrivals of Turkic people to Anatolia and Azerbaijan from Central Asia and the Northern Caucasus lasted for more than two centuries (Vronis, 1971; Esin, 1983; Lewis, 1995).

Studies based on protein (Brega *et al*., 1998), mtDNA (Calafell *et al*.,1996 ; Comas *et al*., 1996) and Y-chromosome (Wells *et al*., 2001) data indicate that Central Asian populations had little genetic effect on the current day Turkish gene pool, thus these results were considered as support to the idea that the Turkic language was imposed by the few elites. However, in none of these studies was Central Asian contribution calculated.

In the present study to quantify the Central Asian contribution to Turkey four different admixture methods were employed; Robert and Hiorns' (1965), Chakraborty *et al.*'s (1992), Bertorelle and Excoffier's (1998) and Chikhi *et al.*'s (2001). In all of these methods, a model with two parents was assumed. The two-parent assumption accommodates all four of the methods. In the model, as in Benedetto *et al.*'s (2001) study, it was assumed that before the invasion of Turkic speaking nomadic groups, genetic make up of the population in Anatolia was similar to that of the Balkan (in the present study, composed of the samples from Greece, Bulgaria, Albania, Hungary and Romania) populations. Since, the very same language replacement was experienced by Azerbaijan, Central Asian contribution to Azerbaijan was also determined with respect to Balkans. For these estimations, we were aware of the fact that the Balkans may not represent the Azerbaijan populations, before Turkic migrations. This issue was tackled by the regression analysis and by the comparative studies in the region to be discussed below.

It was assumed that the source population, which replaced the existing Indo-European language in Anatolia and Azerbaijan, as well as the ones which later arrived to Anatolia by the attraction of Turkic language, were originally Central Asian populations. In turn, Central Asian parental population was composed of the samples from Kazakhstan, Kyrgyzstan, Uighur, Altai, Uzbekistan, Turkmenistan, Tajikistan, Khoremian Uzbeks, and Karakalpaks. These countries harbor the lands from where the long-range spread of the Seljuks started from and / or they are Turkic-speaking countries.

The only exception is the Tajik, who was an Indo-Iranian speaking population. However, previous analysis based on this population showed that the genetic structure of Tajik was not different from the other Turkic speaking Central Asian populations (Comas *et al*., 2004a). The results of the present study, in searching for the presence of outliers, revealed the same conclusion (Tables 3.15 -3.17).

Current admixture methods reveal the cumulative contribution of admixtures from the parents (Jobling *et al*., 2004). Therefore, the obtained admixture estimates for the Turkish and Azerbaijani populations might not solely represent the contributions of the Seljuks and the migrations arriving after the Seljuks. It was hypothesized that if there was a specific contribution related with the language replacement started by the Seljuks then there should be a higher contribution from Central Asia to Turkey and Azerbaijan (region of language replacement, RLR) than that of Indo-European speaking Armenia, Kartvelian speaking Georgia, Afro-Asiatic (Arabic) speaking Syria and Iraq, or Caucasian speaking Northern Caucasus. Hence, admixture estimates, reflecting the Central Asian contributions, if there were any, must be higher in Turkey and Azerbaijan (RLR) and jointly they should appear as an island in a territory of non-Turkic speaking populations. For the non-Turkic speaking countries, contributions of Central Asia was calculated with a reference to the Balkans. Yet, these contributions were judged with respect to the proximity of these countries to RLR.

Analyzing the Central Asian contribution to the different parts of the genome, gives the opportunity of following migration patterns of males and females independently. While, admixture estimates based on mtDNA and Y-chromosome would show the female and male contributions respectively, estimates based on autosomal markers provide joint information about the migrations of males and females and thus enable us to verify scenarios of the sex-related migration patterns.

As indicated before, the only previous study that used admixture analysis to find the Central Asian contribution to Anatolia was that of Benedetto *et al.*'s (2001). They used samples from Bulgaria, Italy, Crete, Greece, and Sicily as the Balkan parent, and samples from Kazakh, Kirghiz, and Uighur populations as the Central Asian parent. Based on the Bertorelle and Excoffier's (1998) admixture method, they determined 30% contribution for mtDNA, 47% for 5 Y-chromosome microsatellites and 35% for 1 autosomal microsatellite. Since the Bertorelle and Excoffier's (1998) method does not incorporate the effect of genetic drift, and since genetic drift has higher effect on rare alleles, by excluding the rare haplogroups / alleles from the data set, they tried to eliminate the effect of genetic drift on their results and found that for all markers the Central Asian contribution to Turkey was about 30%.

However, while trying to eliminate the effect of genetic drift, they also excluded some information present in the allelic distributions. For example, they analyzed five Y-chromosome loci (21 alleles) and determined a 47% Central Asian contribution. With the exclusion of the rare alleles, only six alleles remained in the data and the estimate based on Y decreased to 26%. It is well known that, using the frequencies of only some haplogroups / alleles may result in insufficient use of data and might introduce a bias (Chikhi *et al.*, 2002).

In the present study, for Turkey, despite the considerable differences between the compositions of the parents, molecular markers, and number of analyzed individuals, similar results were obtained to those of Benedetto *et al*. (2001) based on the mtDNA sequences. As was the case in Benedetto *et al*.'s (2001), in the present study, a higher Y-chromosome estimate (64%) than that of mtDNA (33%) and autosomal microsatellites (31%) were obtained. Furthermore, in the present study a ~10% Central Asian contribution based on 7 Alu insertion polymorphisms was found.

When the admixture estimates obtained by all of the applied the methods were considered, it was observed that admixture estimates for a molecular marker were quite different from each other depending on the method used.

Admixture methods are statistical methods that are making several assumptions and simplifications about the evolutionary forces that have been molding the gene pools of the hybrids. Thus, obtaining different admixture proportions for the same molecular marker was not surprising. It must also be emphasized that most of the admixture proportions were associated with wide confidence intervals. Hence, the possible range of one particular estimate may cover another estimate of the same marker or estimate of another marker, making them indifferent from the statistical point of view. Yet, these are the estimates that could be reached using current methods. In the present study, it tried to focus on results of one of the methods, which seemed to provide the most reliable estimations for the studied populations.

Genetic drift is strong in the isolated populations that have small sample sizes. Central Asian populations (Zarjal *et al.*, 2002) and Caucasian populations (Nasidze *et al.*, 2001; 2004) are known to be small and isolated. Consequently, they are prone to genetic drift. Furthermore, due to their inheritance pattern, the effective population sizes of mtDNA and Y-chromosome are only the 1/4 of the autosomal ones (Jobling *et al*., 2004). The higher reproductive variability of males further decreases the effective population size of Y-chromosome to 1/8 of the autosomes (Charlesworth, 2001; Caballero, 1995). Thus, it can be concluded that mtDNA, but especially Y-chromosome, is more prone to genetic drift then the autosomal chromosomes. A high effect of genetic drift on populations under consideration was also shown in the results of previous studies. For example Y-chromosome based drift in Central Asia (Zerjal *et al.,* 2002), in Greece (DiGiacoma *et al*., 2003) in the Caucasus (Nasidze *et al.*, 2003), again in the Caucasus, drift of Alu (Nasidze *et al*., 2001) and mtDNA (Nasidze and Stoneking, 2001) were all reported.

Admixture estimates may exceed 0% - 100% range. Two of the possible reasons of observing an estimate beyond this range could be (i) the presence of strong drift or (ii) the presence of misidentified parental populations (Bertorelle and Excoffier, 1998).

In all methods, except that of Chikhi *et al.*'s (2001), which is the method taking genetic drift into account, admixture estimates exceeding this range were present. It is expected that when parental populations are not identified correctly, by the method of Bertorelle and Excoffier (1998), the probability of obtaining admixture estimates exceeding the 0% – 100% range apply for all of the markers. In parallel to this argument, in our mtDNA based simulations done by randomly composed parents, estimates beyond the 0% - 100% range were observed (Table 3.14). However, results exceeding 100 % contribution of a parent were obtained just for the Y-chromosome data of Bertorelle and Excoffier's (1998) method, which does not consider the effect of drift as a force affecting the gene pools. Furthermore, observed associated standard deviations were not higher than those obtained when the parents were composed of randomly chosen populations.

Results exceeding the normal range were also seen in the Robert and Hiorns' (1965) and Chakraborty *et al.*'s (1992) methods. The negative estimates obtained by the Alu insertion polymorphisms with these two methods may have resulted because of their assumption that the allele frequencies of the hybrid population should be linear combinations of those of the parental populations. In other words, in these two methods the change in haplogroup / allele frequencies from the time of admixture until today was not considered. Consequently, this simplification might have decreased the sensitivity of the method and resulted in negative estimates.

Furthermore, estimates of mtDNA and Y-chromosome might show differences, but estimations of the autosomal markers must be in between them. Otherwise, this can be considered as another evidence for the presence and influence of genetic drift on the gene pool of the analyzed populations. Hence, the case requires the employment of the methods that consider the drift.

However, because the language replacement in Anatolia and Azerbaijan took place a relatively short time ago (around 1000 ya), one may hesitate to consider the effect of drift in this particular case. Yet, in the present study, cumulative contributions of Central Asia were calculated and time span might not necessarily be limited by 1000 years. Contribution could have started as early as 30.000 ya by the migration of Central Asians into Europe and the Near East (Underhill *et al*., 2001)

The method of Chikhi *et al*. (2001) considers the effect of genetic drift in parents as well as hybrids. The other method, among the ones used in the study, which takes into account of drift, is Chakraborty *et al*.'s (1992) method. Yet, in the Chakroborty *et al*.'s (1992) method, drift is not taken into account for the parents. Therefore, in the present study, estimations obtained by the Chikhi *et al*.'s (2001) method are assumed to represent the closest estimates to the true Central Asian contributions.

Indeed, none of the estimates are out of the 0% - 100% range with this model. Furthermore, estimates based on autosomal markers are in between the ones for mtDNA and Y-chromosomes (Turkey, Armenia, and Syria) or they are close to one of them, but not exceeding drastic (Azerbaijan, Georgia, and Northern Caucasus).

The method also measures the amount of drift in the form of time, scaled by the effective population size ($t_{Ai} = T/N_i$) (Chikhi *et al*., 2001). The distribution curves that are narrow and almost identical suggest that the parental populations used in the admixture analysis experienced a limited drift since admixture and had a rather large, long-term population size (Chikhi *et al*., 2001; 2002). In the present study, the distributions of the drift estimations ($t_{Ai} = T/N_i$ distributions) of Chikhi *et al*.'s (2001) indicated that the pooled, composed parental populations (with n=562-453) experienced limited drift between the time of admixture and sampling. Hence, pooling must have hidden the effect of the drift experienced by the components of the parents. The Northern Caucasus was the hybrid on which genetic drift had the most significant effect for both males and females (Figure 4c and 5c) as was observed before (Nasidze *et al*., 2001; Nasidze and Stoneking, 2001; Nasidze *et al*., 2003).

In the present study, Northern Caucasian populations were grouped together (n = 213 (mtDNA), n = 144 (Y-chromosome), n = 205 (Alu insertion polymorphisms)) to avoid the effect of small population sizes of the individual populations. Still, highest genetic drift was seen in this group.

Thus, especially for the methods that do not consider the effect of genetic drift, results for Northern Caucasus should be considered with doubt. Therefore, Chikhi *et al.*'s (2001) method once more is expected to give the most reliable estimates among those obtained through the methods used in the study. Hence, the estimates for the Central Asian contribution in Anatolia are as follows:

	Estimate	*95% Confidence interval*
mtDNA:	0.22	0.03 – 0.56
Y-chromosome:	0.13	0.01 – 0.52
Alu insertion polymorphism:	0.15	0.01 – 0.62
Autosomal microsatellite:	0.24	0.06 – 0.42

The estimates for the Central Asian contribution in Azerbaijan are as follows:

	Estimate	*95% Confidence interval*
mtDNA:	0.18	0.01 – 0.61
Y-chromosome:	0.32	0.02 – 0.87
Alu insertion polymorphism:	0.35	0.02 – 0.94

Estimates for mtDNA are different from those of the Y-chromosome. For all of the markers, and for both of the countries, estimates are associated with large confidence limits. These values may not reflect the actual differences between the contributions in males and females and perhaps one may want to pool the results.

However, evolutionary histories of the two sexes might be different as was hinted by the diversity measures. The Balkans and Caucasus are highly variable in males more than that of Central Asia, and the Near East, whereas the ranking is the opposite for females and autosomal markers. These results suggested that male populations of the Balkans and Caucasus have a higher degree of admixture compared to their female counterparts.

The warriors of the Avars, Huns, Pecheneg, Kipchaks during the incursions to the Balkans (Roux, 1997), might be responsible for the admixtures in the Balkans. Since most of them belong to major branches of Turks (other than the Oghuz) or are composed of people from Eastern Central Asia, their admixture might have generated the similarity between Central Asia and the Balkans as is displayed in Figure 3.2. If these migrations known in history had higher male mediated impacts than that of the females on the Balkans, then the Balkans should not be considered as the identical reference for the males and females and the results cannot be pooled. Therefore, the estimates of the two sexes were analyzed separately.

In terms of contributions in mtDNA, Anatolia was in the 4^{th} and Azerbaijan the 5^{th} rank in descending order. Based on the Y chromosome contributions, Azerbaijan was the 5^{th} and Turkey was the last (9^{th}). Only in Alu insertion polymorphisms, Azerbaijan was the 1^{st} but Turkey was the last out of 6 populations.

Males and females may have different migration histories (Jobling *et al*., 2004). Some migrating populations might be only composed of males. However, females do not migrate alone; instead, they migrate together with males. In other words, the migrations that took place in the past may have been composed of just males or of males and females together. However, lower male contribution than that of females was observed in Turkey and Syria, but especially in Turkey, and this situation could not be explained by higher female migration to these regions. There might be other explanations for the lower male contribution from Central Asia to Turkey.

Contribution of males and females on the genetic structure of a hybrid population can vary if they contribute unequally to the hybrid (Jobling *et al*., 2004). For example, in some cases one sex from incomers might have more chance to incorporate their genetic make up into that of the invaded population, resulting in sex-biased admixture. Based on this scenario, it can be assumed that the number of males and females who entered Turkey are approximately equal (Figure 4.1a).

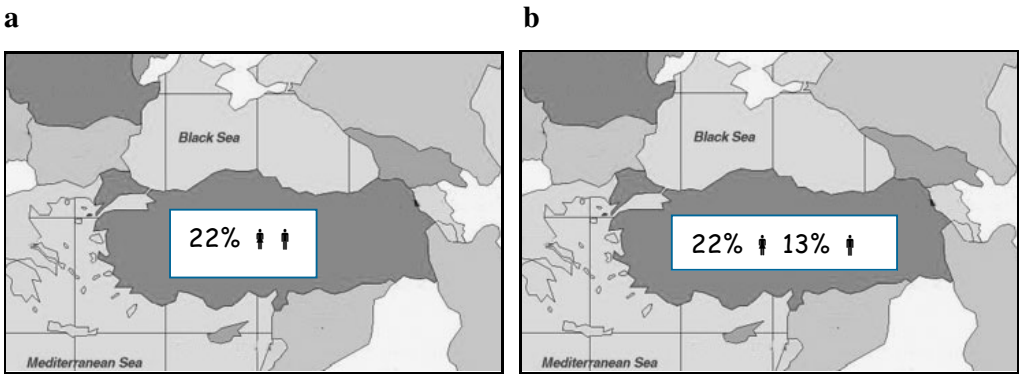


Figure 4.1: Schematic representation explaining the possible mechanism of especially low male contribution in Turkey due to sex-biased admixture in Anatolia. a) Before sex-biased admixture. b) After sex-biased admixture.

Based on the sex-biased admixture, female immigrants could contribute to the recipient gene pool more than that of the males and hence, a lower male admixture proportion was obtained (Figure 4.1.b) when the current day Turkish population was used to estimate the past population processes. However, this scenario might not be so plausible since if it was true, such a pattern should also be seen in the neighboring populations such as in Azerbaijan and in Georgia.

