ALU INSERTION POLYMORPHISMS IN ANATOLIAN TURKS

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

BY

HAVVA DİNÇ

IN PARTIAL FULLFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
IN
THE DEPARTMENT OF BILOGICAL SCIENCES

SEPTEMBER 2003

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 Master of Science.

_____
Prof. Dr. Mesude İŞCAN
Head of the 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 Master of Science.

_____
Prof. Dr. İnci TOGAN
Supervisor

Examining Committee Members

Prof. Dr. İnci TOGAN _____

Prof. Dr. Işık BÖKESOY _____

Prof. Dr. Hüseyin Avni ÖKTEM _____

Assist. Prof. İrfan KANDEMİR _____

Dr. Ayşe ERGÜVEN _____

ABSTRACT

ALU INSERTION POLYMORPHISMS IN ANATOLIAN TURKS

DİNÇ, Havva
M. Sc., Department of Biological Sciences
Supervisor: Prof. Dr. İnci TOGAN

September 2003, 99 pages

In the present study; ten autosomal human-specific *Alu* insertion polymorphisms; ACE, APO, A25, B65, D1, FXIIIB, HS4.32, HS4.69, PV92 and TPA25 were analyzed in approximately 100 unrelated individuals from Anatolia. *Alu* insertion polymorphisms offer several advantages over other nuclear DNA polymorphisms for human evolution studies.

The frequencies of the ten biallelic *Alu* insertions in Anatolians were calculated and all systems were found to be in Hardy-Weinberg equilibrium ($p>0.05$).

By combining the results of this study with results of previous studies done on worldwide populations, the genetic distance (Nei's D_A) between each pair of

populations was calculated and neighbor joining trees were constructed. In general, geographically closer populations were found to be also genetically similar. Principal component analysis (PCA) was performed and Anatolia was found to be in the European cluster. As a result of PCA; it was concluded that FXIIIB, PV92 and ACE were the variables contributing the most to the explanation of the variation between the populations. Additionally; canonical variates analysis (CVA) concluded that the most discriminative markers for the groups of populations were PV92, D1, ACE and HS4.32.

Pair-wise Fst values were also calculated between Anatolians and some of the populations for which the data was available. It was concluded that, Anatolians have non-significant pair-wise Fst values with Swiss and French Acadian populations.

Lastly, heterozygosity *vs.* distance from centroid graph was constructed and it was found that Anatolians and India-Hindu had exactly the expected heterozygosity value predicted by the model of Harpending and Ward (1982).

Keywords: *Alu* insertion polymorphisms, Anatolia, neighbor joining tree, principal component analysis, discriminant function analysis, Fst, heterozygosity *vs.* distance from centroid.

ÖZ

ANADOLU TÜRKLERİ'NDE *ALU* ARA-İLAVE POLİMORFİZİMLERİ

DİNÇ, Havva
Yüksek Lisans, Biyolojik Bilimler Bölümü
Tez Yöneticisi: Prof. Dr. İnci TOGAN

Eylül 2003, 99 sayfa

Bu çalışmada on otozomal *Alu* ara-ilave polimorfizimleri (ACE, APO, A25, B65, D1, FXIIIB, HS4.32, HS4.69, PV92 and TPA25) Anadolu'dan birbiri ile akraba olmayan yaklaşık 100 bireyde incelenmiştir. *Alu* ara-ilave polimorfizimlerinin insan evrimi çalışmalarında, diğer nükleer DNA polimorfizimlerine kıyasla birçok avantajı vardır.

On *Alu* ilavelerinin Anadolu'daki frekansları hesaplanmış ve hepsi de Hardy-Weinberg (H-W) dengesinde bulunmuştur ($p>0.05$).

Bu çalışmanın sonuçlarını daha önce dünya çapında yapılan çalışmaların sonuçları ile birleştirerek, her populasyon çifti arasındaki genetik uzaklık (Nei'nin D_A)

hesaplanmış ve komşu birleştirme ağaçları çizilmiştir. Genel olarak, coğrafi olarak yakın populasyonların genetik olarak da benzer olduğu görülmüştür. Temel öğeler analizi uygulandığında Anadolu'nun Avrupa grubunda olduğu bulunmuştur. Bu analiz sonucunda, FXIIIB, PV92 ve ACE değişkenlerinin populasyonlar arası varyasyonu açıklamada en çok katkısı olan değişkenler oldukları sonucu ortaya çıkmıştır. Ek olarak, ayrışım fonksiyonu analizi sonucunda ise, populasyon grupları için en ayırt edici genetik işaretlerin PV92, D1, ACE ve HS4.32 oldukları görülmüştür.

Anadolu ve verisi mevcut olan bazı populasyonlar arasında ikili Fst değerleri hesaplanmıştır. Sonuç olarak, İsviçre ve Fransız Akadyan populasyonları ve Anadolu arasında istatistiksel olarak anlamsız değerler elde edilmiştir.

Son olarak, heterozigotluk ve merkezden uzaklık grafiği çizildiğinde, Anadolu ve Hindistan-Hindu populasyonlarının heterozigotluk değerlerinin Harpending ve Ward (1982) modeline göre beklenen değerler oldukları gözlenmiştir.

Anahtar kelimeler: *Alu* ara-ilave polimorfizmleri, Anadolu, komşu birleştirme ağacı, temel öğeler analizi, ayrışım fonksiyonu analizi, Fst, heterozigotluk ve merkezden uzaklık.

To My Parents;
Nazlı and Önder Dinç

ACKNOWLEDGEMENTS

First of all, I want to express my appreciation to my dear supervisor Prof. Dr. İnci TOGAN for guiding and understanding me so well, and for being more than a supervisor to me. I am also grateful to Assist. Prof. İrfan KANDEMİR who acted as my co-supervisor and helped a lot while doing the statistical analyses. I would also like to thank Prof. Dr. Mark STONEKING for always replying immediately and supplying what I have asked for.

I would like to thank my lab-mates; Evren, Emel, Ceren who were always with me with their helpful ideas and support when I was dealing with the problems in my experiments. And also thanks to Can, whose presence in the lab always supplied me positive energy. I am grateful to Seral who first taught me the techniques that I have used for this study.

Thanks to Gülen for supplying me the valuable DNA isolation protocol and also Alper for not hesitating to drive me to the other end of Ankara. I am grateful to Damla who thought with me and discussed every problem in my experiments.

I am thankful to Şebnem for her precious effort to help me to collect samples and also to other friends who helped me in the sample collection. I would like to thank everyone who gave me samples without hesitating and made this study real.

Thanks to Aydın for his support and taking care of the first bound copies of this thesis. Of course, I can't deny the support of Kofona!! You are always giving me strength friends. Thank you!

I am grateful to my parents; Nazlı and Önder Dinç, my sister; Ayten, my uncle İlkay Gilanlıoğlu and my whole family for their endless encouragement and support.

What are the needs of a person while writing a thesis? Patience, support, priority, help, friendship, motivation, being loved, companionship, chocolate, coffee and *etc.* It is impossible to expect all of these from people around you, however I owned more than these during the preparation of this thesis. Banu, Aysun, Anıl and Aslı; you were exactly like angels and became an indispensable part of me. It wouldn't be possible without you! THANK YOU FOR BEING WITH ME IN EVERY STEP...

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE

LIST OF TABLES

TABLE

LIST OF ABBREVIATIONS

µl	microliter
µM	micromolar
bp	base pair
BSA	Bovine Serum Albumin
$MgCl_2$	Magnesium Chloride
CAR	Central African Republic
CVA	Canonical Variates Analysis
dH_2O	Distilled Water
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
EDTA	Ethylenediaminetetraacetic acid
Et-Br	Ethidium Bromide
F	Female
HCl	Hydrogen Chloride
HW	Hardy-Weinberg
K_2EDTA	Potassium EDTA
$KHCO_3$	Potassium Bicarbonate
M	Male

ml	mililiter
mM	milimolar
mtDNA	Mitochondrial DNA
mya	million years ago
NaAc	Sodium Acetate
NaCl	Sodium Chloride
NaOH	Sodium Hydroxide
ng	nanogram
NH_4Cl	Ammonium Chloride
NJ	Neighbor Joining
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PNG	Papua New Guinean
SDS	Sodium Dodecyl Sulfate
STR	Short Tandem Repeats
TBE	Tris Borate EDTA
TE	Tris EDTA
UAE	United Arab Emirates
UV	Ultra Violet

CHAPTER 1

INTRODUCTION

The analysis of genetic variation helps to find answers to numerous evolutionary questions. When human populations are studied, it is important to point out that the time scale is relatively small and their evolutionary history gives vital information about the world's historical asset. In this regard, human population genetic studies contribute to human population histories in both a narrow geographical scale and a holistic approach to the origin of human species. Genetic interpretation of the past and present contributes to the interdisciplinary manner of reconstructing the unique history of human populations, which is suggested to be composed of genetics, archeology and linguistics (Renfrew, 1992).