On the other hand, if there were a high degree of homogenization between the males of the Balkans and of Anatolia (Figure 4.2b) this would cause a dilution in the Central Asian contribution for the males of Anatolia. Male mediated gene flow from the Balkans to Anatolia was also hinted at the E, P and Y haplogroups (Table 3.4) that had high frequencies in Balkans and lower frequencies in Turkey but much lower frequencies in Central Asia. A relatively high degree of similarity in males of Turkey to those of the Balkans is also seen in Figure 3.2. Perhaps, males and females entered Turkey in similar numbers, even higher in favor of males from Central Asia, but lower Central Asian admixture for males was obtained just because of the increased similarity between the male genetic compositions of the Balkans and Turkey. It can be suggested that the Janissaries, an army composed of youth, originally Christian boys of Balkans, who were employed for 4 hundred centuries during the Ottoman times (Goodwin, 1994) might be at least partly responsible for this homogenization.

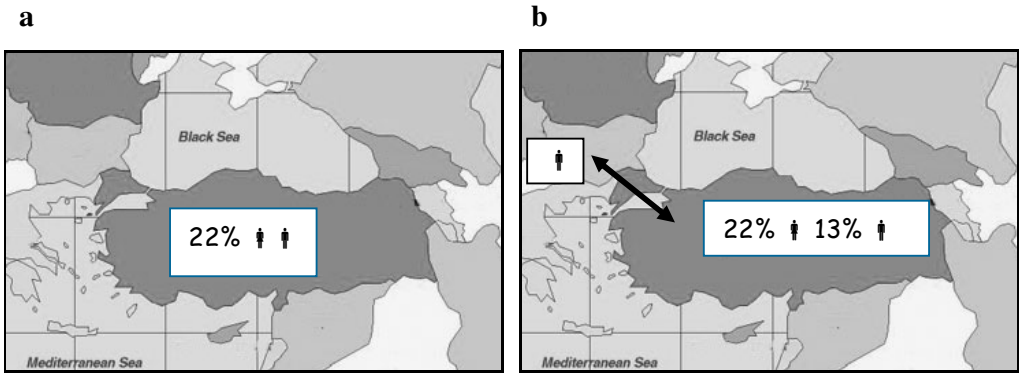


Figure 4.2: Schematic representation explaining the possible mechanism of especially low male contribution in Turkey due to the homogenization of males between Balkans and Anatolia. a) Before homogenization. b) After homogenization.

At this stage of our discussion, it must be emphasized that admixture estimates from different molecular markers indicate the absence of a special contribution of Central Asia, which can be associated with the language replacement. Anticipating that there might be a pattern in Central Asian contribution, changing as a function of distance from Central Asia and generating a background ("language replacement" independent) contribution in the regions, regression analyses were carried out. Only the non-Turkic speaking hybrids were considered and migrations from the south of Caspian Sea up to Azerbaijan and Anatolia, including Armenia (1st scenario), migrations, from the north of the Caspian Sea up to Anatolia and Azerbaijan, (2nd scenario) were assumed.

The reason to invoke the two scenarios was to determine the language independent expectations of the Central Asian contributions to Anatolia and Azerbaijan. The estimated distances from Central Asia in the present study were very crude. The newcomers might not have followed exactly the same route that was assumed in the present study. Moreover, since the distances were calculated from the central points of the countries, in some cases, essentially for the large countries, a more distant neighbor may enter between the two neighbors. For example, Azerbaijan and Turkey are close neighbors but when the central points were used they seemed to be quite distant from each other. Therefore, the midpoint between the geographical distances of Azerbaijan and Turkey was used to represent the region where the language was replaced.

For the Y-chromosome based regression line from the south (1st scenario), a significant relationship between the Central Asian contributions and distances was observed. It is a well-known fact that in 70% of human societies (Burton *et al.,* 1996), females have higher short-distance mobility then males due to patrilocality (see e.g., Ooata, 2001). Therefore, the pattern in females might become blurred when females leave their birthplace to move to a new place when they get married. Based on the significant Y regression line in the 1st scenario, it can be said that an 8% excess in Central Asian contribution for the male population of Azerbaijan was detected, which could be associated with the language replacement contribution.

Furthermore, when the admixtures of "language replacement region" were compared with those of the closest neighbors from the north, excess in both males (~5%) and females (~3%) in Azerbaijan and an excess in females (~7%) of Turkey were observed. However, there were higher admixtures (1-5 % more in females and at least 25 % more in males) in their southern neighbors (Iraq, Iran, Syria, and Lebanon in males; Iraq and Syria in females). Reconsidering the fact that estimates by Chikhi *et al.*'s (2001) method, for Y and Alu, as well as mtDNA except Iraq (13%) and Armenia (6%) for females, the Central Asian contribution to the hybrids (Armenia, Georgia, Northern Caucasus, Syria, Iran, Iraq, Lebanon) were similar or even higher than to those of Turkey and Azerbaijan (Table 3.13). These results suggested that Central Asian people, representing originally Turkic speaking people, are all over the southern (Iran, Iraq, Syria, Lebanon) and northern (Northern Cacasus, Armenia, Georgia) routes as well as in the RLR.

There is no specially high admixture proportion in RLR by Turkic speaking people. Does this mean that language replacement was emposed by the elites? Alternatively, perhaps Central Asian emmigrants arriving to the lands between Central Asia and the Balkans, as a funtion of distance from Central Asia, through centuries were occasionally absorbed by the host populations, sometimes kept their identity as Turkic speaking minorities (as is seen in Iran and Iraq today) and only in RLR did they manage to empose their language. Ethnic and religeous pluralism, together with political and millitary weakness of Byzantinum at the begining of 11th Century CE (Vronis, 1971) in Anatolia might have facilitated the invasion of Seljuks, hence the adoption of Turkic language in contrast to the southern neighbors. If this alternative scenario is true, language may not have been replaced by the elites, as defined by Renfrew (1987). Actually, how small must the group to be considered as "elite"? Certainly the upper limit should not be more than 10% of the total population size. Chikhi *et al.*'s (2001) method revealed that the net effect of the female migrations is equivalent to the effect of single admixture event with a 22% Central Asian contribution. Because of the arguments given above, for the males Central Asian contribution must be at least at the same magnitude.

Although, the details of the arrivals (times, partial contributions) are not known, based on the results of the present study, it was observed that there were no facilitated migrations to the RLR. Therefore, one could argue that before the language replacement episode there was a community from Central Asia in the region, and with the arrivals of Seljuks, language started to be replaced. Unless, the community was less than half of the estimates (<10%), episode can not be assumed as the action of few elites as was envisaged by the model of Renfrew (1987).

As indicated before, studies on protein (Brega *et al.*, 1998), mtDNA (Calafell *et al.*, 1996; Comas *et al.*, 1998) and Y-chromosome (Wells *et al.*, 2001) data indicated that genetically Anatolia was more closely related with Balkans than Central Asia that was used as a support for the elite-dominance model of Renfrew (1987). In the presented study, Central Asian contribution, seemingly 10% in males and 22% in females or in other words more than 80% similarity to Balkans was in accordance with these studies. However, present study pointed out that the language might not be replaced with a small group of elite.

Finally, it could be concluded that there is at least a 22%, contribution, with large confidence intervals, from Central Asia to Turkey with respect to the Balkans. Male and female contributions seem to be different, indicating the possible difference in their evolutionary histories. Moreover, results pointed out that language in Anatolia might not have been replaced by the elites, but by a large group of people. Therefore, it can be concluded that the observations do not support the elite dominance model of Renfrew (1987 ; 1991).

CONCLUSION

In the present study, the Central Asian contribution to the gene pool of Anatolia and Azerbaijan, "the region of language replacement" (RLR), and their non-Turkic speaking neighbors (Northern Caucasus, Armenia, Georgia, Syria, Iraq, Lebanon, and Iran), with reference to the Balkans was determined. Furthermore, by comparing the estimates in RLR and non-Turkic speaking neighbors, the association between the language replacement episode and Central Asian contribution tried to be identified.

Results of the present study indicated that:

1. To be able to follow different evolutionary histories of males and females three sets of markers: mtDNA sequences, Y-chromosome haplogroups and autosomal markers (Alu polymorphisms and microsatellites) were retrieved from databases.

2. For the analyzed populations, haplogroups based on mtDNA HVRI sequences were determined. Furthermore, the mtDNA and Y-chromosome haplogroup frequencies together with the allele frequencies for the Alu insertion polymorphisms and autosomal microsatellites were identified to be used in further analysis.

3. For the mtDNA data, the haplotype and nucleotide diversities of the Turkish population (0.986 ± 0.004 and 0.018 ± 0.010 respectively) were close to those values obtained for the Southern Caucasus populations (0.987 ± 0.003; 0.019 ± 0.010 respectively).

4. The diversity estimates of Alu insertion polymorphisms of Turkey (0.352 ± 0.071) and the Southern Caucasus (0.359 ± 0.063) were very close to each other and slightly lower than that of Central Asian. The lowest values were obtained for the Northern Caucasus as 0.310 ± 0.057.

5. The Y-chromosome diversity estimates, on the other hand, for Balkans (0.8307 ± 0.0058) and Turkey (0.8308 ± 0.0085) were slightly lower than those of the populations of Caucasus and very similar to each other.

6. Determined haplogroup and allele frequencies were used to visualize the genetic relatedness of the examined populations in two dimensions by principle component analysis. The same analysis enabled us to determine the haplogroups / alleles that were differentiating the populations. Based on the principle component analysis, it was observed that genetically, Balkan males highly resembled those of Central Asian populations. Finally, principle component analysis indicated that Anatolian females are genetically closer to those of Caucasians and Southern Caucasians but the males are genetically close to those of the Balkans.

7. Central Asian (Zarjal *et al*., 2002) and Caucasian populations (Nasidze *et* al., 2001; 2004) are known to be small and isolated, hence prone to genetic drift. In the present study, the Northern Caucasus was determined as the population in which genetic drift had the most significant effect for both males and females.

8. In the study, to determine the Central Asian contribution to Anatolia and Azerbaijan (RLR), Robert and Hiorns' (1965), Chakraborty *et al.*'s (1992), Bertorelle and Excoffier's (1998) and Chikhi *et al.*'s (2001) methods were used. In all of these methods, a model with two parents, Central Asia, land of Turkic speaking nomadic groups, and the Balkans, and the representatives of

the gene pools of the hybrids (Anatolia, Azerbaijan, Armenia, Georgia, Northern Caucasus, Syria, Iraq, Lebanon and Iran) before the invasion of Turkic speaking nomadic groups were used. Genetic drift was identified as the major factor operating on the analyzed populations. Therefore, estimations based on the Chikhi *et al.*'s (2001) method, which considers the effect of genetic drift, were assumed to represent the closest estimates to the true Central Asian contributions. Indeed, when the Chikhi *et al.*'s (2001) method was used, estimates did not exceed the 0% - 100% range. Furthermore, estimates based on autosomal markers are in between the ones for mtDNA and Y-chromosomes or they are close to one of them but not exceeding drastic.

9. Based on the Chikhi *et al.*'s (2001) method, it was observed that for the females, the admixture estimates of Turkey (22%) and Azerbaijan (18%) were relatively similar, whereas male contribution from Central Asia is lower in Turkey (13%) than that of females but it is greater than that of the females in Azerbaijan (32%).

10. Diversity measures hinted that evolutionary histories of males and females might be different for the Balkans. Therefore, the Balkans should not be considered as the identical reference for the males and females.

11. Lower male than female contribution from Central Asia to Anatolia was obtained. The situation was explained by invoking the idea of "homogenization between the males of the Balkans and Anatolia". Since females could not migrate alone, the true Central Asian contribution for both males and females were assumed to be 22%.

12. The Central Asian contribution to RLR was determined with regression analysis by assuming two hypothetical migration routes along the northern and southern banks of the Caspian Sea. Results indicated that there were three

negative (females through southern route, males and females through northern route) relationships between the Central Asian contribution and the distance from the Central Asia. The southern route for the males was significant ($p = 0.02$).

13. Based on the significant regression line, an 8% excess in Central Asian contribution to the male population of Azerbaijan was detected, which could be associated with the language replacement contribution.

14. The non-significant relationship between the admixture estimates and geographical distances for females were attributed to patrilocality in which higher short-distance mobility of females blurs the pattern.

15. Although there was a 3-7% excess in RLR than the closest neighbors to the north (Northern Cacasus, Armenia, Georgia), there were higher admixtures (1-5 % more in females and at least 25% more in males) in their southern neighbors (Iraq, Iran, Syria, and Lebanon in males; Iraq and Syria in females). Therefore, comparison of the admixture estimates of RLR with their closest neighbors indicated that there is not an especially high contribution from Central Asia in RLR.

16. Presence of a 20% or more admixture proportion in the RLR and presence of even higher contributions around the region suggested that language might not be replaced in accordance with the elite dominance model of Renfrew (1987, 1991).

REFERENCES

Akbaşak, B.S., Budowle, B., Reeder, D.J., Redman, J., Kline, M.C. (2001). Turkish population data with the CODIS multiplex short tandem repeat loci. *Forensic Science International*. 123:227-9.

Akurgal, E. (2003). Anadolu Uygarlıkları. *Net Turistik Yayınlar AŞ*.

Al-Zahery, N., Semino, O., Benuzzi, G., Magri, C., Passarino, G., Torroni, A., Santachiara-Benerecetti, A. S. (2003). Y-chromosome and mtDNA polymorphisms in Iraq, a crossroad of the early human dispersal and of post-Neolithic migrations. *Molecular Phylogenetics and Evolution.* 28(3):458-72.

Anderson, S., Bankier, T., Barrel, B.G., De Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, S., Schreier, P.H., Smith, A.J.H., Staden, R., Young, I.G. (1981). Sequence and organization of the human mitochondrial genome. *Nature* 290: 457–465.

Anghel, A., Marian, C., Pitulescu, M., Daba, A., Sirbu, I.O., Rusu, V., Budowle, B. (2003). Population genetic study of eight short tandem repeat loci CSF1PO, TPOX, TH01, F13A01, FESFPS, vWA, F13B and LPL in the Western Romanian population. *Forensic Science International.* 131(2-3):218-9.

Antunez-de-Mayolo, G., Antunez-de-Mayolo, A., Antunez-de-Mayolo, P., Papiha, S.S., Hammer, M., Yunis, J.J., Yunis, E.J., Damodaran, C., Martinez de Pancorbo, M., Caeiro, J.L., Puzyrev, V.P., Herrera, R.J. (2002). Phylogenetics of worldwide human populations as determined by polymorphic Alu insertions. *Electrophoresis*. 23(19):3346-56.

Aşıcıoğlu, F., Akyüz, F., Çetinkaya, U., Özbek, U. (2002a). Allele distribution data of nine short tandem repeat loci for Turkish population: D3S1358, vWA,FGA, D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820. *Forensic Science International*. 129:75-77.

Aşıcıoğlu, F., Akyuz, F., Çetinkaya, U., Yılmaz, S., Koluaçık, S., Vural, B., Özbek, U. (2002b). Turkish population data on nine short tandem repeat loci: HumCSF1PO, HumTHO1, HumTPOX, HumFES/FPS, HumF13B, HumVWA, D3S1358, D7S820, D16S539. *Forensic Science International*. 126:252-3.

Barbarii, L.E., Rolf, B., Constantinescu, C., Hohoff, C., Calistru, P., Dermengiu, D. (2004). Allele frequencies of 13 short tandem repeat (STR) loci in the Romanian population. *Forensic Science International*. 141(2-3):171-4.

Barbujani, G., Magagni, A., Minch, E., Cavalli-Sforza, L.L (1997). An apportionment of human DNA diversity. *Proceedings of National Academy of Science USA* 94: 4516–4519.

Batzer, M.A., Gudi, V.A., Mena, J. C., Foltz, D.W., Herrera, R.J., Deninger, P.L. (1991). Amplification dynamics of human-specific (HS) Alu family members. *Nucleic Acids Research*. 19(13): 3619 – 3623.

Batzer, M. A. and Deininger, P. L. (1991). A human-specific subfamily of Alu sequences. *Genomics* 9: 481–487.

Batzer, M.A. and Deininger, P.L. (2002). Alu repeats and human genetic diversity. *Nature Review Genetics*. 3: 370-379.

Beaumont, M.A. and Rannala, B. (2004). The Bayesian revolution in genetics. *Nature Reviews Genetics*. 5: 251-261.

Belledi, M., Poloni, E.S., Casalotti, R., Conterio, F., Mikerezi, I., Tagliavini, J., Excoffier, L. (2000). Maternal and paternal lineages in Albania and the genetic structure of Indo-European populations. *European Journal of Human Genetics* 8:480–486.

Benedetto, G., Erguven, A., Stenico, M., Castrì, L., Bertorelle, G., Togan, I., Barbujani, G. (2001). DNA diversity and population admixture in Anatolia. *America Journal of Physical Anthropology*. 115: 144-156.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Wheeler, D.L. (2003). GenBank. *Nucleic Acids Research*. 31: 23-7.

Bertorelle, G., and L. Excoffier, (1998). Inferring admixture proportions from molecular data. *Molecular Biology and Evolution*. 15(10): 1298-1311.

Brega, A., Scacchi, R., Cuccia, M., Kirdar, B., Peloso, G., Corbo, R.M. (1998). Study of 15 protein polymorphisms in a sample of the Turkish population. *Human Biology*. 70(4):715-28.

Brinkmann, B., Junge, A., Meyer, E., Wiegand, P. (1998). Population genetic diversity in relation to microsatellite heterogeneity. *Human Mutation*. 11:135-144.

Burton, M.L., Moore, C.C., Whiting, J.W.M., and Romney, A.K. (1996). Regions based on social structure. *Current Anthropology*. 37**:** 87–123.

Bowcock, A. M., Ruiz-Linares, A., Tomfohrde, J., Minch, E., Kidd, J. R., Cavalli-Sforza, L. L., (1994) High resolution of human evolutionary trees with polymorphic microsatellites. *Nature 368*: 455–457.

Caballero, A. (1995). On the effective size of populations with separate sexes, with particular reference to sex-linked genes. *Genetics*. 139:1007-1011.

Calafell, F., Underhill, P., Tolun, A., Anglicheva, D., Kalaydjieva, L. (1996). From Asia to Europe: mitochondrial DNA sequence variability in Bulgarian and Turks. *Annals of Human Genetics* 60:35–49.

Calafell, F., Comas, D., Perez-Lezaun, A., Bertranpetit, J. (2000). Genetics and Population History of Central Asia. In Renfrew, A.C., and Boyle, K.V., (Eds), 2000, *Archaeogenetics: DNA and the population prehistory of Europe*. Cambridge: McDonald Institute for Archaeological Research.

Cavalli-Sforza, L.L., Menozzi, P., Piazza, A. (1994). The history and geography of human genes. *Princeton University Press, Princeton*.

Cavalli-Sforza, L.L and Feldman, W.F (2003). The application of molecular genetic approaches to the study of human evolution. *Nature Review Genetics*. 33: 266-275.

Chakraborty, R. (1992) Multiple Alleles and Estimation of genetic parameters: computational equations showing involvement of all alleles. *Genetics.* 130: 231 – 234.

Charlesworth, B. (2001). The effect of life-history and mode of inheritance on neutral genetic variability. *Genetical Research Cambridge*. 77:153-166.

Chikhi, L., Bruford, M. W., Beaumont. M. A. (2001). Estimation of Admixture Proportions: A Likelihood-Based Approach Using Markov Chain Monte Carlo, *Genetics,* 158, 1347-1362.

Chikhi, L., Nichols, R.A., Barbujani, G., Beaumont, M.A. (2002). Y genetic data support the Neolithic demic diffusion model. *Proceedings of National Academy of Science U S A. 99(17):* 11008–11013.

Christian, D. (2001). A history of Russia, Central Asia, and Mongolia. *Blackwell Publishers.*

Choisy, M., Franck, P., Cornuet, M. (2004). Estimating admixture proportions with microsatellites: comparison of methods based on simulated data. *Molecular Ecology.*13: 955-968.

Cinnioğlu, C., King, R., Kivisild, T., Kalfoğlu, E., Atasoy, S., Cavalleri, G.L., Lillie, A.S., Roseman, C.C., Lin, A.A., Prince, K., Oefner, P.J., Shen, P., Semino, O., Cavalli-Sforza, L.L., Underhill, P.A (2004). Excavating Y-chromosome haplotype strata in Anatolia *Human Genetics* 114 : 127–148.

Cohen, C. (1968). Pre-Ottoman Turkey. *Toplinger Publishing Co, New York.*

Comas, D., Calafell, F., Mateu, E., Perez-Lezaun, A., Bertranpettit, J. (1996). Geographic variation in human mitochondrial DNA control region sequence: The population history of Turkey and its relationship to the European populations. *Molecular Biology and Evolution*. 13 (8): 1067 – 1077.

Comas, D., Calafell, F., Mateu, E., Perez-Lezaun, A., Bosch, E., Martinez-Arias, R., Clarimon, J., Facchini, F., Fiori, G., Luiselli, D., Pettener, D., Bertranpetit, J. (1998). Trading Genes along the Silk Road: mtDNA Sequences and the Origin of Central Asian Populations. *American Journal of Human Genetics.* 63:1824–1838.

Comas, D., Calafell, F., Benchemsi, N., Helal, A., Lefrane, G., Stoneking, M., Batzer, M.A., Bertranpetit, J., Sajantila, A. (2000). Alu insertion polymorphisms in NW Africa and the Iberian Peninsula: evidence for a strong genetic boundary through the Gibraltar Strait. *Human Genetics*. 107: 312 – 319.

Comas, D., Plaza, S., Wells, R.S., Yuldaseva, N., Lao, O., Calafell, F., Bertranpettit, J. (2004a). Admixture, migrations and dispersal in Central Asia: evidence from maternal DNA lineages. *European Journal of Human Genetics*. 12(6): 495-504.

Comas, D., Schmid, H., Braeuer, S., Flaiz, C., Busquets, A., Calafell, F., Bertranpetit, J., Scheil, H-G., Huckenbeck, W., Efremovska, L., Schmidt, H. (2004b). Alu insertion polymorphisms in the Balkans and the origins of the Aromuns. *Annals of Human Genetics*, 68:120-127.

Çakır, A.H., Çelebioğlu, A., Altunbaş, S. (2002a). STR data for the AmpFlSTR SGM Plus from Marmara region of Turkey. *Forensic Science International.* 127:240-242.

Çakır, A.H., Çelebioğlu, A., Şimşek, F. (2002b). STR data for the AmpFlSTR SGM Plus from Aegean region of Turkey. *Forensic Science International*. 129:137-139.

Çakır, A.H., Çelebioğlu, A., Altunbaş, S., Yardımcı, E. (2003). Allele frequencies for 15 STR loci in Van-Ağrı districts of the Eastern Anatolia region of Turkey. *Forensic Science International*. 135(1):60-3.

Çakır, A.H., Şimşek, F., Katırcı, N., Taşdelen, B. (2004). STR data for the AmpFISTR SGM Plus from the eastern and western sections of Mediterranean region of Turkey. *Forensic Science International.* 142: 55 – 57.

Çerkezi, A.B., Altunçul, D.H., Yükseloğlu, H., Abaci-Kalfoğlu, E., Atasoy, S. (2002). Allele frequencies for three STR loci in Turkish population of Kosovo. *Journal of Forensic Sciences*. 47:908.

Çetinkaya, G., Ülküer, U., Togan, I. (2003). Turkish population data on the CTTV STR loci. *Journal of Forensic Sciences*. 48(1):218.

Deininger, P.L., Batzer, M.A., Hutchison, C.A., Edgell, M.H. (1992). Master genes in mammalian repetitive DNA amplification. *Trends in Genetics*. 8(9): 307 – 311.

Di Giacomo, F., Luca, F., Anagnou, N., Ciavarella, G., Corbo, R. M., Cresta, M., Cucci, F., Di Stasi, L., Agostiano, V., Giparaki, M., Loutradis, A., Mammi', C., Michalodimitrakis, E. N., Papola, F., Pedicini, G., Plata, E., Terrenato, L., Tofanelli, S., Malaspina, P., Novelletto, A. (2003). Clinal patterns of human Y chromosomal diversity in continental Italy and Greece are dominated by drift and founder effects. *Molecular Phylogenetics and Evolution*. 28: 387 – 395.

Dinç, H., and Togan, I. (2005). Allele frequencies for ten Alu insertion polymorphisms in Anatolian population. *Journal of Forensic Sciences.* 50(5):1221-2.

Dupanloup, I. and Bertorelle, G. (2001). Inferring admixture proportions from molecular data: extension to any number of parental populations. *Molecular Biology and Evolution.* 18(4): 672–675.

Dupanloup, I., Bertorelle, G., Chikhi, L., Barbujani, G. (2004). Estimating the Impact of Prehistoric Admixture on the Genome of Europeans. *Molecular Biology and Evolution*. 21(7):1361–1372.

Egyed, B., Furedi, S., Angyal, M., Boutrand, L., Vandenberghe, A., Woller, J., Padar, Z. (2000). Analysis of eight STR loci in two Hungarian populations. *International Journal of Legal Medicine.* 113:272-275.

Esin, E. (1983). Descriptions of Turks and Tatars (Mongols) of the thirteenth century, in some Anatolian sources. In Otto Harrssowit (ed) Documenta Barbarorum. Wiesbaden.

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

Filoğlu, G., Abacı-Kalfoğlu, E., Atasoy, S. (2002). Allele frequencies for 10 STR loci in Istanbul (Turkey) population. *Journal Forensic Science*. 47:909-910.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457-72.

Griffiths, R.C., Tavare, S. (1994). Sampling Theory for Neutral Alleles in a Varying Environment. *Philosophical Transactions: Biological Sciences*, 344(1310) Mathematical and Statistical Aspects of DNA and Protein Sequence Analysis. 403-410.

Goldstein, D.B. and Chikhi, L. (2002). Human migrations and population structure: what we know and why it matters. *Annual Review of Genomics and Human Genetics* 3:129-52.

Goldwin, G. (1994). The Janissaries. *London, Sagi Books.*

Hammer, M.F., Karafet, T., Rasanayagam, A., Wood, E.T., Altheide, T.K., Jenkins, T., Griffiths, R.C., Templeton, A. R., Zegura, S.L. (1998). Out of Africa and back again: nested cladistic analysis of human Y chromosome variation. *Molecular Biology and Evolution.* 15(4):427-41.

Hammer, M.F., Redd, A.J., Wood, E. T., Bonner, M. R., Jarjanazi, H., Karafet, T., Santachiara-Benerecetti, S., Oppenheim, A., Jobling, M. A., Jenkins, T., Ostrer, H., Bonne-Tamir, B. (2000). Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes. *Procidings of National Academy of Science U S A.* 6;97(12):6769-74.

Hammer, M. F., Karafet, T. M., Redd, A. J., Jarjanazi, H., Santachiara-Benerecetti, S., Soodyall, H., Zegura, S. L. (2001). Hierarchical patterns of global human Y-chromosome diversity. *Molecular Biology and Evolution* 18(7):1189-203.

Handt, O., Meyer, S., von Haeseler, A. (1998). Compilation of human mtDNA control region sequences. *Nucleic Acids Research.* 26(1): 126 – 129.

Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*. 405: 907-913.

Hewitt, G. (2004). Genetic Consequences of Climatic Oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences.* 359 (1442): 183-195.

Higgs, P.G and Attwood, T.K. (2005). Bioinformatics and Molecular Evolution. *Blackwell Publishing*.

Higgins, D., Thompson, J., Gibson, T., Thompson, J.D., Higgins, D.G., Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*. 22:4673-4680.

İnalcık, K. (1997). Turkey and Europe: A historical Perspective. *Perceptions, Journal of International Affairs*. Volume II.

Ingman, M., Kaessmann, H, Paabo, S., Gyllensten, U. (2000). Mitochondrial genome variation and the origin of modern humans. *Nature*. 408: 708 – 713.

Iwasa, M., Wiegand, P., Rand, S., Schurenkamp, M., Atasoy, S., Brinkmann, B.(1997). Genetic variation at five STR loci in subpopulations living in Turkey. *International Journal of Legal Medicine*. 110:170-172.

Jobling, M.A., Tyler-Smith, C. (2003). The human Y-chromosome: An evolutionary marker comes of age. *Nature Review Genetics*. 4(8):598-612.

Jobling, M.A., Tyler-Smith, C Hurles, M. (2004). Human Evolutionary Genetics: Origins, Peoples and Disease. *Garland Science.*

Jorde, L.B., Bamshad, M., Rogers, A.R. (1998). Using mitochondrial and nuclear DNA markers to reconstruct human evolution. *BioEssays* 20:126–136.

Jorde, B.L., Watkins, W.S., Bamshad, M.J., Dixon, M.E., Ricker, C.E., Seielstad, M.T., Batzer, M.A. (2000). The Distribution of Human Genetic Diversity: A Comparison of Mitochondrial, Autosomal, and Y-Chromosome Data. *American Journal of Human Genetics.* 66:979–988.

Karafet, T.M., Zegura, S.L., Posukh, O., Osipova, L., Bergen, A., Long, J., Goldman, D., Klitz, W., Harihara, S., de Knijff, P., Wiebe, V., Griffiths, R. C., Templeton, A. R., Hammer, M. F. (1999). Ancestral Asian source(s) of new world Y-chromosome founder haplotypes. *American Journal of Human Genetics.* 64(3):817-31.

Karafet, T., Xu, L., Du, R., Wang, W., Feng, S., Wells, R. S., Redd, A. J., Zegura, S. L., Hammer, M. F. (2001). Paternal population history of East Asia: sources, patterns, and microevolutionary processes. *American Journal of Human Genetics*. 69(3):615-28.

Khitrinskaya, I.Yu., Stepanov, V.A., Puzyrev, V.P., Spiridonova, M.G., Voevoda, M.I. (2003). Genetic differentiation of residents of Central Asia from autosomal marker data. *Genetika*. 39(10):1389-97.

Kolman, C.J., Sambuughin, N., Bermingham, E. (1996). Mitochondrial DNA analysis of Mongolian populations and implications for the origin of New World founders. *Genetics* 142:1321–1334.

Kondopoulou, H., Loftus, R., Kouvatsi, A., Triantaphyllidis, C. (1999). Genetic studies in 5 Greek population samples using 12 highly polymorphic DNA loci. *Human Biology*. 71:27-42.

Kouvatsi, A., Karaiskou, N., Apostolidis, A., Kirmizidis, G. (2001). Mitochondrial DNA sequence variation in Greeks. *Human Biology*. 73(6):855-69.

Kovatsi, L., Parsons, T.J., Just, R.S., Irwin, J.A.(2006). Genetic variation for 15 autosomal STR loci (PowerPlex 16) in a population sample from northern Greece. *Forensic Science International.* 159(1):61-3.

Kubat, M., Skavic, J., Behluli, I., Nuraj, B., Bekteshi, T., Behluli, M., Klaric, I.M., Pericic, M. (2004). Population genetics of the 15 AmpF lSTR Identifiler loci in Kosovo Albanians. *International Journal of Legal Medicine*. 118:115.

Lahermo, P., Laıtınen, V., Sıstonen, P., Beres, J., Karcag, V., Savontaus, M. (2000). MtDNA polymorphism in the Hungarians: comparison to three other Finno-Ugric-speaking populations. *Hereditas.* 132: 35-42.

Lahr, M.M. and Foley, R.A. (1998). Towards a theory of modern human origins: geography, demography, and diversity in recent human evolution. *Yearbook of Physical Anthropology*. 41: 137 – 176.

Lewis, B. (1995). The Middle East: A brief history of the last 2000 years. *A Touchstone book, New York.*

Long, J.C. (1991). The genetic structure of admixed populations. *Genetics.* 127: 417 – 428.

Macaulay,V., Richards, M., Hickey, E., Vega, E., Cruciani, F., Guida, V., Scozzari, R., Bonne´-Tamir, B., Sykes, B., Torroni,A. (1999). The Emerging Tree of West Eurasian mtDNAs: A Synthesis of Control-Region Sequences and RFLPs. *American Journal of Human Genetics.* 64:232–249.