Population genetics aims at understanding the effects of evolutionary forces in a gene pool and helps to quantify genetic relatedness of populations. Until recent developments in molecular genetics, analysis of the gene pool was usually indirect and limited to a small number of markers. These were based on the study of gene products that are mostly polymorphic proteins, which are usually referred to as "classical markers". However, polymorphism based on direct observation of DNA is more useful,

because polymorphism is higher in DNA sequences (Cavalli-Sforza, 1998). The other advantages of studying DNA are that: (i) very small amounts of DNA are enough for the study with the aid of polymerase chain reaction (PCR) amplification, (ii) non-coding DNA regions can be employed, (iii) non-coding regions can be expected as free of selection at least more than that of coding regions and (iv) mutations not resulting in electrophoretic mobility changes can be detected by techniques using DNA. Examples of DNA markers include single nucleotide polymorphisms (SNPs), biallelic polymorphisms (BAs), restriction fragment length polymorphisms (RFLP), short tandem repeats (STR), microsatellites, minisatellites, mitochondrial DNA (mtDNA) and Y chromosome markers.

In this study, ten *Alu* insertion polymorphisms, which are DNA markers, were used to determine the *Alu* insertion variability of Anatolia that is situated in a unique and evolutionarily important geographical location. In this study, only the genetics component of the interdisciplinary manner of Renfrew (1992) for reconstructing the unique history of human populations was used. Combination of the results of this study with archeology and linguistics can supply a better understanding of the history of Anatolia.

1.1 Brief Review of Evolutionary History of Modern Human Being

Various studies have been performed on the origin of modern human beings by using different genetic markers (Cann *et al.*, 1987, Vigilant *et al.*, 1991, Harpending *et al.*, 1993, Nei and Roychoudhury, 1993, Stoneking *et al.*, 1997, Jorde *et al.*, 2000, Watkins *et al.*, 2001, Templeton, 2002). According to Nei and Roychoudhury (1993), anatomically modern humans originated in Africa approximately 200,000 years ago, started to move out of Africa 100,000 years ago and migrated towards Asia 70,000-50,000 years ago. However, in a recent study; Templeton (2002) performed formal statistical analysis of human haplotype trees for mtDNA, Y-chromosomal DNA, two X-linked regions and six autosomal regions by using Nested Clade Analysis. As a result, it was concluded that there were at least two major movements of modern humans out of Africa after the original spread of *Homo erectus* about 1.7 mya. 95% confidence interval for the older and more recent out-of-Aftrica range-expansion events was found to be 0.42 to 0.84 mya and 0.08 to 0.15 mya, respectively.

In the Upper Paleolithic period (60,000-10,000 years ago), it was suggested that several major demographic expansions, which can impose significant imprints on genetic landscape, have influenced the Asian and Balkan genetic make-up as a result of the first introduction of modern humans into Europe (Calafell *et al.*, 1996, Comas *et al.*, 1996 and references therein).

Moreover, Anatolia is one of the oldest settlement areas; especially in Çatalhöyük near Konya, Turkey (8,500 before present). A second introduction into

Europe is suggested by genetic studies proposing the spread of farming in the Neolithic (10,000-5,000 years ago) from the Middle East and Anatolia through the Balkans and into Europe (Calafell *et al.*, 1996 and references therein). Zerjal *et al.*, (1997) state that Y chromosome provides both information about population relationships in Asia and evidence for a substantial paternal genetic contribution of Asians to northern European populations. Therefore, Anatolia was not the only route for the migration to Europe.

Last but not least, many migratory events also took place in Anatolia in recent millennia. In this time scale, Anatolia was well populated by various civilizations (Assyrians, Hittites, Phrygians, Lydians, Urartians, Persians, Romans, Byzantines and Vennetians) until the arrival of the Oghuz Turks (Turkic nomadic people) in the 11^{th} century AD. The Oghuz Turks, who imposed the language of this Turkic group to Anatolia, were from the area between Mongolia and the Caspian Sea from the 9^{th} century AD (Roberts, 1993).

The unique geographical location and prosperous history of Anatolia emphasized above have been triggering factors for studying the genetic structure of the Anatolian Turkish population. Therefore, there are numerous valuable population genetic studies performed on Anatolians by using several genetic markers in order to understand the genetic background of Anatolia. Some of the studies based on ABO blood groups, red blood cell enzymes and proteins have been performed by Saatçioğlu (1979), Togan and Ergüven (1994), Önde and Kence (1995), Togan *et al.* (1996) and Ergüven (1997). There are also other studies based on STR systems and PCR-based genetic markers in Anatolia (Isawa *et al.*, 1997, Takeshita *et al.*, 1997, Vural *et al.*,

1998). There are additional studies based on mtDNA and other polymorphic markers (Calafell *et al.*, 1996, Comas *et al.*, 1996, Benedetto *et al.,* 2001 and İçener, 2001). Up till now, Gerçeker's (1998) is the only study based on polymorphic *Alu* insertions in Anatolian Turks.

1.2 Specialty of *Alu* Insertion Polymorphisms

Nongenic DNA, which is genomic DNA that fails to encode proteins, is classified into two main groups: Highly repetitive DNA, which constitutes about 5% of the human genome, and moderately repetitive DNA. Moderately repetitive DNA consists of either tandomly repeated (e.g. STR, microsatellites and minisatellites) or interspersed (e.g. mobile elements) sequences. Interspersed elements can be either long or short as illustrated in Figure 1 (Klug and Cummings, 1997). In humans, the most prominent example for long interspersed elements (LINEs) is a family called L1, whose members are as large as 7000 base pairs long. Short interspersed elements (SINEs) range in size from 90 to 400 base pairs and are present in the human genome at a high copy number. The most abundant SINEs are the *Alu* repeats, which are the largest family of the mobile elements in the human genome. These repeats, which are approximately 300 base pairs long, have more than 1 million copies. Since they have a high copy number, the *Alu* gene family comprises more than 10% of the mass of the human genome and as *Alu* sequences accumulate preferentially in gene-rich regions, they are not uniformly distributed in the human genome (Deiniger *et al.,* 1992, Batzer and Deiniger, 2002).

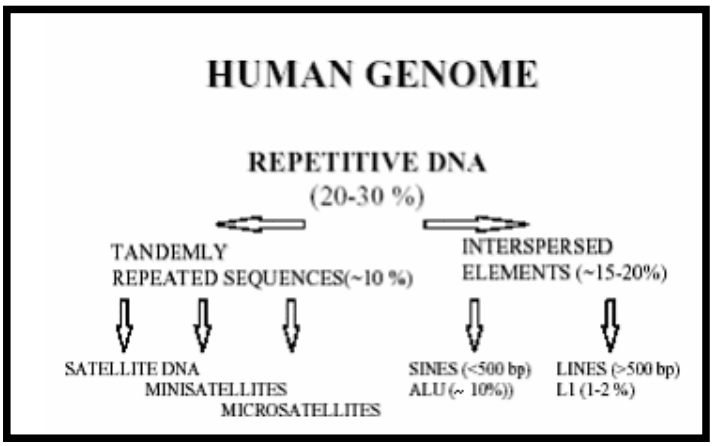


Figure 1. Schematic organization of repetitive DNA in the human genome (with modification from http://www.promega.com/geneticidproc/eusymp2proc/01.pdf).

The origin and amplification of *Alu* elements, which are restricted to primate genomes, are evolutionarily recent events that coincided with the radiation of primates in the past 65 million years (Batzer *et al.,* 1996a,b and references therein). *Alu* elements are ancestrally derived from the 7SL RNA gene, which forms part of the ribosome complex (Ullu and Tschudi, 1984). In other words, the background of *Alu* elements can be traced back to an initial gene duplication early in primate evolution and to the subsequent amplification of these elements (Batzer and Deininger, 2002).

The spread of *Alu* elements involves a retrotransposition from a so-called "master" *Alu* element (Shen *et al.,* 1991). Briefly, the steps in retrotransposition involve the synthesis of an RNA copy from the master gene by RNA polymerase III, the reverse transcription to form a cDNA copy, and then the insertion of this cDNA copy into a new location in the genome. This daughter copy is generally not capable of generating

further copies, unless it by chance happens to be inserted into a location near a sequence capable to permit an initiation of RNA polymerase III transcription, or if a subsequent mutation creates an initiation site near a previously silent *Alu* element (Deiniger *et al.,* 1992).

During evolutionary time, the master *Alu* element accumulates new mutations and generates new daughter elements. As shown in Figure 2, once a new mutation arises in the master *Alu* element, all subsequent daughter elements will contain that mutation. This leads to the formation of families of *Alu* elements that can be distinguished based on the hierarchical accumulation of diagnostic or subfamily-specific substitutions (Shen *et al.,* 1991, Deiniger *et al.,* 1992, Batzer *et al.,*1996b, Stoneking *et al.*, 2001).

One of the most recently formed groups of *Alu* elements within the human genome has been termed as Human Specific (HS) (Batzer and Deininger, 1991). Almost all of the recently integrated human *Alu* elements belong to one of several small and closely related 'young' *Alu* subfamilies, known as Y, Yc1, Yc2, Ya5, Ya5a2, Ya8, Yb8 and Yb9 (Batzer *et al.*, 1990, Carroll *et al.*, 2001, Roy-Engel *et al.*, 2001, Batzer and Deininger, 1991, Batzer *et al.*, 1995 and Jurka, 1993). With the exception of the *Alu* Y-family members and of a small number of elements from the other 'young' subfamilies, individual members of these young *Alu* subfamilies that are present in the human genome are not found at orthologous positions in the genomes of other great apes. These largely human-specific *Alu* subfamilies represent only ~0.5% of all the *Alu* repeats in the human genome (Batzer and Deininger, 2002).