Mansoor, A., Mazhar, K., Khaliq, S., Hameed, A., Rehman, S., Siddiqi, S., Papaioannou, M., Cavalli-Sforza, L.L., Mehdi, S.Q., Ayub, Q. (2004) Investigation of the Greek ancestry of populations from northern Pakistan. *Human Genetics* 114(5):484-90.

Manz, B.F (1994). "Historical Background" in Manz, B.F (ed) Central Asia in Historical perspective. Boulder, Colorado. *Westview Press.*

MINITAB13 package program , Minitab Inc., State College, PA, USA.

Miyazaki, S., Sugawara, H., Gojobori, T., Tateno, Y. (2003). DNA Databank of Japan (DDBJ) in XML. *Nucleic Acids Research*. 31: 13-16.

Nasidze, I., Risch, G.M., Robichaux, M., Sherry, S.T., Batzer, M.A., Stoneking, M. (2001). Alu insertion polymorphisms and the genetic structure of human populations from the Caucasus. *European Journal of Human Genetics*. 9: 267 – 272.

Nasidze, I ve Stoneking, M. (2001). Mitochondrial DNA variation and language replacement in the Caucasus. *Proceedings of Royal Society London B.* 268: 1197-1206.

Nasidze, I., Sarkisian, T., Kerimov, A., Stoneking, M. (2003). Testing hypotheses of language replacement in the Caucasus: evidence from the Y-chromosome. *Human Genetics*. 112(3):255-61.

Nasidze, I., Ling, Y., Quinque, D., Dupanloup, I., Cordaux, R., Rychkov, S., Naumova, O., Zhukova, O., Sarraf-Zadegan, N., Naderi, G., Asgary, S., Sardas, S., Farhud, D., Sarkisian, T., Asadov, C., Kerimov, A., Stoneking, M. (2004). Mitochondrial DNA and Y-Chromosome variation in the Caucasus. *Annals of Human Genetics*. 68: 205-221.

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

Oota, H., Settheetham-Ishida, W., Tiwawech, D., Ishida, T., Stoneking, M. (2001). Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. *Nature Genetics*. 29: 20-21.

Ota, T. (1993). DISPAN: Genetic distance and phylogenetic analysis. State College, PA: Institute of Molecular Evolutionary Genetics, The Pennsylvania State University.

Rajeevan, H., Cheung, K., Gadagkar, R., Stein, S., Soundararajan, U., Kidd, J. R., Pakstis, A. J., Miller, P.L., Kidd, K.K. (2005) ALFRED: An Allele Frequency Database for Microevolutionary Studies. *Nucleic Acids Research.*31(1):270-271.

Ramachandran, S., Deshpande, O., Roseman, C.C., Rosenberg, N.A., Feldman, M.W., Cavalli-Sforza, L.L (2005). Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proceedings of National Academy of Science U S A*. 102: 15942–15947.

Renfrew, C. (1987) Archaeology and Language. The puzzle of Indo-European origins. *Cambridge University Press.*

Renfrew, C (1991). Before Babel: Speculations on the origins of linguistic diversity. *Cambridge Archeological Journal* 1(1): 3-23.

Renfrew, C (2000). At the edge of Knowability: Towards a prehistory of language. *Cambridge Archeological Journal.* 10 (1): 7-34.

Rholf, F.G. (2000). Numerical Taxonomy and Multivariate Analysis System. NTSYSpc. Version 2.1. Exeter Software.

Richards, M., Macaulay,V., Hickey, E., Vega, E., Sykes, B., Guida, V., Rengo, C., Sellitto, D., Cruciani, F., Kivisild,T., Villems, R., Thomas,M., Rychkov, S., Rychkov,O., Rychkov, Y., Gölge,M., Dimitrov,D., Hill,E., Bradley,D., Romano, V., Calı`, F., Vona,G., Demaine,A., Papiha, S., Triantaphyllidis,C., Stefanescu,G., Hatina,J., Belledi,M., Di Rienzo,A., Novelletto,A., Oppenheim,A., Nørby,S., Al-Zaheri,N., Santachiara-Benerecetti,S., Scozzari,R., Torroni,A., Bandelt, H. (2000) Tracing European Founder Lineages in the Near Eastern mtDNA Pool. *American Journal of Human Genetics*. 67: 1251-1276.

Robert, D.F. and Hiorns, R.W. (1965). Methods of analysis of the genetic composition of a hybrid population. *Human Biology* 37:38–43.

Roewer, L., Krawczak, M., Willuweit, S., Nagy, M., Alves, C., Amorim, A., Anslinger, K., Augustin, C., Betz, A., Bosch, E., Caglia, A., Carracedo, A., Corach, D., Dekairelle, A.F., Dobosz, T., Dupuy, B.M., Furedi, S., Gehrig, C., Gusmao, L., Henke, J., Henke, L., Hidding, M., Hohoff, C., Hoste, B., Jobling, M.A., Kargel, H.J., de Knijff, P., Lessig, R., Liebeherr, E., Lorente, M., Martinez-Jarreta, B., Nievas, P., Nowak, M., Parson, W., Pascali, V.L., Penacino, G., Ploski, R., Rolf, B., Sala, A., Schmidt, U., Schmitt, C., Schneider, P.M., Szibor, R., Teifel-Greding, J., Kayser, M. (2001). Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes. *Forensic Science International.* 118(2-3):106-13.

Rolf, B., Röhl, A., Forster, P., Brinkmann, B. (1999). On the genetic origins of the Turks study of six Y-chromosomal short tandem repeats. In: Papiha S, Deka R, Chakraborty R (eds) Genomic diversity:applications in human population genetics. Kluwer Academic/Plenum, New York, pp 75–82.

Romualdi, C., Balding, D., Nasidze, I.S., Risch, G., Robichaux, M., Sherry, S.T., Stoneking, M., Batzer, M.A., Barbujani, G (2002). Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. *Genome Research* 12: 602-612.

Rossabi, M. (1994) "The Legacy of the Mongols" in Manz, B.F (ed) Central Asia in Historical perspective. Boulder, Colorado. *Westview Press.*

Rosser, Z.H., Zerjal, T., Hurles, M. E., Adojaan, M., Alavantic, D., Amorim, A., Amos, W., Armenteros, M., Arroyo, E., Barbujani, G., Beckman, G., Beckman, L., Bertranpetit, J., Bosch, E., Bradley, D. G., Brede, G., Cooper, G., Corte-Real, H. B., de Knijff, P., Decorte, R., Dubrova, Y. E., Evgrafov, O., Gilissen, A., Glisic, S., Golge, M., Hill, E. W., Jeziorowska, A., Kalaydjieva, L., Kayser, M., Kivisild, T., Kravchenko, S. A., Krumina, A., Kucinskas, V., Lavinha, J., Livshits, L. A., Malaspina, P., Maria, S., McElreavey, K., Meitinger, T. A., Mikelsaar, A. V., Mitchell, R. J., Nafa, K., Nicholson, J., Norby, S., Pandya, A., Parik, J., Patsalis, P. C., Pereira, L., Peterlin, B., Pielberg, G., Prata, M. J., Previdere, C., Roewer, L., Rootsi, S., Rubinsztein, D. C., Saillard, J., Santos, F. R., Stefanescu, G., Sykes, B. C., Tolun, A., Villems, R., Tyler-Smith, C., Jobling, M. A. (2000). Y-chromosomal diversity in Europe is clinal and influenced primarily by geography, rather than by language. *American Journal of Human Genetics*. 67(6):1526-43.

Roux, J-P. (1997). Türklerin Tarihi: Büyük Okyanus'tan Akdeniz'e İki Bin Yıl. *Milliyet Yayınları.*

Roy-Engel, A.M., Carroll, M.L., Vogel, E., Garber, R.K., Nguyen, S.V., Salem, A., Batzer, M.A., Deininger, P.L. (2001). Alu Insertion Polymorphisms for the Study of Human Genomic Diversity. *Genetics* 159: 279–290.

Ruhlen, M. (1991). A guide to the World`s languages. Volume 1: Classification. *Standford University Press*. Standford, California.

Salman, H. (2004). "Türk adı, Türklerin anayurdu ve göçleri" in Öztürk, C (ed) Türk Tarihi ve kültürü. *Pegem A Yayıncılık.*

Sanchez-Diz, P., Lareu, M.V., Brion, M., Skitsa, I., Carracedo, A. (2002). STR data for the AmpFlSTR Profiler Plus loci from Greece. *Forensic Science International.* 126:265-6.

Semino, O., Passarino, G., Oefner, P.J., Lin, A.A., Arbuzova, S., Beckman, L.E., benedictis, G.D., Francalacci, P., Kouvatis, A., Limborska, S., Marcikiae, M., Mika, A., Mika, B., Primorac, D., Santachiara-Benerecetti, A.S., Cavalli-Sforza, L.L., Underhill, P.A. (2000). The genetic legacy of Paleolithic Homo sapiens sapiens in extant Europeans: A Y-chromosome perspective. *Science.* 290: 1155-1159.

Shen, M.R., Batzer, M.A., Deininger, P.L. (1991). Evolution of the master Alu gene(s). *Journal of Molecular Evolution*. 33(4):311-20.

Shields, G.F,, Schmiechen, A.M,, Frazier, B.L., Redd, A., Voevoda, M.I., Reed, J.K., Ward, R. H. (1993). mtDNA sequences suggest a recent evolutionary divergence for Beringian and northern North American populations. *American Journal of Human Genetics.* 53(3):549-62.

Skitsa, I., Salas, A., Lareu, M.V., Carracedo, A. (2003). STR-CODIS typing in Greece. *Forensic Science International*. 137(1):104-6.

Starikovskaya, Y.B., Sukernik, R.I., Schurr, T.G., Kogelnik, A.M., Wallace, D.C. (1998). MtDNA diversity in Chukchi and Siberian Eskimos: implications for the genetic history of ancient Beringia and the peopling of the New World. *American Journal of Human Genetics*. 63:1473–1491.

Stoesser, G., Baker, W., Van der Broek, A., Garcia-Pastor, M., Kanz, C., Kulikova, T., Leinonen, R., Lin, Q., Lombard, P., Tuli, M.A., Tzouvara, K., Vaughan, R. (2003). The EMBL Nucleotide Sequence Database: Major new developments. *Nucleic Acids Research*. 31: 17-22.

Szabo, A., Schurenkamp, M., Huhne, J. (1998). Hungarian population data for six STR loci. *International Journal of Legal Medicine*. 111:49-51.

Şekeryapan, C. (2005). An expended study on the Alu insertion polymorphisms in Anatolian human populations. MSc Thesis, METU.

Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105:437-460.

Takeshita, H., Meyer, E., Brinkmann, B. (1997). The STR loci HumTPO and HumLPL: population genetic data in eight populations. *International Journal of Legal Medicine*. 110:331-3.

Tambets, K., Kivisild, T., Metspalu, E., Parik, J., Kaldma, K., Laos, S., Tolk, H.V., Golge, M., Demirtas, H., Geberhiwot, T., Papiha, S.S., de Stefano, G.F., Villems, R., (2000). The topology of the maternal lineages of the Anatolian and Trans-Caucasus populations and the peopling of Europe: some preliminary considerations. In: Renfrew, C., Boyle, K. (Eds.), Archaeogenetics: DNA and the Population Prehistory of Europe. Cambridge, pp. 219–235.

Ullu, E. and Tschudi, C. (1984) Alu sequences are processed 7SL RNA genes. *Nature.* 8(14);312(5990):171-2.

Umar, B. (1999). İlkçağda Türkiye Halkı. *İnkılap Yayınları*.

Underhill, P.A., Passarino, G., Lin, A.A., Shen, P., Lahr, M.M., Foley, R.A., Oefner, P.J., Cavalli-Sforza, L.L. (2001). The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Annals of Human Genetics.* 65: 43-62.

Ülküer, U., Kurtuluş-Ülküer, M., Elma, C., Kesici, T., Menevse, S. (2004). Short tandem repeat (STR) polymorphisms in Turkish population. *Journal of Genetics.* 83:197-199.

Vernesi, C., Benedetto, G., Caramelli, D., Secchieri, E., Simoni, L., Katti, E., Malaspina, P., Novelletto, A., Terribile Wiel Marin, V., Barbujani, G. (2001). Genetic characterization of the body attributed to the evangelist Luke. *Proceedings of National Academy of Science USA*. 98: 13460-13463.

Vryonis, S. (1971). The Decline of medieval Hellenism in Asia Minor and the process of Islamization from the eleventh through the fifteenth century. *University of California press.*

Vural, B., Poda, M., Atlıoğlu, E., Kolusayın, O., Cenani, A., Morling, N., Tümer, Z. (1998). Turkish population data on the short tandem repeat locus TPOX. *International Journal of Legal Medicine*. 111:105-106.

Wang, J. (2003). Maximum-Likelihood Estimation of Admixture Proportions From Genetic Data. *Genetics* 164: 747–765.

Wells, R. S., Yuldasheva, N., Ruzibakiev, R., Underhill, P. A., Evseeva, I., Blue-Smith, J., Jin, L., Su, B., Pitchappan, R., Shanmugalakshmi, S., Balakrishnan, K., Read, M., Pearson, N. M., Zerjal, T., Webster, M. T., Zholoshvili, I., Jamarjashvili, E., Gambarov, S., Nikbin, B., Dostiev, A., Aknazarov, O., Zalloua, P., Tsoy, I., Kitaev, M., Mirrakhimov, M., Chariev, A., Bodmer, W. F. (2001) The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proceedings of National Academy of Science U S A.* 98(18):10244-9.

Xiao, F.X., Yang, J.F., Cassiman, J.J., Decorte, R. (2002) Diversity at eight polymorphic Alu insertion loci in Chinese populations shows evidence for European admixture in an ethnic minority population from northwest China. *Human Biology*. 74(4):555-68

The Y Chromosome Consortium (2002). A Nomenclature System for the Tree of Human Y-Chromosomal Binary Haplogroups. *Genome Research.* 12(2): 339-348.

Yao, Y., Lü, X., Luo, H., Li, W., Zhang, Y. (2000). Genetic admixture in the Silk Road region of China: Evidence for mtDNA and melanocortin 1 receptor polymorphism. *Genes and Genetic Systems.* 75: 173-178.

Yavuz, I. and Sarıkaya, A. (2005). Turkish population data for 15 STR Loci by Multiplex PCR . *Journal Forensic Science*. 50:737-738.

Zerjal T, Wells RS, Yuldasheva N, Ruzibakiev R, Tyler-Smith C (2002). A genetic landscape reshaped by recent events: Y-chromosomal insights into Central Asia. *American Journal of Human Genetics.* 71:466-482.

Zhivotovsky ,L.A., Rosenberg, N.A., Feldman, M.W. (2003). Features of Evolution and Expansion of Modern Humans, Inferred from Genomewide Microsatellite Markers. *American Journal of Human Genetics*. 72:1171–1186.

Zhu, J., Li, J., Guo, Y., Liu, K., Zhu, B., Liu, Y. (2005). Population data of 15 STR in Chinese Han population from north of Guangdong. *Journal of Forensic Sciences*. 50:1510-1511.

APPENDIX A: Frequencies of the mtDNA haplotypes for each population

Number of individuals from parental and hybrid populations in each haplotype

Haplotype	Turkey	Central Asia	Balkans	Georgia	Azerbaijan	Armenia	Northern Caucasus	Syria	Iraq
1	31	20	64	17	3	23	31	8	8
2	1	0	0	0	0	1	0	0	0
3	1	0	0	0	0	0	0	0	0
4	1	0	0	0	0	0	0	0	0
5	2	0	1	0	0	1	0	0	0
6	1	0	1	0	0	1	0	0	0
7	1	0	0	0	0	0	0	0	0
8	3	2	18	1	0	0	2	2	0
9	1	0	0	0	0	0	0	0	0
10	1	0	0	0	0	0	0	0	0
11	1	0	0	0	0	0	0	0	1
12	3	0	1	0	0	1	0	0	0
13	1	0	0	0	0	0	0	1	0
14	1	0	0	0	0	0	0	0	0
15	1	0	0	0	0	0	0	0	0
16	1	0	0	0	0	0	0	0	0
17	2	0	0	0	0	0	0	0	0
18	1	0	17	0	0	1	0	0	0
19	1	0	0	0	0	0	0	0	0
20	1	0	2	0	0	0	0	0	0
21	5	1	22	1	0	0	6	2	0
22	1	0	1	0	0	0	0	0	1
23	1	1	0	0	0	0	0	0	0
24	2	0	1	0	0	0	0	0	0
25	1	2	0	2	1	2	1	0	0
26	1	0	1	0	0	2	0	1	0
27	1	0	0	0	0	0	0	0	0
28	4	2	1	5	0	2	5	1	1
29	1	0	0	0	0	0	0	0	1
30	1	0	0	0	0	0	0	0	0
31	1	0	0	0	0	0	0	0	0
32	1	0	0	0	0	0	0	0	0
33	1	0	0	0	0	0	0	0	0
34	1	0	0	0	0	0	0	0	0
35	1	0	0	0	0	0	0	0	0
36	1	3	1	0	0	0	0	0	0
37	6	0	14	0	2	4	0	2	2
38	3	0	0	1	0	0	0	1	0
39	1	0	0	0	0	0	0	0	0
40	1	0	0	0	0	0	0	0	0
41	1	0	0	1	0	0	0	0	0
42	2	0	0	0	1	1	0	1	0

Appendix A (continued)

43	1	0	0	0	1	0	0	0	0
44	1	1	2	0	0	0	0	0	0
45	1	0	0	0	0	0	0	0	0
46	1	0	0	0	0	0	0	0	0
47	1	0	0	0	0	0	0	0	0
48	1	0	0	0	0	0	0	0	0
49	3	0	4	2	0	2	0	0	0
50	1	0	0	0	0	0	0	0	0
51	1	0	0	0	0	1	0	0	0
52	1	0	0	0	0	0	0	0	0
53	1	0	0	0	0	0	0	0	0
54	1	0	0	0	0	0	0	0	0
55	1	0	0	0	0	0	0	0	0
56	1	0	0	0	0	0	0	0	0
57	1	0	0	0	0	0	0	0	0
58	1	0	0	0	0	0	0	0	0
59	4	1	1	0	0	1	0	0	0
60	1	2	0	0	0	0	0	0	0
61	1	3	0	0	0	0	0	0	0
62	1	2	0	0	0	0	0	0	0
63	3	24	2	2	0	0	2	0	0
64	1	0	0	0	0	0	0	0	0
65	1	0	0	0	0	0	0	0	0
66	1	0	0	0	0	0	0	0	0
67	1	0	0	0	0	0	0	0	0
68	1	0	0	0	0	0	0	0	0
69	1	0	0	0	0	0	0	0	0
70	1	0	0	0	0	0	0	0	0
71	1	0	0	0	0	0	0	0	0
72	1	0	0	0	0	1	0	0	0
73	1	0	0	0	0	0	0	0	0
74	1	0	0	0	0	0	0	0	0
75	1	0	0	0	0	0	0	0	0
76	1	0	1	0	0	0	0	0	0
77	2	0	0	0	0	1	0	0	0
78	1	0	0	0	0	0	0	0	0
79	1	1	2	0	0	0	7	0	0
80	1	0	0	0	0	0	0	0	0
81	1	4	0	0	0	0	0	0	0
82	1	0	0	0	0	0	0	0	0
83	1	0	0	0	0	0	0	0	0
84	1	0	0	0	0	0	0	0	0
85	1	0	0	0	0	0	0	0	0
86	1	5	0	2	1	1	0	0	0
87	1	0	0	0	0	0	0	0	0
88	1	0	4	0	0	0	0	0	0
89	1	0	0	0	0	1	0	0	0
90	2	0	1	0	0	0	2	0	0
91	1	1	0	0	0	0	1	0	0
92	1	0	0	0	0	0	0	0	0

Appendix A (continued)

93	2	9	14	5	3	1	3	0	1
94	1	0	0	0	0	0	0	0	0
95	1	0	0	0	0	0	0	0	0
96	1	0	0	0	0	0	0	0	0
97	1	0	0	0	0	0	0	0	0
98	1	0	0	0	0	0	0	0	0
99	1	0	0	0	1	0	0	0	0
100	1	0	0	0	0	0	0	0	0
101	1	0	0	0	0	0	0	0	0
102	6	1	13	2	1	8	0	2	2
103	1	0	0	0	0	0	0	0	0
104	1	0	0	0	0	0	0	0	0
105	1	0	0	0	0	0	0	0	0
106	1	0	0	0	0	0	0	0	0
107	1	0	0	0	0	0	0	0	0
108	2	0	2	3	1	2	1	1	1
109	1	0	0	0	0	0	0	0	0
110	1	0	1	0	0	0	0	0	1
111	1	0	0	0	0	0	0	0	0
112	1	0	0	0	0	0	0	0	0
113	1	0	0	0	0	1	0	0	0
114	1	1	9	2	1	0	0	0	0
115	1	1	0	0	0	0	0	0	0
116	1	0	1	0	0	0	0	0	0
117	1	0	0	0	0	0	0	0	0
118	6	4	15	0	1	4	0	2	6
119	1	0	0	0	0	0	0	0	0
120	2	0	0	0	0	0	0	0	0
121	1	0	0	0	0	0	0	0	0
122	1	0	1	0	0	0	0	1	0
123	1	0	0	0	0	0	0	0	0
124	1	0	6	0	0	0	0	0	0
125	1	0	2	0	0	0	0	0	0
126	1	0	0	0	0	0	0	0	0
127	1	0	0	0	0	0	0	0	0
128	1	0	0	0	0	0	0	0	0
129	1	0	1	0	0	2	0	0	0
130	1	0	0	0	0	0	0	0	0
131	1	0	0	0	0	0	0	0	0
132	1	1	1	0	0	0	0	0	0
133	2	0	1	0	1	0	0	0	1
134	1	0	0	0	0	0	0	0	0
135	1	0	0	0	0	0	0	0	0
136	1	1	3	0	0	0	0	0	2
137	1	0	0	0	0	0	0	0	0
138	1	0	0	0	0	0	0	0	0
139	1	0	0	0	0	0	0	0	0
140	1	0	1	0	0	1	0	0	0
141	1	0	0	0	0	0	0	0	0
142	1	0	0	0	0	0	0	0	1

Appendix A (continued)

143	4	0	14	1	1	1	2	3	2
144	1	0	1	0	0	0	0	0	0
145	1	0	0	0	0	0	0	0	0
146	1	0	0	0	0	0	0	0	0
147	1	0	2	0	0	0	0	0	0
148	1	0	2	0	0	0	0	0	0
149	1	0	7	0	0	0	0	0	0
150	1	0	1	0	0	4	1	0	0
151	1	0	0	0	0	0	0	0	0
152	1	0	0	0	0	0	0	0	0
153	1	0	0	0	0	0	0	0	0
154	1	0	0	0	0	0	0	0	0
155	1	0	1	0	0	0	0	1	0
156	1	0	0	0	0	0	0	0	0
157	1	0	0	0	0	1	0	0	1
158	1	0	3	0	0	0	0	0	0
159	1	0	0	0	0	0	0	0	0
160	1	0	0	0	0	0	1	0	0
161	1	0	0	0	0	0	0	0	0
162	1	0	0	0	0	0	0	0	0
163	1	0	0	0	0	0	0	0	0
164	1	0	0	0	0	0	0	0	0
165	1	0	0	0	0	0	0	0	0
166	3	3	14	0	0	0	1	1	1
167	1	1	0	0	0	0	0	0	0
168	1	0	1	0	0	0	0	0	0
169	1	0	0	0	0	0	0	0	0
170	1	0	0	0	0	0	0	0	0
171	7	4	16	0	1	2	2	1	1
172	2	2	2	1	1	0	9	1	0
173	1	0	0	0	0	0	0	0	0
174	1	0	0	0	0	0	0	0	0
175	1	1	4	0	0	0	0	0	0
176	2	0	4	0	1	2	0	0	0
177	1	0	0	0	0	1	0	0	0
178	1	0	0	0	1	0	0	0	0
179	2	0	2	0	0	1	1	1	0
180	1	0	0	0	0	0	0	0	0
181	2	0	0	0	2	0	0	0	0
182	1	0	0	0	0	0	0	0	0
183	1	0	0	0	0	0	0	0	0
184	1	0	0	0	0	0	0	0	0
185	1	0	0	0	0	0	0	0	0
186	3	0	0	0	0	2	1	0	0
187	1	0	0	0	0	0	0	0	0
188	1	0	0	0	0	0	0	0	0
189	1	0	0	0	0	1	0	0	0
190	1	0	1	0	0	0	0	1	0
191	1	0	0	0	0	0	0	0	0
192	1	0	0	0	0	0	0	0	0

Appendix A (continued)

193	1	0	8	0	0	0	0	0	0
194	1	0	0	0	0	0	0	0	0
195	1	0	0	0	0	0	0	0	0
196	1	0	0	0	0	0	0	0	0
197	1	0	0	0	0	0	0	0	0
198	1	0	0	0	0	0	0	0	0
199	0	1	0	0	0	0	0	0	0
200	0	4	0	0	1	0	1	0	0
201	0	1	0	0	0	0	0	0	0
202	0	1	0	0	0	0	0	0	0
203	0	2	0	0	0	0	0	0	0
204	0	1	0	0	0	0	0	0	0
205	0	1	0	0	0	0	0	0	0
206	0	1	0	0	0	0	0	0	0
207	0	6	0	0	1	0	0	0	0
208	0	1	0	0	0	0	0	0	0
209	0	3	0	1	0	0	0	0	0
210	0	5	0	0	0	0	0	0	0
211	0	3	0	0	0	0	0	0	0
212	0	1	0	0	0	0	0	0	0
213	0	1	0	0	0	0	0	0	0
214	0	1	0	0	0	0	0	0	0
215	0	1	0	0	0	0	0	0	0
216	0	1	0	0	0	0	0	0	0
217	0	1	0	0	0	0	0	0	0
218	0	1	0	0	0	0	0	0	0
219	0	3	0	0	0	0	0	0	0
220	0	1	0	0	0	0	0	0	0
221	0	1	0	0	0	0	0	0	0
222	0	1	0	0	0	0	0	0	0
223	0	2	0	0	0	0	0	0	0
224	0	1	0	0	0	0	0	0	0
225	0	1	0	0	0	0	0	0	0
226	0	3	0	0	0	0	0	0	0
227	0	3	0	0	0	0	0	0	0
228	0	1	0	0	0	0	0	0	0
229	0	2	0	0	0	0	0	0	0
230	0	2	0	0	0	0	2	0	0
231	0	1	0	0	0	0	0	0	0
232	0	1	0	0	0	0	0	0	0
233	0	5	0	0	0	0	0	0	0
234	0	1	0	0	0	0	0	0	0
235	0	1	0	0	0	0	0	0	0
236	0	1	0	0	0	0	0	0	0
237	0	3	0	0	0	0	0	0	0
238	0	1	0	0	0	0	0	0	0
239	0	1	0	0	0	0	0	0	0
240	0	1	0	0	0	0	0	0	0
241	0	1	0	0	0	0	0	0	0
242	0	1	0	0	0	0	0	0	0

Appendix A (continued)

243	0	2	11	0	3	0	3	1	0
244	0	2	0	0	0	0	0	0	0
245	0	1	0	0	0	0	0	0	0
246	0	1	0	0	0	0	0	0	0
247	0	1	0	0	0	0	0	0	0
248	0	2	0	0	0	0	0	0	0
249	0	1	0	0	0	0	0	0	0
250	0	3	0	0	0	0	0	0	0
251	0	1	7	0	0	0	0	0	0
252	0	2	0	0	0	0	0	0	0
253	0	2	0	0	0	0	0	0	0
254	0	1	0	0	0	0	0	0	0
255	0	2	0	0	0	0	0	0	0
256	0	1	2	0	0	0	0	0	0
257	0	1	0	0	0	0	0	0	0
258	0	1	0	0	0	0	1	1	0
259	0	1	0	0	0	0	0	0	0
260	0	1	0	0	0	0	0	0	0
261	0	1	0	0	0	0	0	0	0
262	0	1	0	0	0	0	0	0	0
263	0	7	0	0	0	0	0	0	0
264	0	2	0	0	0	0	0	0	0
265	0	1	0	0	0	0	0	0	0
266	0	1	0	0	0	0	0	0	0
267	0	1	0	0	0	0	0	0	0
268	0	1	0	0	0	0	0	0	0
269	0	1	0	0	0	0	0	0	0
270	0	1	0	0	0	0	0	0	0
271	0	2	0	0	0	0	0	0	0
272	0	2	0	0	0	0	0	0	0
273	0	1	0	0	0	0	0	0	0
274	0	1	0	0	0	0	0	0	0
275	0	1	0	0	0	0	0	0	0
276	0	1	0	0	0	0	0	0	0
277	0	2	0	0	0	0	0	0	0
278	0	1	0	0	0	0	0	0	0
279	0	1	0	0	0	0	1	0	0
280	0	1	0	0	0	0	0	0	0
281	0	1	5	2	0	1	5	1	0
282	0	3	0	0	0	0	0	0	0
283	0	1	0	0	0	0	0	0	0
284	0	1	0	0	0	0	0	0	0
285	0	1	0	0	0	0	0	0	0
286	0	1	0	0	0	0	0	0	0
287	0	1	0	0	0	0	0	0	0
288	0	6	0	0	0	0	0	0	0
289	0	2	1	0	0	0	0	0	0
290	0	1	0	0	0	0	0	0	0
291	0	2	8	1	0	0	1	0	0
292	0	2	0	0	0	0	0	0	0

Appendix A (continued)

293	0	2	8	3	1	0	4	2	1
294	0	1	0	0	0	0	0	0	0
295	0	1	0	0	0	0	0	0	0
296	0	2	3	0	1	0	2	2	0
297	0	2	0	0	0	0	0	0	0
298	0	1	0	0	0	0	0	0	0
299	0	1	0	0	0	0	0	0	0
300	0	1	0	0	0	0	0	0	0
301	0	1	0	0	0	0	0	0	0
302	0	1	0	0	0	0	0	0	0
303	0	2	0	0	0	0	0	0	0
304	0	1	0	0	0	0	0	0	0
305	0	1	2	0	0	0	0	0	0
306	0	1	0	0	0	0	0	0	0
307	0	7	0	0	0	0	0	0	1
308	0	1	0	0	0	0	0	0	0
309	0	1	0	0	0	0	0	0	0
310	0	2	0	0	0	0	0	0	0
311	0	1	0	0	0	0	0	0	0
312	0	1	0	0	0	0	0	0	0
313	0	1	0	0	0	0	0	0	0
314	0	2	0	0	0	0	0	0	0
315	0	3	0	0	0	0	0	0	0
316	0	1	0	0	0	0	0	0	0
317	0	2	0	0	0	0	0	0	0
318	0	1	0	0	0	0	0	0	0
319	0	1	0	0	0	0	0	0	0
320	0	1	0	0	0	0	0	0	0
321	0	1	0	0	0	0	0	0	0
322	0	1	0	0	0	0	0	0	0
323	0	1	0	0	0	0	0	0	0
324	0	1	0	0	0	0	0	0	0
325	0	1	0	0	0	0	0	0	0
326	0	1	0	0	0	0	0	0	0
327	0	1	0	0	0	0	0	0	0
328	0	1	0	0	0	0	0	0	0
329	0	1	0	0	0	0	0	0	0
330	0	1	0	0	0	0	0	0	0
331	0	1	0	0	0	0	0	0	0
332	0	1	0	0	0	0	0	0	0
333	0	1	0	0	0	0	0	0	0
334	0	1	0	0	0	0	0	0	0
335	0	1	0	0	0	0	0	0	0
336	0	3	0	0	0	0	0	0	0
337	0	1	0	0	0	0	0	0	0
338	0	3	1	0	0	0	0	0	0
339	0	1	0	0	0	0	0	0	0
340	0	1	0	0	0	0	0	0	0
341	0	1	0	0	0	0	0	0	0
342	0	1	0	0	0	0	0	0	0

Appendix A (continued)