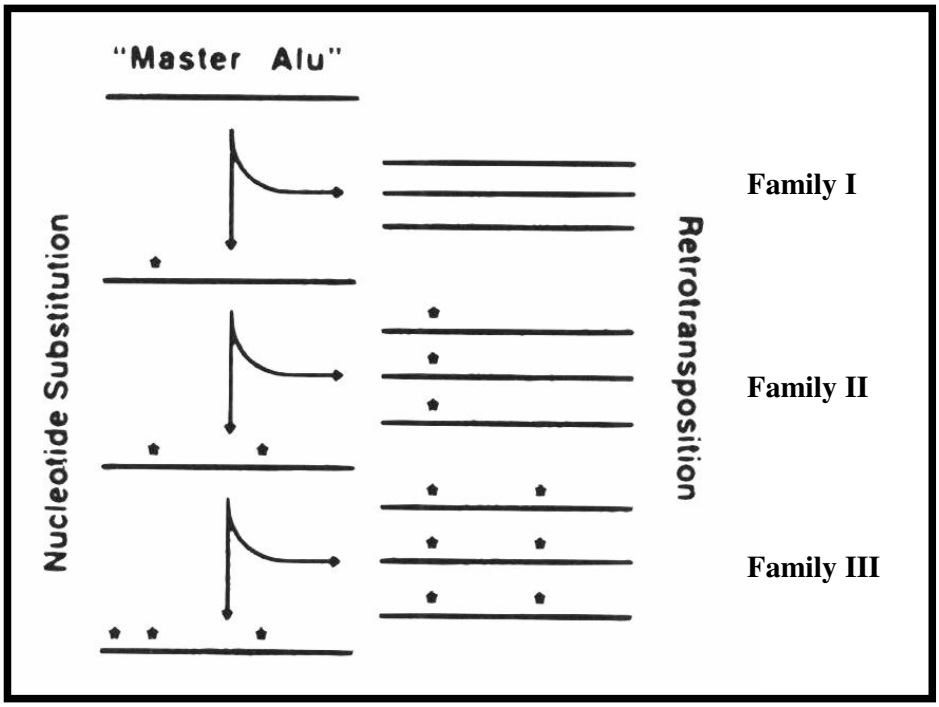


Figure 2. Model of how nucleotide substitutions in *Alu* "master" elements lead to the formation of families of daughter elements formed by retrotransposition. Nucleotide substitutions are indicated by asterisk (Stoneking *et al.*, 2001).

After a new, neutral *Alu* insertion integrates in the genome, it is subject to random genetic drift. So, the probability that it will be lost from the population is initially quite high, depending on the size of the population. However, in accordance with chance events, the *Alu* element may increase in frequency in the population. During the time until fixation or loss, the population is polymorphic for the presence or the absence of the *Alu* element at that specific chromosomal location, and this type of polymorphism is known as an *Alu* insertion polymorphism (Batzer and Deininger, 1991, Batzer *et.al.,* 1991, Batzer *et.al.*, 1994).

The 'young' *Alu* subfamilies (~25 mya) started to integrate into the genome much before the divergence of humans from African apes (~5 mya), as illustrated in Figure 3. Thus, most *Alu* repeats became monomorphic for their insertion sites among diverse human genomes before the divergence of humans. In addition, approximately 25% of the young *Alu* repeats inserted into the human genome so recently that they are dimorphic for the presence or absence of the insertion, which makes them a useful source of genomic polymorphism (Batzer *et al.*, 1991, Batzer and Deininger, 1991, Batzer *et al.*, 1995, Carroll *et al.*, 2001, Roy-Engel *et al.*, 2001).

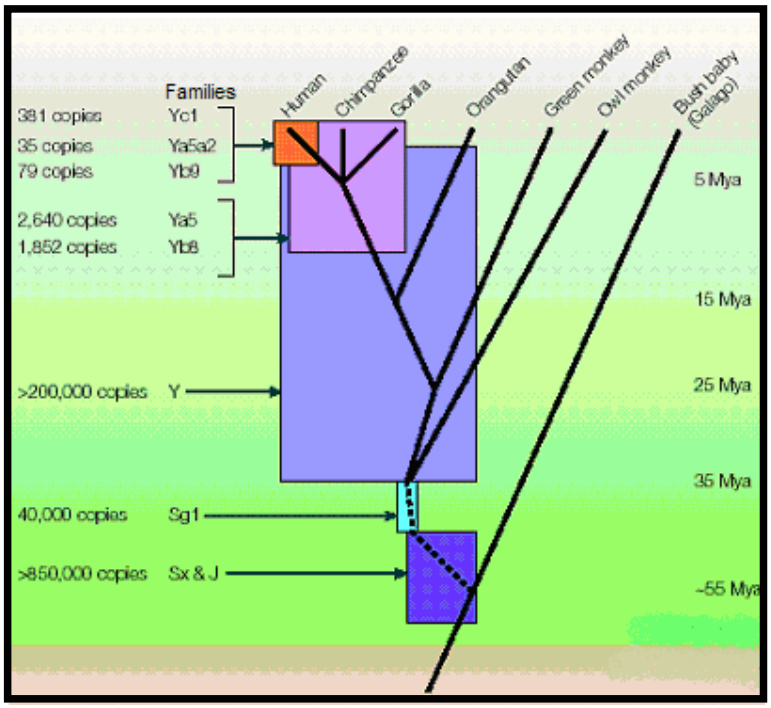


Figure 3. The expansion of *Alu* elements in primates. The 'young' *Alu* subfamilies are denoted with a capital Y in the name of the subfamily (Batzer and Deininger, 2002).

Alu insertion polymorphisms have several attractive features that make them unique elements for the study of human population genetics. First of all, the genotypes of *Alu* insertion polymorphisms are easy to determine by typing with rapid, nonradioactive, simple PCR based assays. They are biallelic polymorphisms with three possible genotypes: homozygous for the presence of the *Alu* element, heterozygous with one chromosome having the *Alu* element and the other lacking it and homozygous for the absence of the *Alu* element (Novick, *et al.*, 1996).

Secondly, once inserted into a new location, an *Alu* element is rarely subject to deletion. Even if deleted, it would not be an exact excision, but instead it would leave behind a molecular signature of the original insertion event by either retaining a part of the *Alu* element and/or deleting some of the flanking region (Edwards and Gibbs, 1992). Therefore, *Alu* insertion polymorphisms are stable markers that reflect a unique evolutionary event, which is the insertion of an *Alu* element into a new chromosomal location.

Thirdly, *Alu* insertion polymorphisms display unique events that occurred during human evolution. Since there are 3 billion nucleotides in the haploid human genome, the probability that an *Alu* element would insert between the exact same two nucleotides at two different times during evolution is insignificant. Therefore, there is no parallel gain or loss of *Alu* elements at a particular chromosomal location, so all chromosomes that carry a polymorphic *Alu* element must be identical by descent. Hence, polymorphic *Alu* insertions reflect population relationships more accurately than other genetic markers, such as RFLP, SNPs, STRs, microsatellites, mtDNA markers, *etc.* The

disadvantage of these latter genetic markers lies in the fact that they have arisen as the result of several independent parallel mutations at different times. Therefore, they are identical by state rather than descent and thus may have not been inherited from a common ancestor. This is because the same allele could arise independently at different times during human evolution (Batzer and Deininger, 2002).

Lastly, the ancestral state of *Alu* insertion polymorphisms is known to be the absence of the *Alu* element at a particular chromosomal location and the derived state is the presence of the *Alu* element. Alleles that are identical by descent must have been inherited from a common ancestor. The precise knowledge of the ancestral state of a genomic polymorphism, which is very important in phylogenetic analyses, permits the construction of phylogenetic trees without making too many assumptions (Batzer *et al.*, 1994, 1996a, Stoneking *et al.,* 1997).

Furthermore, *Alu* insertion polymorphisms are autosomal markers that reflect both the maternal and paternal history of a population. When these polymorphisms are compared with other genetic markers, it emerges again that they are very useful. For instance, mtDNA is used in human population genetic studies because it has a high mutation rate, is a haploid and is the maternal mode of inheritance. However, inferences about human evolution based on mtDNA are quite limited because mtDNA is a single genetic locus. Consequently, the history of such a single gene can differ from the history of the species, due to selection or chance events involving that gene. Apart from this, RFLP loci are relatively time-consuming to assay and require relatively large amounts of DNA.

A*lu* insertion polymorphisms in different populations have been analyzed in several studies such as; Batzer *et al.* (1994, 1996a), Stoneking *et al.* (1997), Gerçeker (1998), Comas *et al.* (2000), İçener (2001), Watkins *et al.* (2001), Nasidze *et al.* (2001), Donaldson *et al.* (2002) and Romualdi *et al.* (2002). This study will be complementary for these studies.

Analyses performed in this study were all based on the allele frequencies of the *Alu* insertion polymorphisms data. However, there are also other ways to analyze *Alu* insertion polymorphisms. For instance; the nucleotide diversity of the *Alu* elements are also used to examine human population structure and to estimate when the insertion event occurred (Knight *et al.,* 1996). Another alternative approach is to determine haplotypes consisting of *Alu* insertion polymorphisms and one or more closely linked hypervariable loci (such as an STR locus); thereby allowing one to compare the diversity on chromosome with and without the *Alu* element (Tishkoff *et.al.,* 2000).

1.3 The Objective of the Study

New researches on the Anatolian Turkish population with polymorphic genetic markers are needed to understand genetic background of Anatolia. Hence, this study includes the analysis of ten *Alu* insertion polymorphisms (ACE, APO, A25, B65, D1, FXIIIB, HS4.32, HS4.69, PV92 and TPA25) in approximately 100 Anatolian Turks. The incorporation of the information on the Anatolian Turkish population obtained in this study will contribute to filling in the gap in our knowledge of *Alu* insertion

polymorphism variation in Anatolia and will allow a more continuous interpretation of the evolution of *Alu* insertion.

The purpose of this study can be summarized as follows;

- to determine the *Alu* insertion variability of the Anatolian Turks,
- to resolve the genetic relatedness of Anatolian Turks to other populations,
- to find out the cluster that Anatolia belongs to among the worldwide populations according to *Alu* insertion polymorphisms,
- to show the discrimination power of ten *Alu* insertion polymorphisms in differentiating populations,
- to contribute new knowledge to the evolutionary history of Anatolia.

CHAPTER 2

MATERIALS AND METHODS

2.1 MATERIALS

Samples for this study were collected from 102 individuals from different regions of Turkey. The regions of sampling for all of the individuals studied and their gender are presented in Appendix A. Samples were collected from healthy individuals and special care was taken to avoid sampling from related individuals. Individuals whose families had been from a same particular area for at least three generations were selected and this was recorded with the questionnaires filled by the donors. A sample of this questionnaire is given in Appendix B. The permission of the DNA donors was also taken and a form to obtain the consent of the sampled individuals is shown in Appendix B. Ingredients of the chemical solutions used are presented in Appendix E.

2.2 METHODS

2.2.1 DNA Isolation

DNA isolation was performed from both blood and buccal samples. DNA isolation from whole blood was performed with a method called the Phenol-Chloroform-Isoamylalcohol (25:24:1) isolation method. However, DNA isolation from buccal samples was performed by using two different isolation methods.

2.2.1.1 DNA Isolation from Blood Samples

Blood samples (≅ 10 ml) were collected into EDTA containing tubes to prevent coagulation and stored at +4 °C until use. 10 ml blood was completed to 50 ml with 2X lysis buffer to lyse the red blood cells. Tubes were mixed for 10 minutes by inversion and then centrifuged at 3000 rpm for 10 minutes at +4 °C to precipitate nuclei. The pellet was resuspended in 3 ml of salt-EDTA buffer by vortexing. Then, 0.3 ml of 10% SDS and 150 µl proteinase-K (10 mg/ml) were added and the tubes were incubated at +55 °C for 3 hours. After the incubation, 3 ml phenol was added and this suspension was centrifuged at 3000 rpm for 10 minutes at +4 °C. The supernatant was mixed with 3 ml phenol-chloroform-isoamylalcohol solution (25:24:1) and centrifuged at 3000 rpm for 10 minutes at +4 °C. Then, the upper phase was collected carefully and transferred into a new glass tube by a transfer pipette. Then 1 ml 3 M Sodium Acetate (NaAc) and

about 2 volumes of 99% ethanol were added to precipitate and collect the DNA. Glass tubes were mixed gently by inversion and then the DNA was transferred to eppendorf tubes containing 500 µl Tris-EDTA (TE) buffer (pH: 7.5) and stored at -20 °C.

2.2.1.2 DNA Isolation from Buccal Samples: Method I

Buccal cells were collected from the inside of both of the cheeks with a sterile brush by scraping for at least one minute. This should be done strongly and carefully to collect enough cells. The head of the brush was placed in a 0.5 ml eppendorf tube containing 400 µl of 50 mM sodium hydroxide (NaOH) solution and incubated at 97 °C for 15 minutes. After removing the head of the brush from the tube, 150 µl 1 M Tris-HCl (pH: 8) solution was added and stored at -20 °C. (This technique was performed according to the information gathered by personal communication from the Veterinary Genetics Laboratory of University of California, Davis, USA.)

2.2.1.3 DNA Isolation from Buccal Samples: Method II

Epithelial cells were collected from the inside of both of the cheeks by scraping with a sterile brush for at least one minute. The scraping should be strong enough to collect as many cells as possible. The head of the brush was placed into a 1.5 ml eppendorf tube containing 700 µl 2X lysis buffer. The cells were digested overnight by

adding 10 µl proteinase-K (20 mg/ml). After digestion, the brush was not removed since it still contained a lot of material and the solution was extracted by adding 500 µl phenol-chloroform-isoamylalcohol solution (25:24:1). The mixture was centrifuged at 10000 rpm for 2 minutes at room temperature (RT). The upper aqueous phase was transferred into a new tube and the extraction was repeated. The second aqueous phase was then transferred into a new tube and extracted once more with 500 µl chloroform-isoamylalcohol solution (24:1). The phases were separated by centrifuging at 10000 rpm for 2 minutes at RT. This new aqueous phase was collected into a new tube and 0.1 volumes of 3 M Sodium Acetate (NaAc) together with 0.6 volumes of isopropanol were added to the solution. The tube was incubated overnight at -20 °C. The DNA was precipitated by centrifuging at 10000 rpm for 10 minutes. The pellet was rinsed with 500 µl 70% ethanol and the DNA was precipitated again by centrifuging at 10000 rpm for 5 minutes at RT. After discarding the supernatant, the pellet was dried at RT and resuspended in 100 µl of Tris-EDTA (TE) buffer. The DNA samples were stored at -20 °C. (This technique was performed according to the information obtained by personal communication from the Department of Molecular Biology and Genetics, Bosphorus University, Istanbul, Turkey.)

2.2.2 Checking the Presence of DNA

Presence of DNA was checked by agarose gel electrophoresis in order to be sure that DNA was isolated successfully. 0.6 % agarose gel was prepared by boiling agarose in 0.5X TBE buffer. The gel was poured into an electrophoresis plate and left in room temperature for about 30 minutes for polymerization. 1 µl of newly isolated genomic DNA, 6 µl of 6X loading buffer (bromophenol blue dye) and 6 µl of dH_2O were mixed and then loaded into the wells of the gel. The gel was run at 100 V for about 30 minutes in 0.5X TBE buffer, stained in 0.5 µl/ml ethidium bromide (Et-Br) solution and then was visualized under UV light. The presence or absence of smears and the migration patterns of the bands on the gel corresponds to the presence and the quality of DNA.

2.2.3 Amplification of DNA with Polymerase Chain Reaction (PCR)

The PCR amplification of the DNA samples for each *Alu* insertion polymorphism (ACE, APO, A25, B65, D1, FXIIIB, HS4.32, HS4.69, PV92 and TPA25) was performed in 25-µl amplification reactions using;

- 1X PCR Buffer (50mM KCl and 10mM Tris-HCl)
- 1 mM BSA (10 mg/ml)
- 1.5 mM $MgCl_2$ for ACE, APO, FXIIIB, HS4.32, HS4.69, PV92, TPA25 and 4 mM $MgCl_2$ for A25, B65, D1
- 200 µM dNTP (5 mM)

- 0.32 µM Forward and Reverse Primers (10 µM)
- 1 unit Taq Polymerase (5 units/µl)
- 50 ng of Genomic DNA
- dH_2O

The specific oligonucleotide primer sequences together with the annealing temperatures used for each *Alu* insertion are given in Table 1.

Hot start at 94 °C for 5 minutes is only applied to genomic DNA just before the addition of the reaction mixture to improve the accuracy of primer annealing. The amplification conditions for the oligonucleotides are shown in Table 2. In the denaturation step, double stranded DNA is dissociated into single strands, while in the annealing step, primers are annealed to the single stranded DNA. In the extension step, the Taq polymerase extends the oligonucleotide primers in the 5'-to-3' direction using the single-stranded DNA bound to the primer as a template. So, at the end of these cycles, the enzymatic amplification of microgram quantities of specific DNA sequences were maintained (Klug and Cummings, 1997).

2.2.4 Analysis of the PCR Products

Analysis of the PCR products was done by agarose gel electrophoresis. 2 % agarose gel was prepared by boiling agarose in 0.5X TBE buffer, pouring it into an electrophoresis plate and leaving it at RT for 30 minutes for polymerization. 9 µl of each PCR product was mixed with 7 µl of 6X bromophenol blue dye and loaded into the wells of the gel. The gel was run in 0.5X TBE buffer at 100 V until the bands reached the end of the gel. Then, gel was stained in 0.5 µl/ml ethidium bromide (Et-Br) solution and the amplification products were directly visualized by UV fluorescence. The photograph of the gel was obtained by a gel image system. Some examples of gel photographs showing the result of the amplification of *Alu* insertions with three different genotypes are shown in Figures 4-8. The length of the PCR products; according to whether the *Alu* element is present or absent, the chromosomal locations and subfamilies of each *Alu* insertion are given in Table 3.