343	0	1	0	0	0	0	0	0	0
344	0	1	0	0	0	0	0	0	0
345	0	1	0	0	0	0	0	0	0
346	0	1	0	0	0	0	0	0	0
347	0	3	0	0	0	0	0	0	0
348	0	1	0	0	0	0	0	0	0
349	0	1	0	0	0	0	0	0	0
350	0	1	0	0	0	0	0	0	0
351	0	2	0	0	0	0	0	0	0
352	0	1	0	0	0	0	0	0	0
353	0	1	0	0	0	0	0	0	0
354	0	1	0	0	0	0	0	0	0
355	0	1	0	0	0	0	0	0	0
356	0	2	0	0	0	0	0	0	0
357	0	1	0	0	0	0	0	0	0
358	0	1	0	0	0	0	0	0	0
359	0	1	0	0	0	0	0	0	0
360	0	1	0	0	0	0	0	0	0
361	0	2	0	0	0	0	0	0	0
362	0	1	2	0	0	0	0	0	0
363	0	1	0	0	0	0	0	0	0
364	0	1	1	0	0	1	1	0	0
365	0	3	0	0	0	0	0	0	0
366	0	2	0	0	0	0	0	0	0
367	0	1	0	0	0	0	0	0	0
368	0	1	0	0	0	0	0	0	0
369	0	1	0	0	0	0	0	0	0
370	0	1	0	0	0	0	0	0	0
371	0	1	0	0	0	0	0	0	0
372	0	1	0	0	0	0	0	0	0
373	0	1	0	0	0	0	0	0	0
374	0	1	0	0	1	0	0	0	0
375	0	1	0	0	0	0	0	0	0
376	0	2	0	0	0	0	0	0	0
377	0	1	0	0	0	0	0	0	0
378	0	1	0	0	0	0	0	0	0
379	0	1	0	0	0	0	0	0	0
380	0	1	0	0	0	0	0	0	0
381	0	1	0	0	0	0	0	0	0
382	0	2	0	0	0	0	0	0	0
383	0	1	0	0	0	0	0	0	0
384	0	1	0	0	0	0	0	0	0
385	0	1	0	0	0	0	0	0	0
386	0	2	0	0	0	0	0	0	0
387	0	1	0	0	0	0	0	0	0
388	0	1	0	0	0	0	0	0	0
389	0	1	0	0	0	0	0	0	0
390	0	1	0	0	0	0	0	0	0
391	0	1	0	0	0	0	0	0	0
392	0	1	0	0	0	0	0	0	0

Appendix A (continued)

393	0	1	0	0	0	0	0	0	0
394	0	1	0	0	0	0	0	0	0
395	0	1	0	0	0	0	0	0	0
396	0	1	0	0	0	0	0	0	0
397	0	1	0	0	0	0	0	0	0
398	0	1	0	0	0	0	0	0	0
399	0	1	0	0	0	0	0	0	0
400	0	1	0	0	0	0	0	0	0
401	0	1	0	0	0	0	0	0	0
402	0	1	0	0	0	0	0	0	0
403	0	2	0	0	0	0	0	0	0
404	0	1	0	0	0	0	0	0	0
405	0	1	2	0	0	0	0	0	0
406	0	1	0	0	0	0	0	0	0
407	0	1	0	0	0	0	0	0	0
408	0	1	0	0	0	0	0	0	0
409	0	1	0	0	0	0	0	0	0
410	0	1	0	0	0	0	0	0	0
411	0	1	0	0	0	0	0	0	0
412	0	1	0	0	0	0	0	0	0
413	0	1	0	0	1	2	2	2	1
414	0	1	0	0	0	0	0	0	0
415	0	1	0	0	0	0	0	0	0
416	0	1	0	0	0	0	0	0	0
417	0	1	0	0	0	0	0	0	0
418	0	1	1	0	0	0	0	0	0
419	0	1	0	0	0	0	0	0	0
420	0	1	0	0	0	0	0	0	0
421	0	1	0	0	0	0	0	0	0
422	0	1	0	0	0	0	0	0	0
423	0	1	0	0	0	0	0	0	0
424	0	1	0	0	0	0	0	0	1
425	0	1	0	0	0	0	0	0	0
426	0	1	0	0	0	0	0	0	0
427	0	1	0	0	0	0	0	0	0
428	0	1	0	0	0	0	0	0	0
429	0	1	0	0	0	0	0	0	0
430	0	1	0	0	0	0	0	0	0
431	0	1	0	0	0	0	0	0	0
432	0	1	0	0	0	0	0	0	0
433	0	1	0	0	0	0	0	0	0
434	0	1	0	0	0	0	0	0	0
435	0	1	0	0	0	0	0	0	0
436	0	2	0	0	0	0	0	0	0
437	0	1	2	0	0	0	0	0	0
438	0	1	0	0	0	0	0	0	0
439	0	1	0	0	0	0	0	0	0
440	0	1	0	0	0	0	0	0	0
441	0	1	0	0	0	0	0	0	0
442	0	2	0	0	0	0	0	0	0

Appendix A (continued)

443	0	1	0	0	0	0	0	0	0
444	0	1	0	0	0	0	0	0	0
445	0	1	0	0	0	0	0	0	0
446	0	1	0	0	0	0	0	0	0
447	0	1	0	0	0	0	0	0	0
448	0	1	0	0	0	0	0	0	0
449	0	1	0	0	0	0	0	0	0
450	0	1	0	0	0	0	0	0	0
451	0	1	0	0	0	0	0	0	0
452	0	1	0	0	0	0	0	0	0
453	0	1	0	0	0	0	0	0	0
454	0	1	0	0	0	0	0	0	0
455	0	0	1	0	0	0	0	0	0
456	0	0	1	0	0	0	0	0	0
457	0	0	1	0	0	0	0	0	0
458	0	0	1	0	0	0	0	0	0
459	0	0	3	0	0	0	0	0	0
460	0	0	1	0	0	0	0	0	0
461	0	0	1	0	0	0	0	0	0
462	0	0	1	0	0	0	0	0	0
463	0	0	1	0	0	0	0	0	0
464	0	0	1	0	0	0	0	0	0
465	0	0	2	0	0	0	0	0	0
466	0	0	1	0	0	0	0	0	0
467	0	0	1	0	0	0	0	0	0
468	0	0	1	0	0	0	0	0	0
469	0	0	1	0	0	0	0	0	0
470	0	0	2	0	0	0	0	0	0
471	0	0	1	0	0	0	0	0	0
472	0	0	1	0	0	0	0	0	0
473	0	0	1	0	0	0	0	0	0
474	0	0	1	0	0	0	0	0	0
475	0	0	1	0	0	0	0	0	0
476	0	0	1	0	0	0	0	0	0
477	0	0	2	0	0	0	0	0	0
478	0	0	2	0	0	0	0	0	0
479	0	0	1	0	0	0	0	0	0
480	0	0	1	0	0	0	0	0	0
481	0	0	1	0	0	1	0	0	0
482	0	0	1	0	0	0	0	0	0
483	0	0	1	0	0	0	0	0	0
484	0	0	1	0	0	0	0	0	0
485	0	0	3	0	0	0	0	0	0
486	0	0	2	0	0	0	0	0	0
487	0	0	1	0	0	1	0	0	0
488	0	0	7	1	0	1	0	0	0
489	0	0	1	0	0	0	0	0	1
490	0	0	1	0	0	0	0	0	0
491	0	0	1	0	0	0	0	0	0

Appendix A (continued)

492	0	0	1	0	0	0	0	0	0
493	0	0	1	0	0	0	0	0	0
494	0	0	1	0	0	0	0	0	0
495	0	0	1	0	0	0	0	0	0
496	0	0	1	0	0	1	2	1	0
497	0	0	1	0	0	0	0	0	0
498	0	0	2	0	0	0	0	0	0
499	0	0	2	0	0	0	0	0	0
500	0	0	4	0	0	0	0	1	0
501	0	0	1	0	0	0	0	0	1
502	0	0	1	0	0	0	0	0	0
503	0	0	3	0	1	0	2	0	0
504	0	0	10	2	2	0	0	2	0
505	0	0	1	0	0	1	0	0	0
506	0	0	1	0	0	0	0	0	0
507	0	0	3	0	0	0	0	0	0
508	0	0	3	0	0	0	0	0	0
509	0	0	2	0	0	0	0	0	0
510	0	0	1	0	0	0	0	0	0
511	0	0	1	0	0	0	0	0	0
512	0	0	6	0	0	0	0	1	0
513	0	0	4	0	0	0	0	0	2
514	0	0	1	0	0	0	0	0	0
515	0	0	1	0	0	0	0	0	0
516	0	0	1	0	0	0	0	0	0
517	0	0	1	0	0	0	0	0	0
518	0	0	1	0	0	0	0	0	0
519	0	0	1	0	0	0	0	0	0
520	0	0	1	0	0	1	0	0	0
521	0	0	1	0	0	0	0	0	0
522	0	0	2	0	0	0	0	0	0
523	0	0	2	0	0	0	0	0	0
524	0	0	1	0	0	0	0	0	0
525	0	0	1	0	0	0	0	0	0
526	0	0	1	0	0	0	0	0	0
527	0	0	1	0	0	0	0	0	0
528	0	0	2	0	0	0	0	0	0
529	0	0	1	0	0	0	0	2	0
530	0	0	1	0	0	0	0	0	0
531	0	0	1	0	0	0	0	0	0
532	0	0	2	0	0	0	0	0	0
533	0	0	1	0	0	0	0	0	0
534	0	0	1	0	0	0	0	0	0
535	0	0	1	0	0	0	0	0	0
536	0	0	2	0	0	0	0	0	0
537	0	0	1	0	0	0	0	0	0
538	0	0	2	0	0	0	0	0	0
539	0	0	1	0	0	0	0	0	0
540	0	0	2	0	0	1	0	0	0
541	0	0	1	0	0	0	0	0	0

Appendix A (continued)

542	0	0	1	0	0	0	0	0	0
543	0	0	1	0	0	0	0	0	0
544	0	0	1	0	0	0	0	0	0
545	0	0	1	0	0	0	0	0	0
546	0	0	1	0	0	0	0	0	0
547	0	0	1	0	0	0	0	0	0
548	0	0	1	0	0	0	0	0	0
549	0	0	1	0	0	0	0	0	0
550	0	0	3	0	0	0	0	0	0
551	0	0	1	0	0	0	0	0	0
552	0	0	1	0	0	0	0	0	0
553	0	0	1	0	0	0	0	0	0
554	0	0	1	0	0	0	0	0	0
555	0	0	1	0	0	0	0	0	0
556	0	0	1	0	0	0	0	0	0
557	0	0	1	0	0	0	0	0	0
558	0	0	1	0	0	0	0	0	0
559	0	0	2	0	0	0	0	0	0
560	0	0	1	0	0	0	0	0	0
561	0	0	1	0	0	0	0	0	0
562	0	0	1	0	0	0	0	0	0
563	0	0	3	0	0	0	0	1	0
564	0	0	1	0	0	0	0	0	0
565	0	0	1	0	0	0	0	0	0
566	0	0	1	0	0	0	0	0	0
567	0	0	1	0	0	0	0	0	0
568	0	0	1	0	0	0	0	0	0
569	0	0	1	0	0	0	0	0	0
570	0	0	1	0	0	0	0	0	0
571	0	0	1	0	0	0	0	0	0
572	0	0	1	0	0	0	0	0	0
573	0	0	1	0	0	0	0	0	0
574	0	0	2	0	0	0	0	0	0
575	0	0	1	0	0	2	0	0	0
576	0	0	1	0	0	0	0	0	0
577	0	0	1	0	0	0	0	0	0
578	0	0	1	0	0	0	0	0	0
579	0	0	1	0	0	0	0	0	0
580	0	0	1	0	0	0	0	0	0
581	0	0	1	0	0	0	0	0	0
582	0	0	1	0	0	0	0	0	0
583	0	0	1	0	0	0	0	0	0
584	0	0	1	0	0	0	0	0	0
585	0	0	1	0	0	0	0	0	0
586	0	0	1	0	0	0	0	0	0
587	0	0	1	0	0	0	0	0	0
588	0	0	1	0	0	0	0	0	0
589	0	0	1	0	0	0	0	0	0
590	0	0	1	0	0	0	0	0	0
591	0	0	1	0	0	0	0	0	0

Appendix A (continued)

592	0	0	3	0	0	0	0	0	0
593	0	0	3	0	0	0	0	0	0
594	0	0	1	0	0	0	0	0	0
595	0	0	2	0	0	0	0	0	0
596	0	0	1	0	1	0	2	0	0
597	0	0	1	0	0	0	0	0	0
598	0	0	2	0	0	0	0	0	1
599	0	0	1	0	0	0	0	0	0
600	0	0	1	0	0	0	0	0	0
601	0	0	1	0	0	0	0	0	0
602	0	0	2	0	0	0	0	0	0
603	0	0	1	0	0	0	0	1	0
604	0	0	1	0	0	0	0	0	0
605	0	0	1	0	0	0	0	0	0
606	0	0	3	0	0	0	0	0	0
607	0	0	1	0	0	0	0	0	0
608	0	0	1	0	0	0	0	0	0
609	0	0	1	0	0	0	0	0	0
610	0	0	1	0	0	0	0	0	0
611	0	0	1	0	1	0	5	0	0
612	0	0	1	0	0	0	0	0	0
613	0	0	1	0	0	0	0	0	0
614	0	0	1	0	0	0	0	0	0
615	0	0	1	0	0	0	0	0	0
616	0	0	1	0	0	0	0	0	0
617	0	0	1	0	0	0	0	0	0
618	0	0	1	0	0	0	0	0	0
619	0	0	1	0	0	0	0	0	0
620	0	0	2	0	0	0	0	0	0
621	0	0	1	0	0	0	0	0	0
622	0	0	1	0	0	0	0	0	0
623	0	0	1	0	0	0	0	0	0
624	0	0	1	0	0	0	0	0	0
625	0	0	1	0	0	0	0	0	0
626	0	0	1	0	0	0	0	0	0
627	0	0	1	0	0	0	0	0	0
628	0	0	1	0	0	0	0	0	0
629	0	0	1	0	0	0	0	0	0
630	0	0	1	0	0	0	0	0	0
631	0	0	1	0	0	0	0	0	0
632	0	0	1	0	0	0	0	0	0
633	0	0	1	0	0	0	0	0	0
634	0	0	1	0	0	0	0	0	0
635	0	0	1	0	0	0	0	0	0
636	0	0	2	0	0	2	0	0	0
637	0	0	1	0	0	0	0	0	0
638	0	0	1	0	0	0	0	0	0
639	0	0	1	0	0	0	0	0	0
640	0	0	1	0	0	0	0	0	0
641	0	0	1	0	0	0	0	0	0

Appendix A (continued)

642	0	0	1	0	0	0	0	0	0
643	0	0	1	0	0	0	0	0	0
644	0	0	1	0	0	0	0	0	0
645	0	0	1	0	0	0	0	0	0
646	0	0	2	0	0	0	0	0	0
647	0	0	1	0	0	0	0	0	0
648	0	0	1	0	0	0	0	2	0
649	0	0	1	0	0	0	0	0	0
650	0	0	1	0	0	0	0	0	0
651	0	0	1	0	0	0	0	0	0
652	0	0	1	0	0	0	0	0	0
653	0	0	1	0	0	0	0	0	0
654	0	0	1	0	0	0	0	0	0
655	0	0	1	0	0	0	0	0	0
656	0	0	0	1	0	0	0	0	0
657	0	0	0	1	0	0	5	1	0
658	0	0	0	1	0	0	0	0	0
659	0	0	0	1	0	0	0	0	0
660	0	0	0	1	0	0	0	0	0
661	0	0	0	1	0	0	0	0	0
662	0	0	0	2	0	0	0	0	0
663	0	0	0	1	0	0	3	1	0
664	0	0	0	1	0	0	0	0	0
665	0	0	0	1	0	1	1	0	2
666	0	0	0	1	0	0	1	0	0
667	0	0	0	1	0	0	0	0	0
668	0	0	0	1	0	0	0	0	0
669	0	0	0	1	0	0	0	0	0
670	0	0	0	1	0	0	0	0	0
671	0	0	0	1	0	0	0	0	0
672	0	0	0	2	0	0	0	0	0
673	0	0	0	1	0	0	0	0	0
674	0	0	0	1	0	0	0	0	0
675	0	0	1	1	0	0	0	0	0
676	0	0	0	1	0	0	0	0	0
677	0	0	0	1	0	2	1	1	1
678	0	0	0	1	0	0	0	0	0
679	0	0	0	1	0	0	1	0	0
680	0	0	0	1	0	0	0	0	0
681	0	0	0	1	0	0	0	0	1
682	0	0	0	1	0	0	0	0	0
683	0	0	0	1	0	0	0	0	0
684	0	0	0	1	0	0	0	0	0
685	0	0	0	1	0	0	0	0	0
686	0	0	0	3	0	0	0	0	0
687	0	0	0	1	0	0	0	0	0
688	0	0	0	1	0	0	0	0	0
689	0	0	0	1	0	0	0	0	0
690	0	0	0	1	0	0	0	0	0
691	0	0	0	1	0	0	0	0	0