Table 1. Oligonucleotide primer sequences and annealing temperatures for each *Alu* insertion.

***Alu* Insertion**	**5' Primer Sequence (5'-3')**	**3' Primer Sequence (5'-3')**	**Annealing Temperature**
ACE	CTGGAGACCACTCCCATCCTTTCT	GATGTGGCCATCACATTCGTCAGAT	58 ^{o}C
APO	AAGTGCTGTAGGCCATTTAGATTAG	AGTCTTCGATGACAGCGTATACAGA	50 ^{o}C
A25	CCACAAATAGGCTCATGTAGAAC	TATAATATGGCCTGGATTATACC	60 ^{o}C
B65	ATATCCTAAAAGGGACACCA	AAAATTTATGGCATGCGTAT	60 ^{o}C
D1	TGCTGATGCCCAGGGTTAGTAAA	TTTCTGCTATGCTCTTCCCTCTC	60 ^{o}C
FXIIIB	TCAACTCCATGAGATTTTCAGAAGT	CTGGAAAAAATGTATTCAGGTGAGT	56 ^{o}C
HS4.32	GTTTATTGGGCTAACCTGGG	TGACCAGCTAACTTCTACTTTAACC	57 ^{o}C
HS4.69	GTCTGAATGTTCGTGTCCCC	GTCCAAGTTCAAGGCACCAG	60 ^{o}C
PV92	AACTGGGAAAATTTGAAGAAAAGT	TGAGTTCTCAACTCCTGTGTGTTAG	54 ^{o}C
TPA25	GTAAGAGTTCCGTAACAGGACAGCT	CCCCACCCTAGGAGAACTTCTCTTT	58 ^{o}C

Table 2. PCR condition of the 10 *Alu* insertions.

Duration	Step	Temperature	Number of Cycles
5 minutes	Denaturation	94 °C	1
1 minute	Denaturation	94 °C	32
1 minute	Annealing	At specified annealing temperature	
1 minute	Extension	72 °C	
10 minutes	Final Extension	72 °C	1

Table 3. Sizes of the PCR products in the presence and absence of *Alu* insertion, the chromosomal location and subfamily of each insertion.

Alu Insertion	Insertion Positive (bp)	Insertion Negative (bp)	Chromosomal Location	Subfamily
ACE	490	190	17	Ya5
APO	433	122	11	Ya5
A25	552	268	8	Ya5
B65	423	81	11	Ya5
D1	622	333	3	Yb8
FXIIIB	725	425	1	Ya5
HS4.32	601	289	12	Ya5
HS4.69	572	262	6	Ya5
PV92	416	101	16	Ya5
TPA25	424	125	8	Ya8

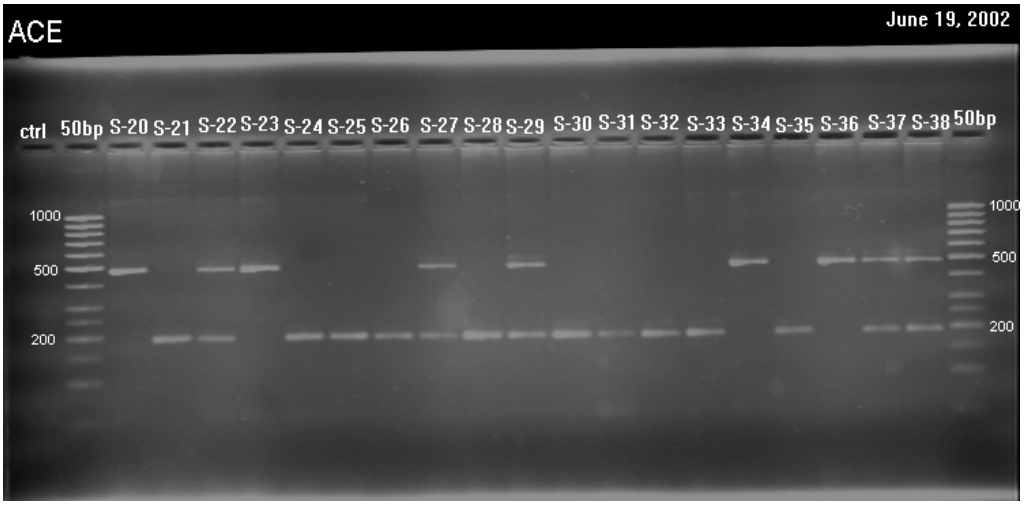


Figure 4. Photograph of a 2 % agarose gel containing the PCR products of ACE.
ctrl: Negative control
50bp: 50bp DNA size marker
S-20, S-23, S-34, S-36: Homozygous individuals for the presence of ACE insertion (+/+)
S-22, S-27, S-29, S-37, S-38: Heterozygous individuals (+/-)
S-21, S-24, S-25, S-26, S-28, S-30, S-31, S-32, S-33, S-35: Homozygous individuals for the absence of ACE insertion (-/-)

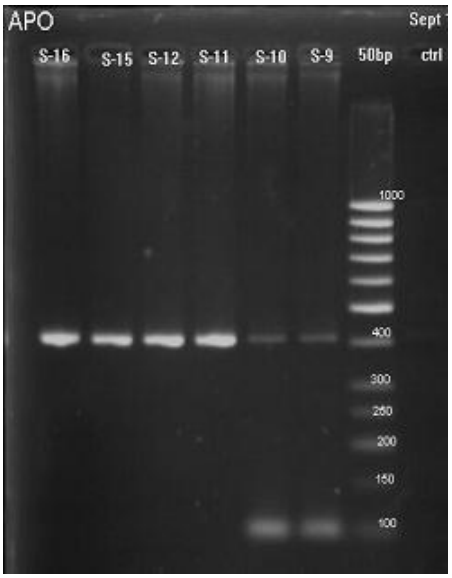


Figure 5. Photograph of a 2 % agarose gel containing the PCR products of APO.
ctrl: Negative control
50bp: 50bp DNA size marker
S-16, S-15, S-12, S-11: Homozygous individuals for the presence of APO insertion (+/+)
S-10, S-9: Heterozygous individuals (+/-)

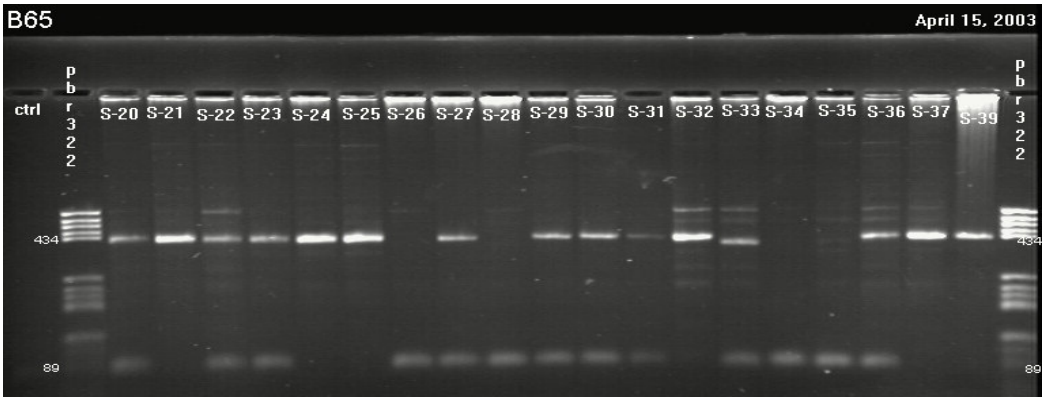


Figure 6. Photograph of a 2 % agarose gel containing the PCR products of B65.
ctrl: Negative control
PBR 322: pBR 322 DNA size marker
S-21, S-24, S-25, S-32, S-37, S-39: Homozygous individuals for the presence of B65 insertion (+/+)
S-20, S-22, S-23, S-27, S-29, S-30, S-31, S-33, S-36: Heterozygous individuals (+/-)
S-26, S-28, S-34, S-35: Homozygous individuals for the absence of B65 insertion (-/-)

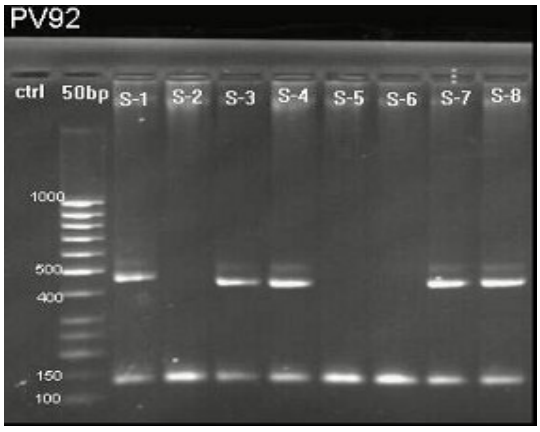


Figure 7. Photograph of a 2 % agarose gel containing the PCR products of PV92.
ctrl: Negative control
50bp: 50bp DNA size marker
S-1, S-3, S-4, S-7, S-8: Heterozygous individuals (+/-)
S-2, S-5, S-6: Homozygous individuals for the absence of PV92 insertion (-/-)

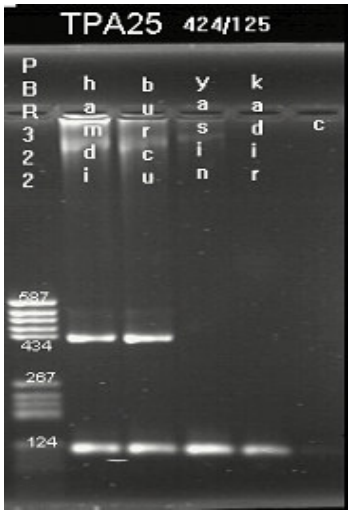


Figure 8. Photograph of a 2 % agarose gel containing the PCR products of TPA25.
c: Negative control
PBR 322: pBR 322 DNA size marker
h, b: Heterozygous individuals (+/-)
y, k: Homozygous individuals for the absence of TPA25 insertion (-/-)

2.2.5 Statistical Analysis of Data

2.2.5.1 Allele Frequencies, Heterozygosities and Evaluation of Hardy-Weinberg Equilibrium

The statistical analysis of the data started with the calculation of relative frequencies of each *Alu* insertion polymorphism in the studied population. This calculation gives the allele frequencies of each *Alu* insertion, which are used frequently in the subsequent steps of the data analysis. Allele frequencies of the insertions were calculated according to the formula given below;

$$\text{Frequency of an allele (p)} = \frac{(2 \times \text{number of homozygotes}) + (\text{number of heterozygotes})}{2 \times \text{total number of individuals (N)}}$$

Heterozygosity which helps to measure the genetic variation in a population was calculated with the formula below;

$$\text{Observed Heterozygosity} = \frac{\text{Number of heterozygotes}}{\text{Total number of individuals}}$$

Calculation of the allele frequencies of each *Alu* insertion in Anatolian Turks and the observed heterozygosities were performed with the help of GENETIX 4.0 software, which is available at http://www.univ-montp2.fr/~genetix/genetix.htm (Belkhir *et al.*, 1996-2001).

The evaluation of the Hardy-Weinberg (HW) equilibrium in Anatolian Turks was performed by calculating the expected genotype frequencies and comparing them with the observed ones. HW equilibrium is based on the following assumptions: (i) mating is random, (ii) allelic frequencies are conserved from generation to generation, (iii) no significant migrations occur, (iv) mutation, selection, genetic drift and gene flow are negligible (Hedrick, 2000). All of these requirements were assumed to be true while calculating the expected genotype frequencies, which were calculated as follows;

Expected frequency of homozygotes = p^2 and q^2

Expected frequency of heterozygotes = $2pq$, where $p + q = 1$

The agreement between observed and expected values was tested by the chi-square (χ^2) test statistic (Daniel, 1999). The general formula for χ^2 is;

$\chi^2 = \Sigma$ (Observed values - Expected values)^2 / Expected values

Chi-square values for each *Alu* insertion were calculated and these values were evaluated in chi-square distribution (df = 1, $\chi^2_{.95}$ = 3.841). The decision on the null hypothesis (H_o), which states that observed and expected frequencies are not different from each other, was made.

2.2.5.2 Nei's Genetic Distance (D_A) and Neighbor Joining (NJ) Tree

Genetic distance analysis, which focuses on average genetic distance between populations, is quite efficient while constructing an evolutionary tree from allele frequency data. Genetic distances, Nei's genetic distances, between the pairs of Anatolian Turkish population and the populations studied in Stoneking *et al.* (1997), Nasidze *et al.* (2001), Romualdi *et al.* (2002) calculated with the GENDIST program in the PHYLIP program package (Felsenstein, 1993). Neighbor-joining trees were produced by using the NEIGHBOR program, 1000 bootstrap replicates were generated by the SEQBOOT program and a consensus tree was built with the CONSENSE program as implemented in PHYLIP 3.6 which is available at http://evolution.genetics.washington.edu/phylip.html (Felsenstein, 1993).

2.2.5.3 Principal Component Analysis (PCA)

To analyze population relationships and to determine the relative positions of populations in 3 dimensional space, the principal component analysis (PCA) was performed with the help of a computer program called NTSYS: Numerical Taxonomy and Multivariate Analysis System (Rohlf, 1993).

In PCA from a mass of variables a set of independent compound axes are synthesized and relative positions of the populations in our study are visualized in the space generated by these axes. The first axis will explain the highest variation of the all data that can be accounted by the compound axes; the second will explain the next highest variation, and so on. Inspection of the weightings of the first few axes will show which variables contribute most to the differences between individuals (Dytham, 2003). As a result, the relative importance of the variables (different *Alu* insertions) for the discrimination of the groups was also studied by means of the PCA.

2.2.5.4 Discriminant Function Analysis

Discriminant function analysis was used to obtain set of weightings that allow the groups (formed by at least two populations) to be distinguished. Multigroup discriminant analysis is also called as Canonical Variates Analysis, which produces weightings that allow the identification of the variables that are the most different between groups and discards the ones that are the same. Therefore, these results lead to

the determination of the variables; that explain the most variation between the groups of the populations and that have higher power of discrimination (Dytham, 2003).

This technique requires that the populations be divided into groups. Thus, the data set in this study was divided into groups according to the clusters obtained from Principal Component Analysis. Canonical variates analysis was performed by using the software program SYNTAX (Podani, 1993).

2.2.5.5 The Fst Analysis

Fst is a measure of degree of genetic differentiation between subpopulations and it can be estimated by the following formula:

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

Where;

H_S = average expected heterozygosity in the subpopulations

H_T = average heterozygosity of the total population

(Nei and Kumar, 2000)

The F indices Wright (1951) does not consider the unequal finite sample sizes and there is some disagreement on the interpretation of the quantities and on the method of evaluating them. Weir and Cockerham (1984) revised the F coefficients in order to

offer some unity to various estimations formulae suggested by different authors. They used θ for Fst. This estimator does not make assumption concerning numbers of populations, sample sizes or heterozygote frequencies and they are suited to small data sets.

Therefore, Weir and Cockerham's approach was used in the calculation of Fst-related genetic distances and it was computed between two populations with the help of the GENETIX 4.0 software (Belkhir *et al.*, 1996-2001).

The Fst values between Anatolia and among the available the geographically closest populations were calculated. The Fst values between Anatolian Turks and populations that are genetically close to Anatolia were also calculated. The data were permutated for 1000 times in order to test the significance of the pair-wise Fst values.

2.2.5.6 Heterozygosity *vs*. Distance from Centroid

The expected heterozygosity under Hardy-Weinberg equilibrium of each population was plotted against the distance of the population from the centroid, which is the arithmetic mean of the allele frequencies, to determine the relative amount of gene flow experienced by and/or size of each population. In this model of Harpending and Ward (1982), the distance from centroid r_i for a population *i* is;

$$r_i = (p_i - P)^2 / (P).(1-P)$$

where p_i and P are the frequency of the *Alu* insertion in population *i*th and in the total population, respectively. This equation was used to compute the distance from centroid for each locus separately and these values were then averaged over the eight loci. Then, heterozygosity *vs.* distance from centroid graph was plotted. In this analysis, Anatolian Turks data together with the data of Stoneking *et al.* (1997) and Nasidze *et al.* (2001) for eight *Alu* insertion polymorphisms (ACE, APO, A25, B65, D1, FXIIIB, PV92, TPA25) were used.

CHAPTER 3

RESULTS

3.1. Distribution of ten *Alu* Insertion Polymorphisms in Anatolia

An average of 100 individuals from Anatolian Turks were typed for each of the ten human-specific *Alu* insertion polymorphisms and all were found to be polymorphic in the Anatolian Turkish population. All of the loci are biallelic and the observed allele frequencies for each *Alu* insertion together with the observed heterozygosities and observed genotype numbers are given in Table 4.

The presence of Hardy-Weinberg equilibrium was tested for Anatolian Turks in ten systems: ACE, APO, A25, B65, D1, FXIIIB, HS4.32, HS4.69, PV92, TPA25 and Anatolian Turks were found to be in Hardy-Weinberg equilibrium in all of them ($p>0.05$). The expected genotype numbers, expected heterozygosities and χ^2 values are also shown in Table 4. The average observed and expected heterozygosity values for Anatolian Turks population was found to be 0.3637 and 0.3752, respectively for ten *Alu* loci.

The sampling sites of the individuals and their genotypes for ACE, APO, A25, B65, D1, FXIIIB, HS4.32, HS4.69, PV92 and TPA25 are illustrated in Figures 9-18 to visualize the distribution patterns of the markers in Turkey. The ACE, B65, D1, HS4.32 and TPA25 insertions, with relatively intermediate insertion frequency values, seemed to be homogeneously distributed all over Turkey. The APO insertion, which has the highest insertion frequency in Turkey (0.9510), was observed to be present extensively in every sampled region of Turkey with homozygous individuals for the presence. Conversely, the A25 insertion, which has the lowest frequency in Turkey (0.0693), was only present as heterozygous individuals and they were homogeneously distributed all over Turkey. Moreover, the FXIIIB insertion with relatively intermediate frequency in Turkey (0.4604) was mainly localized in the central and southern regions of Turkey. Lastly, the HS4.69 and PV92 insertions with frequencies 0.3021 and 0.1520 respectively, seemed to be present mostly as heterozygous individuals and they were distributed homogeneously all over Turkey.