Appendix A (continued)

692	0	0	0	1	0	0	0	0	0
693	0	0	0	1	0	0	0	1	0
694	0	0	0	2	0	0	0	0	0
695	0	0	0	0	1	0	0	0	0
696	0	0	0	0	1	0	0	0	0
697	0	0	0	0	1	0	0	0	0
698	0	0	0	0	1	0	0	0	0
699	0	0	0	0	1	0	0	0	0
700	0	0	0	0	1	0	0	0	0
701	0	0	0	0	1	0	0	0	0
702	0	0	0	0	1	0	0	0	0
703	0	0	0	0	2	0	0	0	0
704	0	0	7	0	1	0	0	0	0
705	0	0	2	0	1	0	0	0	0
706	0	0	0	0	1	0	0	0	0
707	0	0	0	0	1	0	0	0	0
708	0	0	0	0	1	0	0	0	0
709	0	0	0	0	1	0	0	0	0
710	0	0	0	0	1	0	0	0	0
711	0	0	0	0	1	0	0	0	0
712	0	0	0	0	1	0	0	0	0
713	0	0	0	0	1	0	0	0	0
714	0	0	0	0	1	0	0	0	0
715	0	0	0	0	1	0	0	0	0
716	0	0	0	0	1	0	0	0	0
717	0	0	0	0	1	0	0	0	0
718	0	0	0	0	1	1	1	0	1
719	0	0	0	0	1	0	0	0	0
720	0	0	0	0	1	0	0	0	0
721	0	0	0	0	1	0	0	0	0
722	0	0	0	0	1	0	0	0	0
723	0	0	0	0	1	0	0	0	0
724	0	0	0	0	1	0	0	0	0
725	0	0	0	0	1	0	0	0	0
726	0	0	0	0	1	0	0	0	0
727	0	0	0	0	1	0	0	0	0
728	0	0	0	0	1	0	0	0	0
729	0	0	0	0	1	0	0	0	0
730	0	0	0	0	1	0	0	0	0
731	0	0	0	0	1	0	0	0	0
732	0	0	0	0	1	0	0	0	0
733	0	0	0	0	2	0	0	0	0
734	0	0	0	0	1	0	0	0	0
735	0	0	0	0	1	0	0	0	0
736	0	0	0	0	1	0	0	0	0
737	0	0	0	0	1	0	0	0	0
738	0	0	0	0	1	0	0	0	0
739	0	0	0	0	1	0	0	0	0
740	0	0	0	0	1	0	0	0	0
741	0	0	0	0	0	5	0	1	1

Appendix A (continued)

742	0	0	0	0	0	1	0	0	0
743	0	0	0	0	0	1	0	0	0
744	0	0	0	0	0	1	0	0	0
745	0	0	0	0	0	1	0	0	0
746	0	0	0	0	0	1	0	0	0
747	0	0	0	0	0	1	0	0	0
748	0	0	0	0	0	1	0	0	0
749	0	0	0	0	0	1	0	0	0
750	0	0	0	0	0	1	0	0	0
751	0	0	0	0	0	1	0	1	1
752	0	0	0	0	0	1	0	0	0
753	0	0	0	0	0	1	0	0	0
754	0	0	0	0	0	1	0	0	1
755	0	0	0	0	0	1	0	0	0
756	0	0	0	0	0	1	0	0	0
757	0	0	0	0	0	1	0	0	0
758	0	0	0	0	0	1	0	0	0
759	0	0	0	0	0	1	0	0	0
760	0	0	0	0	0	3	0	0	0
761	0	0	0	0	0	1	0	0	0
762	0	0	0	0	0	1	0	0	1
763	0	0	0	0	0	1	0	0	0
764	0	0	0	0	0	1	0	0	0
765	0	0	0	0	0	1	0	0	0
766	0	0	0	0	0	1	0	0	0
767	0	0	0	0	0	1	0	0	0
768	0	0	0	0	0	1	0	0	0
769	0	0	0	0	0	1	0	0	0
770	0	0	0	0	0	1	0	0	0
771	0	0	0	0	0	1	0	0	0
772	0	0	0	0	0	3	0	0	0
773	0	0	0	0	0	3	0	0	0
774	0	0	0	0	0	1	0	0	0
775	0	0	0	0	0	1	0	0	0
776	0	0	0	0	0	1	0	0	0
777	0	0	0	0	0	3	0	0	0
778	0	0	0	0	0	1	0	0	0
779	0	0	0	0	0	1	0	0	0
780	0	0	0	0	0	1	0	0	0
781	0	0	0	0	0	1	0	0	0
782	0	0	0	0	0	1	0	0	0
783	0	0	0	0	0	1	0	0	0
784	0	0	0	0	0	1	0	0	0
785	0	0	0	0	0	1	0	0	0
786	0	0	0	0	0	1	0	0	0
787	0	0	0	0	0	2	0	0	0
788	0	0	0	0	0	1	0	0	0
789	0	0	0	0	0	1	0	0	0
790	0	0	0	0	0	1	0	0	0
791	0	0	0	0	0	1	0	0	0

Appendix A (continued)

792	0	0	0	0	0	1	0	0	0
793	0	0	0	0	0	1	0	0	0
794	0	0	0	0	0	3	0	0	3
795	0	0	0	0	0	1	0	0	0
796	0	0	0	0	0	1	0	0	0
797	0	0	0	0	0	3	0	0	0
798	0	0	0	0	0	1	0	0	0
799	0	0	0	0	0	2	0	1	0
800	0	0	0	0	0	1	0	0	0
801	0	0	0	0	0	1	0	0	0
802	0	0	0	0	0	1	0	0	0
803	0	0	0	0	0	2	0	0	0
804	0	0	0	0	0	1	0	0	0
805	0	0	0	0	0	1	0	0	0
806	0	0	0	0	0	2	0	0	0
807	0	0	0	0	0	2	0	0	0
808	0	0	0	0	0	1	0	0	0
809	0	0	0	0	0	1	0	0	0
810	0	0	0	0	0	1	0	0	0
811	0	0	0	0	0	1	0	0	0
812	0	0	0	0	0	1	1	0	0
813	0	0	0	0	0	1	0	0	0
814	0	0	0	0	0	1	0	0	0
815	0	0	0	0	0	1	0	0	0
816	0	0	0	0	0	1	0	0	0
817	0	0	0	0	0	2	0	0	0
818	0	0	0	0	0	1	0	0	0
819	0	0	0	0	0	1	0	0	0
820	0	0	0	0	0	1	0	0	0
821	0	0	0	0	0	1	0	0	0
822	0	0	0	0	0	2	0	0	0
823	0	0	0	0	0	1	0	0	0
824	0	0	0	0	0	1	0	0	0
825	0	0	0	0	0	1	0	0	0
826	0	0	0	0	0	1	0	0	0
827	0	0	0	0	0	1	0	0	0
828	0	0	0	0	0	1	0	0	0
829	0	0	0	0	0	1	0	0	0
830	0	0	0	0	0	1	0	0	0
831	0	0	0	0	0	2	0	0	0
832	0	0	7	0	0	1	0	0	0
833	0	0	0	0	0	1	0	0	0
834	0	0	0	0	0	4	0	0	0
835	0	0	0	0	0	2	0	0	0
836	0	0	0	0	0	1	0	0	0
837	0	0	0	0	0	1	0	0	0
838	0	0	0	0	0	2	0	0	0
839	0	0	0	0	0	2	0	0	0
840	0	0	0	0	0	1	0	0	0
841	0	0	0	0	0	1	0	0	0

Appendix A (continued)

842	0	0	0	0	0	1	0	0	0
843	0	0	0	0	0	1	0	0	0
844	0	0	0	0	0	0	1	0	0
845	0	0	0	0	0	0	1	0	0
846	0	0	0	0	0	0	1	0	0
847	0	0	0	0	0	0	1	0	0
848	0	0	0	0	0	0	4	0	0
849	0	0	0	0	0	0	1	0	0
850	0	0	0	0	0	0	1	0	0
851	0	0	0	0	0	0	2	0	0
852	0	0	0	0	0	0	1	0	0
853	0	0	0	0	0	0	1	0	0
854	0	0	0	0	0	0	1	0	0
855	0	0	0	0	0	0	1	0	0
856	0	0	0	0	0	0	1	0	0
857	0	0	0	0	0	0	1	0	0
858	0	0	0	0	0	0	1	0	0
859	0	0	0	0	0	0	1	0	0
860	0	0	0	0	0	0	1	0	0
861	0	0	0	0	0	0	1	0	0
862	0	0	0	0	0	0	1	0	0
863	0	0	0	0	0	0	2	0	0
864	0	0	0	0	0	0	1	0	0
865	0	0	0	0	0	0	1	0	0
866	0	0	0	0	0	0	1	0	0
867	0	0	0	0	0	0	1	0	0
868	0	0	0	0	0	0	1	0	0
869	0	0	0	0	0	0	4	0	0
870	0	0	0	0	0	0	1	0	0
871	0	0	0	0	0	0	1	0	0
872	0	0	0	0	0	0	1	0	0
873	0	0	0	0	0	0	1	0	0
874	0	0	0	0	0	0	1	0	0
875	0	0	0	0	0	0	1	0	0
876	0	0	0	0	0	0	1	0	0
877	0	0	0	0	0	0	1	0	0
878	0	0	0	0	0	0	1	0	0
879	0	0	0	0	0	0	1	0	0
880	0	0	0	0	0	0	1	0	0
881	0	0	0	0	0	0	1	0	0
882	0	0	0	0	0	0	1	0	0
883	0	0	0	0	0	0	1	0	0
884	0	0	0	0	0	0	1	0	0
885	0	0	0	0	0	0	1	0	0
886	0	0	0	0	0	0	1	0	0
887	0	0	0	0	0	0	1	0	0
888	0	0	0	0	0	0	1	0	0
889	0	0	0	0	0	0	1	0	0
890	0	0	0	0	0	0	1	0	0
891	0	0	0	0	0	0	1	0	0

Appendix A (continued)

892	0	0	0	0	0	0	1	0	0
893	0	0	0	0	0	0	1	0	0
894	0	0	0	0	0	0	1	0	0
895	0	0	0	0	0	0	1	0	0
896	0	0	0	0	0	0	1	0	0
897	0	0	0	0	0	0	1	0	0
898	0	0	0	0	0	0	1	0	0
899	0	0	0	0	0	0	1	0	0
900	0	0	0	0	0	0	1	0	0
901	0	0	0	0	0	0	1	0	0
902	0	0	0	0	0	0	1	0	0
903	0	0	0	0	0	0	1	0	0
904	0	0	0	0	0	0	1	0	0
905	0	0	0	0	0	0	1	0	0
906	0	0	0	0	0	0	1	0	0
907	0	0	0	0	0	0	1	0	0
908	0	0	0	0	0	0	1	0	0
909	0	0	0	0	0	0	1	0	0
910	0	0	0	0	0	0	1	0	0
911	0	0	0	0	0	0	1	0	0
912	0	0	0	0	0	0	1	0	0
913	0	0	0	0	0	0	1	0	0
914	0	0	0	0	0	0	1	0	0
915	0	0	0	0	0	0	1	0	0
916	0	0	0	0	0	0	1	0	0
917	0	0	0	0	0	0	1	0	0
918	0	0	0	0	0	0	1	0	0
919	0	0	0	0	0	0	1	0	0
920	0	0	0	0	0	0	1	0	0
921	0	0	0	0	0	0	1	0	0
922	0	0	0	0	0	0	0	1	0
923	0	0	0	0	0	0	0	1	0
924	0	0	0	0	0	0	0	1	0
925	0	0	0	0	0	0	0	1	0
926	0	0	0	0	0	0	0	1	0
927	0	0	0	0	0	0	0	1	0
928	0	0	0	0	0	0	0	1	0
929	0	0	0	0	0	0	0	1	0
930	0	0	0	0	0	0	0	1	0
931	0	0	0	0	0	0	0	1	0
932	0	0	0	0	0	0	0	1	0
933	0	0	0	0	0	0	0	1	0
934	0	0	0	0	0	0	0	1	0
935	0	0	0	0	0	0	0	1	0
936	0	0	0	0	0	0	0	1	0
937	0	0	0	0	0	0	0	1	0
938	0	0	0	0	0	0	0	1	0
939	0	0	0	0	0	0	0	1	0
940	0	0	0	0	0	0	0	1	0
941	0	0	0	0	0	0	0	1	0

Appendix A (continued)

942	0	0	0	0	0	0	0	1	0
943	0	0	0	0	0	0	0	1	1
944	0	0	0	0	0	0	0	1	0
945	0	0	0	0	0	0	0	1	0
946	0	0	0	0	0	0	0	1	0
947	0	0	7	0	0	0	0	1	0
948	0	0	0	0	0	0	0	1	0
949	0	0	0	0	0	0	0	1	0
950	0	0	0	0	0	0	0	1	0
951	0	0	0	0	0	0	0	1	0
952	0	0	0	0	0	0	0	1	0
953	0	0	0	0	0	0	0	1	0
954	0	0	0	0	0	0	0	1	0
955	0	0	0	0	0	0	0	1	0
956	0	0	0	0	0	0	0	1	0
957	0	0	0	0	0	0	0	1	0
958	0	0	0	0	0	0	0	1	0
959	0	0	0	0	0	0	0	1	0
960	0	0	0	0	0	0	0	1	0
961	0	0	0	0	0	0	0	1	0
962	0	0	0	0	0	0	0	1	0
963	0	0	2	0	0	0	0	1	0
964	0	0	0	0	0	0	0	1	0
965	0	0	0	0	0	0	0	1	0
966	0	0	0	0	0	0	0	1	0
967	0	0	0	0	0	0	0	1	0
968	0	0	0	0	0	0	0	1	0
969	0	0	0	0	0	0	0	1	0
970	0	0	0	0	0	0	0	1	0
971	0	0	0	0	0	0	0	1	0
972	0	0	0	0	0	0	0	1	0
973	0	0	0	0	0	0	0	1	0
974	0	0	0	0	0	0	0	1	0
975	0	0	0	0	0	0	0	2	0
976	0	0	0	0	0	0	0	2	0
977	0	0	0	0	0	0	0	0	1
978	0	0	0	0	0	0	0	0	1
979	0	0	0	0	0	0	0	0	1
980	0	0	0	0	0	0	0	0	1
981	0	0	0	0	0	0	0	0	1
982	0	0	0	0	0	0	0	0	1
983	0	0	0	0	0	0	0	0	1
984	0	0	0	0	0	0	0	0	2
985	0	0	0	0	0	0	0	0	1
986	0	0	0	0	0	0	0	0	1
987	0	0	0	0	0	0	0	0	1
988	0	0	0	0	0	0	0	0	1
989	0	0	0	0	0	0	0	0	1
990	0	0	0	0	0	0	0	0	1
991	0	0	0	0	0	0	0	0	1

Appendix A (continued)

992	0	0	0	0	0	0	0	0	1
993	0	0	0	0	0	0	0	0	1
994	0	0	0	0	0	0	0	0	1
995	0	0	0	0	0	0	0	0	1
996	0	0	0	0	0	0	0	0	1
997	0	0	0	0	0	0	0	0	1
998	0	0	0	0	0	0	0	0	2
999	0	0	0	0	0	0	0	0	1
1000	0	0	0	0	0	0	0	0	1
1001	0	0	0	0	0	0	0	0	1
1002	0	0	0	0	0	0	0	0	1
1003	0	0	0	0	0	0	0	0	2
1004	0	0	0	0	0	0	0	0	1
1005	0	0	0	0	0	0	0	0	1
1006	0	0	0	0	0	0	0	0	1
1007	0	0	0	0	0	0	0	0	1
1008	0	0	0	0	0	0	0	0	1
1009	0	0	0	0	0	0	0	0	1
1010	0	0	0	0	0	0	0	0	1
1011	0	0	0	0	0	0	0	0	1
1012	0	0	0	0	0	0	0	0	1
1013	0	0	0	0	0	0	0	0	1
1014	0	0	7	0	0	0	0	0	1
1015	0	0	0	0	0	0	0	0	1
1016	0	0	0	0	0	0	0	0	1
1017	0	0	0	0	0	0	0	0	1
1018	0	0	0	0	0	0	0	0	1
1019	0	0	0	0	0	0	0	0	1
1020	0	0	0	0	0	0	0	0	1
1021	0	0	0	0	0	0	0	0	1
1022	0	0	0	0	0	0	0	0	1
1023	0	0	0	0	0	0	0	0	1
1024	0	0	0	0	0	0	0	0	1
1025	0	0	0	0	0	0	0	0	1
1026	0	0	1	0	0	0	0	0	1
1027	0	0	0	0	0	0	0	0	1
1028	0	0	0	0	0	0	0	0	1
1029	0	0	0	0	0	0	0	0	1
1030	0	0	0	0	0	0	0	0	1
1031	0	0	0	0	0	0	0	0	1
1032	0	0	0	0	0	0	0	0	1
1033	0	0	0	0	0	0	0	0	1
Total	**290**	**453**	**676**	**102**	**87**	**233**	**213**	**118**	**116**

APPENDIX B: Principle Component Analysis for mtDNA

Weights of first three principle components of the variables for the mtDNA HVRI haplogroup frequencies.

	PC1	PC2	PC3
CRS	0,18	0,71	0,50
A	-0,27	-0,13	0,00
B	-0,22	0,20	-0,19
B5	-0,51	-0,41	-0,22
C	-0,76	-0,03	-0,12
D	-0,81	-0,36	-0,14
E	-0,63	-0,09	-0,33
I	0,10	0,52	0,35
W	0,24	-0,07	0,11
M	-0,48	0,12	0,40
X	0,66	0,14	-0,42
K	0,71	-0,08	0,04
F	-0,11	0,06	-0,14
V	-0,05	0,09	-0,60
R1	0,01	0,24	-0,12
PREHV	0,17	-0,77	-0,08
J	0,60	0,15	-0,03
J1	0,49	-0,44	-0,01
J2	-0,27	-0,17	0,08
T	0,45	-0,01	-0,34
T1	0,68	-0,01	-0,20
T2	0,15	-0,15	0,31
T3	0,17	-0,32	0,27
T4	0,27	0,10	-0,53
T5	-0,18	0,20	-0,07
U1	-0,06	-0,62	0,48
U2	-0,06	-0,28	0,12
U3	0,36	-0,22	0,43
U4	0,34	0,40	-0,26
U5	0,25	0,68	0,25
U7	0,33	-0,12	-0,39
L1	0,43	-0,67	0,06
L3	0,57	-0,61	0,10

APPENDIX C: Principle Component Analysis for Y-chromosome

Weights of first three principle components of the variables for the Y-chromosome haplogroup frequencies.

	PC1	PC2	PC3
A	0,02	0,05	0,75
C	0,64	-0,51	-0,10
D	0,71	-0,33	-0,23
E	-0,41	0,57	0,19
G	-0,52	-0,58	-0,07
H	0,39	0,03	0,22
I	-0,22	-0,03	-0,15
J	-0,54	0,12	0,51
L	0,26	0,23	0,17
N	0,20	-0,23	0,64
O	0,68	-0,27	0,16
Q	0,35	-0,01	0,60
F	-0,56	-0,45	-0,20
K	-0,11	-0,80	-0,01
P	0,38	0,38	-0,43
R	0,73	0,33	-0,04
Y	-0,01	0,62	-0,24

APPENDIX D: Principle Component Analysis for Alu insertion Ploymorphisms

Weights of first three principle components of the variables for the Alu insertion polymorphisms frequencies.

	PC1	PC2	PC3
A25	0,55	0,20	0,71
B65	-0,14	-0,79	0,52
ACE	0,96	0,06	-0,14
APO	-0,18	-0,67	-0,23
PV9	0,79	-0,08	0,07
TPA25	0,63	-0,62	-0,22
FXIIIB	0,96	0,06	-0,14

APPENDIX E: Principle Component Analysis for Autosomal Microsatellites

Weights of first three principle components of the variables for the autosomal microsatellite frequencies.

Allele	PC1	PC2	PC3
TH01-5	0,45	0,42	0,46
THO1-6	0,79	-0,38	-0,40
THO1-7	-0,88	-0,39	-0,28
THO1-8	0,34	0,24	0,14
THO1-9	-0,81	-0,37	-0,28
THO1-9.3	0,31	0,20	0,32
THO1-10	-0,73	-0,31	-0,25
THO1-11	0,12	0,31	0,26
THO1-12	0,08	0,43	0,36
TPOX-6	0,55	0,28	-0,77
TPOX-7	0,71	-0,63	-0,07
TPOX-8	-0,54	-0,64	0,47
TPOX-9	0,97	-0,28	-0,22
TPOX-10	0,82	-0,06	-0,65
TPOX-11	-0,01	0,94	-0,23
TPOX-12	-0,97	0,11	0,36
TPOX-13	0,69	-0,78	-0,16
D13S-7	0,46	-0,09	0,08
D13S-8	-0,62	0,65	-0,53
D13S-9	-0,94	0,45	-0,04
D13S-10	-1,05	-0,07	-0,03
D13S-11	0,62	0,70	0,22
D13S-12	0,81	-0,55	-0,09
D13S-13	0,67	-0,41	0,61
D13S-14	0,75	-0,35	0,66
D13S-15	0,19	-0,74	-0,33
D5S-7	0,82	-0,02	0,62
D5S-8	-0,15	-0,96	0,24
D5S-9	-0,15	-0,91	0,46
D5S-10	-0,90	-0,48	0,08
D5S-11	-0,93	-0,12	0,45
D5S-12	0,58	0,65	-0,29
D5S-13	0,70	0,38	-0,57
D5S-14	0,76	0,32	-0,01
D5S-15	0,56	-0,44	-0,07
D5S-16	-0,92	-0,17	0,35
D3S=12	0,48	-0,62	-0,26
D3S=13	0,52	-0,13	0,35
D3S=14	0,64	0,41	0,43

Appendix E (continued)

D3S=15	-0,78	-0,44	0,34
D3S=16	-0,53	-0,05	-0,81
D3S=17	0,62	-0,04	-0,41
D3S=18	0,69	0,52	-0,33
D3S=19	0,69	-0,22	0,67
D3S=20	0,69	-0,23	0,32
D7S=6	0,19	-0,74	-0,33
D7S=7	0,02	0,16	-0,96
D7S=8	-0,65	0,46	-0,72
D7S=9	0,91	-0,04	0,49
D7S=91	0,42	0,24	0,64
D7S=10	0,91	-0,09	0,52
D7S=11	0,76	-0,24	-0,57
D7S=12	-0,92	-0,37	0,35
D7S=13	0,62	-0,53	-0,50
D7S=14	0,90	-0,34	0,39
D7S=15	0,19	-0,74	-0,33
VWA-11	-0,07	0,51	-0,82
VWA-12	0,48	-0,57	-0,23
VWA-13	0,37	-0,75	-0,22
VWA-14	-0,50	0,63	0,58
VWA-15	0,90	0,00	-0,27
VWA-16	0,09	-0,78	-0,47
VWA-17	-0,68	0,07	0,55
VWA-18	-0,76	0,02	0,56
VWA-19	0,66	-0,15	-0,01
VWA-20	0,25	0,72	-0,57
VWA-21	0,16	-0,77	-0,40
VWA-22	0,16	-0,77	-0,40
D8S-7	0,16	-0,77	-0,40
D8S-8	0,28	-0,34	0,24
D8S-9	0,56	0,65	-0,45
D8S-10	-0,64	-0,44	0,29
D8S-11	0,58	0,37	0,54
D8S-12	0,62	0,32	0,53
D8S-13	0,11	0,64	-0,09
D8S-14	0,04	-0,68	0,41
D8S-15	-0,33	-0,05	-0,93
D8S-16	-0,42	-0,62	-0,36
D8S-17	-0,78	-0,49	0,25
D8S-18	0,69	-0,53	0,11
D2S13-15	0,10	0,32	-1,00
D2S13-16	0,56	0,52	-0,67
D2S13-17	0,70	0,55	-0,39
D2S13-18	-0,80	-0,55	0,18
D2S13-19	-0,66	-0,62	0,35
D2S13-20	-0,34	-0,56	0,68
D2S13-21	0,49	-0,49	0,74

Appendix E (continued)

D2S13-22	-0,61	-0,80	-0,21
D2S13-23	-0,20	0,46	-0,90
D2S13-24	-0,14	1,02	0,27
D2S13-25	0,20	-0,09	0,81
D2S13-26	0,02	0,04	1,05
D2S13-27	0,53	-0,85	-0,22
D2S13-28	0,31	-0,89	-0,44
D18S-9	0,16	-0,77	-0,40
D18S-10	0,56	0,19	-0,44
D18S-102	0,16	-0,77	-0,40
D18S-11	0,36	-0,63	0,20
D18S-12	0,54	-0,48	0,56
D18S-13	0,63	0,66	-0,01
D18S-132	0,16	-0,77	-0,40
D18S-14	-0,67	0,02	-0,61
D18S-142	0,16	-0,77	-0,40
D18S-15	-0,84	0,23	-0,25
D18S-16	-0,11	0,61	-0,49
D18S-17	0,56	-0,14	0,79
D18S-18	0,82	-0,03	-0,10
D18S-19	-0,25	-0,29	0,02
D18S-20	-0,20	-0,02	0,58
D18S-21	-0,69	0,64	-0,01
D18S-22	0,69	0,55	-0,37
D18S-23	0,31	0,10	0,38
D18S-24	0,16	-0,77	-0,40
D18S-25	-0,09	0,64	-0,72
D2S11-242	0,16	-0,77	-0,40
D2S11-25	0,33	-0,83	-0,41
D2S11-26	0,60	0,12	-0,52
D2S11-27	0,83	0,37	0,11
D2S11-28	0,80	-0,02	-0,07
D2S11-282	-0,36	-0,29	0,18
D2S11-29	-0,65	0,37	-0,63
D2S11-292	0,38	0,24	-0,47
D2S11-30	-0,79	-0,33	0,34
D2S11-302	0,39	0,45	0,64
D2S11-31	0,67	-0,42	0,03
D2S11-312	0,44	0,34	-0,78
D2S11-32	0,56	-0,02	0,40
D2S11-322	-0,50	-0,38	0,46
D2S11-33	0,42	0,04	-0,55
D2S11-332	0,05	0,03	0,53
D2S11-34	0,16	-0,77	-0,40
D2S11-342	-0,20	-0,22	-0,05
D2S11-35	0,48	-0,63	-0,27
D2S11-352	-0,09	0,64	-0,72
D2S11-36	0,16	-0,77	-0,40

Appendix E (continued)

D2S11-38	0,16	-0,77	-0,40
FGA-16	0,36	0,16	0,48
FGA-17	0,16	-0,77	-0,40
FGA-18	-0,67	0,23	0,68
FGA-182	-0,89	-0,19	0,24
FGA-19	0,84	0,17	0,36
FGA-20	0,70	0,49	-0,20
FGA-202	-0,62	0,42	-0,61
FGA-21	0,62	-0,36	0,58
FGA-212	-0,06	-0,04	-0,76
FGA-22	0,53	0,78	0,19
FGA-222	0,37	0,44	0,77
FGA-223	0,40	-0,15	-0,01
FGA-23	-0,73	-0,28	-0,55
FGA-232	-0,29	0,48	0,38
FGA-24	-0,78	-0,34	0,33
FGA-242	0,60	0,12	0,63
FGA-25	-0,18	-0,73	-0,54
FGA-252	0,57	0,01	0,44
FGA-26	-0,46	-0,03	-0,76
FGA-27	0,20	-0,19	0,17
FGA-28	-0,01	0,48	-0,85
FGA-29	0,16	-0,77	-0,40

CURRICULUM VITAE

Personal Information

Surname, Name: BERKMAN CANER, Ceren
Nationality: Turkish (TC)
Date and Place of Birth: 16 April 1976, Ankara
Marital Status: Married
Phone: +90 312 427 04 02
e-mail: cerenberkman@yahoo.com

Education

B.Sc., METU Department of Biological Sciences, 1995-1999

M.Sc., METU Department of Biological Sciences, 1999-2001

PhD., METU Department of Biological Sciences, 2001 – 2006

Work experience

1999 – : Worked as a research assistant in Department of Biological Sciences.

Foreign Languages

English

Projects contributed

Göçlerin Anadolu genetik yapısı üzerine etkileri (project no: BAP- 2003-07-02-00-39, 2003 - 2005)

Göçlerin Anadolu'nun genetik yapısına etkilerinin incelenmesinde yeni yaklaşımlar METU-BAP-2004-07-04-02 (2004-2005)

Conservation of rainbow trout (*Oncorhynchus mykiss*) stocks in South Aegean and Western Mediterranean regions in Turkey based on molecular methods. Supported by TUBİTAK, (1998-2001).

Publications

Journal Papers

Togan, I., Soysal, M.I., Berkman, C.C. and Koban, E. 2005. Molecular markers in conservation of breeds. *Tekirdağ Ziraat Fakültesi Dergisi,* 2(1): 44-49. (in Turkish)

Book Chapter

Togan, I., Berkman, C., Koban, E. (2004). Korumada Moleküler İşaretler. Book chapter in Autochthonous Breeds of Domestic Animals In Turkey. Edited by Soysal, M.I. Tekirdağ. 271-277.

Technical Report

Togan, İ., Ergüven, A., Berkman, C., Koban, E., (2001). Conservation of rainbow trout (Oncorhynchus mykiss) stocks in South Aegean and Western Mediterranean regions in Turkey based on molecular methods. (TUBITAK, TOGTAG/TARP-1811).

Conference papers

Berkman, C.C., Togan, Inci (2006) Inferences about Past Population Processes: The Asian Contribution to Turkish Population With Respect to Balkans. Workshop on Networks in Computational Biology Ankara, Turkey. (Accepted).

Berkman, C.C., Togan, Inci (2005). Resemblance of Turkish human population to the populations of Balkans and Central Asia. International Symposium of Healt Informatics and Bioinformatics, Turkey'05. 35-41

Gökçek, Ç., Koban, E.,Berkman, C.C., Togan, Inci (2005). Mitochondrial DNA (mtDNA) Sequence Analysis of KAngal Dogs in Turkey. International Symposium of Healt Informatics and Bioinformatics, Turkey'05. 243.

Togan, I., Koban, E. and Berkman, C.C. (2005). Molecular markers in management of domestic animals 1st Molecular Genetic Applications Workshop Samsun, Turkey,

Togan, I., Berkman, C.C., Koban, E. (2004). Molecular Markers in Conservation. Workshop on Conservation of Native AnGR, Tekirdağ, Turkey.

Berkman, C., Koban, E., Togan, İ. (2002). Conservation of Rainbow trout (Oncorhynchus mykiss) stocks in Southern Aegean and Western Mediterranean Regions in Turkey based on molecular methods. XVI. National Biology Congress, Malatya, Turkey.

Özer,F., Plan, E., Koban, E., Berkman, C., Emre, Y., Ergüven, A., Togan, İ. (2002). Genetic diversity of brown trout (Salmo trutta L.) in Anatolia. International symposium on Phylogeography in Southern European Refugia: Evolutionary perspectives on the origins and conservation of European biodiversity, Vairão, Portugal, p 46.

Koban, E., Plan, E., Gezgin, F., Caner, C., Ergüven, A. and Togan, İ., 2000. Genetic relation among brown trout (Salmo trutta L.) populations of Anatolia and neighboring countries. 15th National Congress on Biology (with international participation), Ankara, Turkey, pp 33.