Table 4. Population statistics for 10 *Alu* insertion polymorphisms in Anatolian Turks.

ANATOLIAN TURKS								
***Alu* Insertion**	**N**	**Frequency of the insertion**	**Genotype**	**Observed Number**	**Expected Number**	**χ^2**	**Observed Heterozygosity**	**Expected Heterozygosity**
			+/+	15	11.33			
ACE	102	0.3333	+/-	38	45.33	2.670	0.3725	0.4444
			-/-	49	45.34			
			+/+	92	92.25			
APO	102	0.9510	+/-	10	9.51	0.266	0.0980	0.0932
			-/-	0	0.24			
			+/+	0	0.48			
A25	101	0.0693	+/-	14	13.03	0.555	0.1386	0.1290
			-/-	87	87.49			
			+/+	26	23.77			
B65	101	0.4851	+/-	46	50.45	0.786	0.4554	0.4996
			-/-	29	26.78			
			+/+	17	13.93			
D1	101	0.3713	+/-	41	47.15	1.716	0.4059	0.4669
			-/-	43	39.92			

Table 4. (continued)

ANATOLIAN TURKS								
***Alu* Insertion**	**N**	**Frequency of the insertion**	**Genotype**	**Observed Number**	**Expected Number**	χ^2	**Observed Heterozygosity**	**Expected Heterozygosity**
			+/+	26	21.41			
FXIIIB	101	0.4604	+/-	41	50.18	3.380	0.4059	0.4969
			-/-	34	29.41			
			+/+	41	42.44			
HS4.32	95	0.6684	+/-	45	42.11	0.448	0.4737	0.4433
			-/-	9	10.45			
			+/+	7	8.76			
HS4.69	96	0.3021	+/-	44	40.48	0.726	0.4583	0.4217
			-/-	45	46.76			
			+/+	1	2.36			
PV92	102	0.1520	+/-	29	26.29	1.088	0.2843	0.2577
			-/-	72	73.35			
			+/+	21	23.29			
TPA25	101	0.4802	+/-	55	50.42	0.833	0.5446	0.4992
			-/-	25	27.29			
Average Heterozygosities							0.3637	0.3752

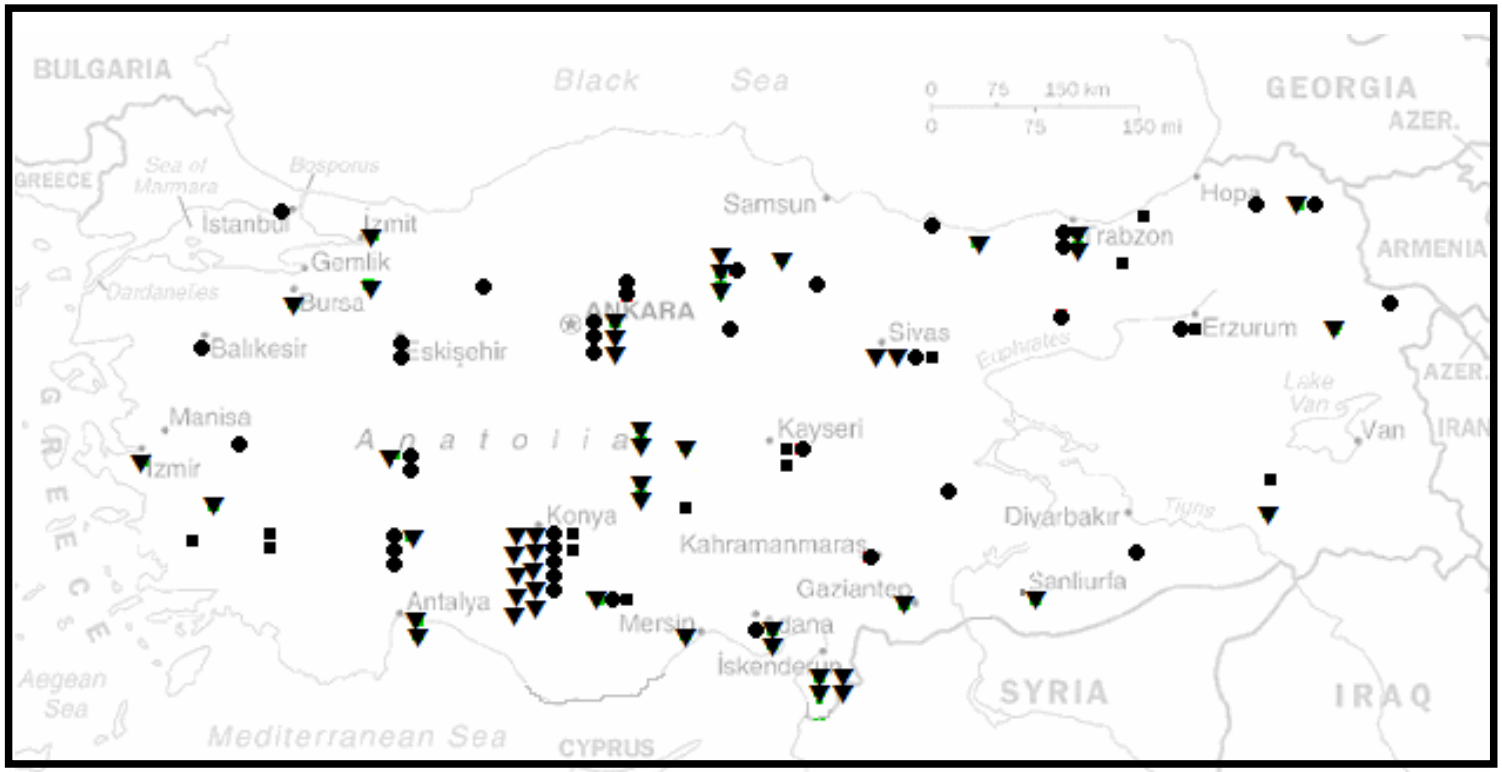


Figure 9. The distribution pattern of ACE insertion in Turkey.
Square : Homozygous individual for the ACE insertion (+/+)
Circle : Heterozygous individual (+/-)
Triangle : Homozygous individual for the lack of ACE insertion (-/-)

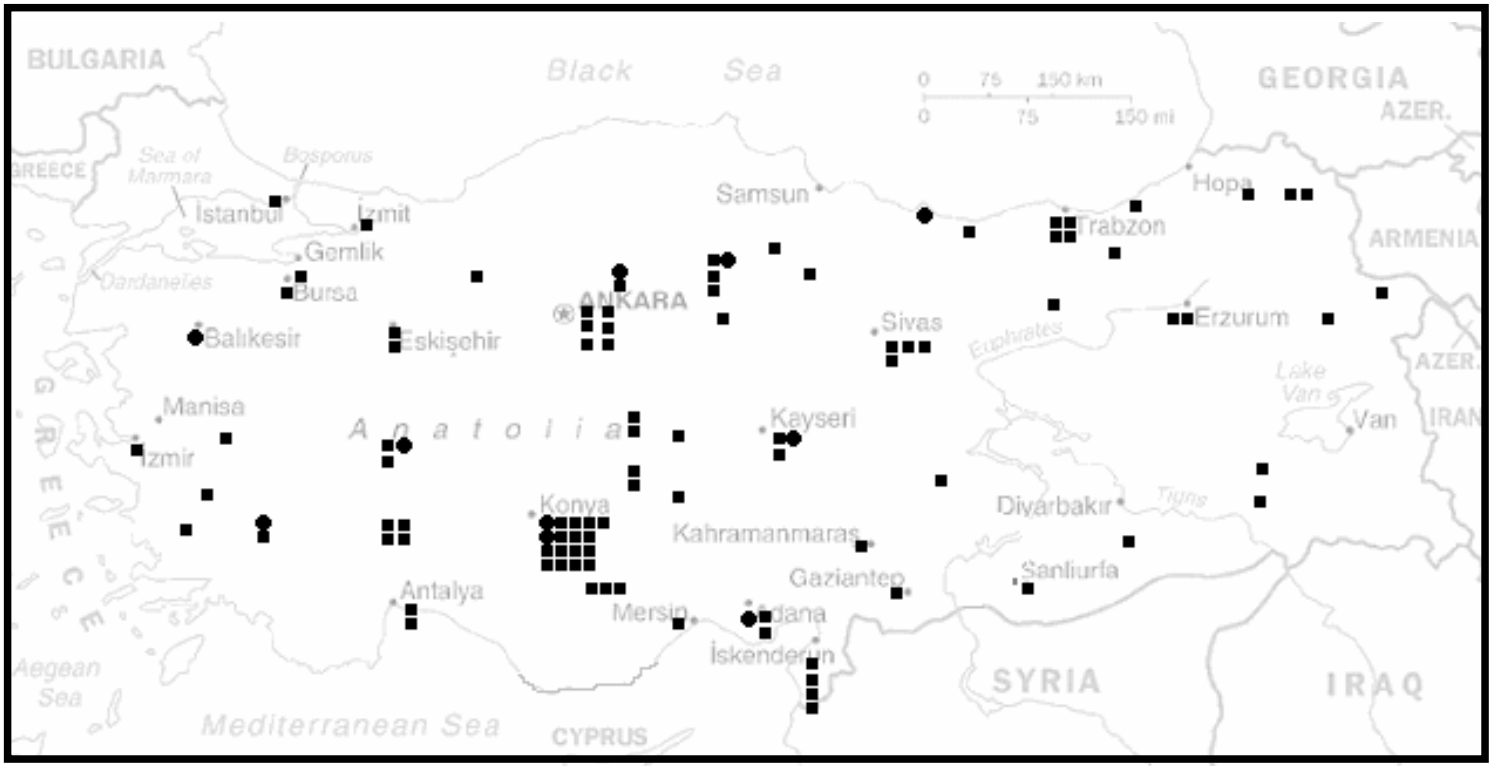


Figure 10. The distribution pattern of APO insertion in Turkey.
Square : Homozygous individual for the APO insertion (+/+)
Circle : Heterozygous individual (+/-)
Triangle : Homozygous individual for the lack of APO insertion (-/-)

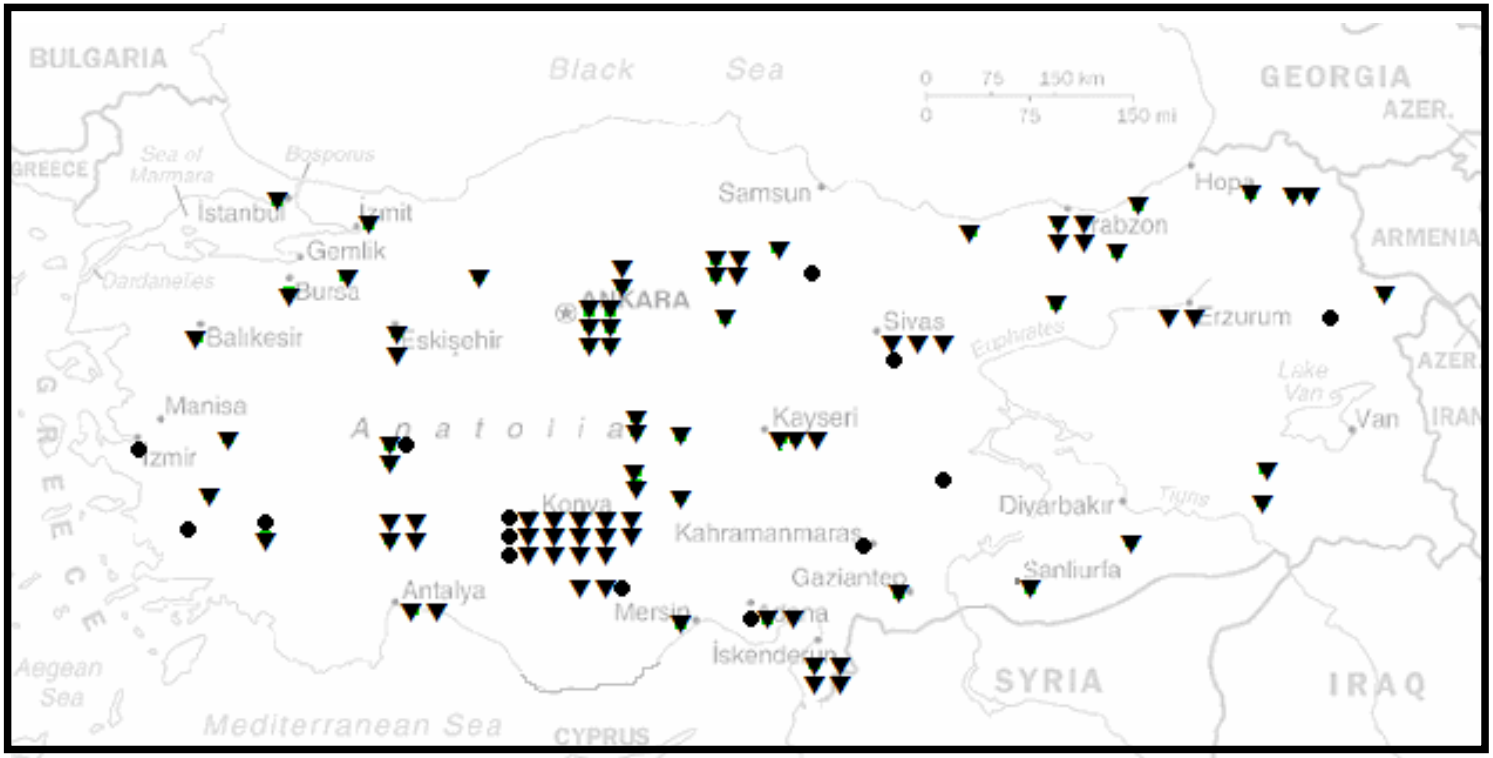


Figure 11. The distribution pattern of A25 insertion in Turkey.
Square : Homozygous individual for the A25 insertion (+/+)
Circle : Heterozygous individual (+/-)
Triangle : Homozygous individual for the lack of A25 insertion (-/-)

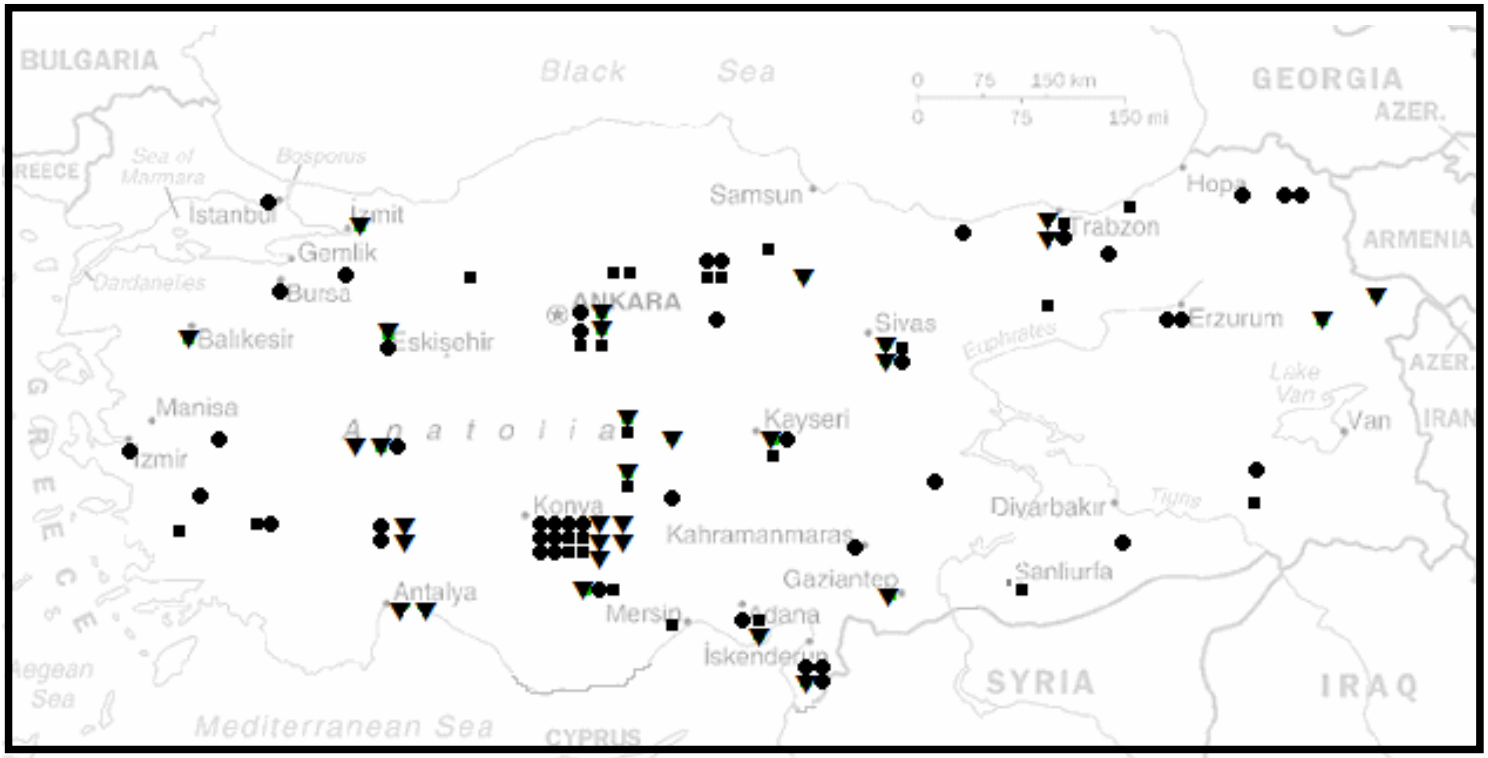


Figure 12. The distribution pattern of B65 insertion in Turkey.
Square : Homozygous individual for the B65 insertion (+/+)
Circle : Heterozygous individual (+/-)
Triangle : Homozygous individual for the lack of B65 insertion (-/-)

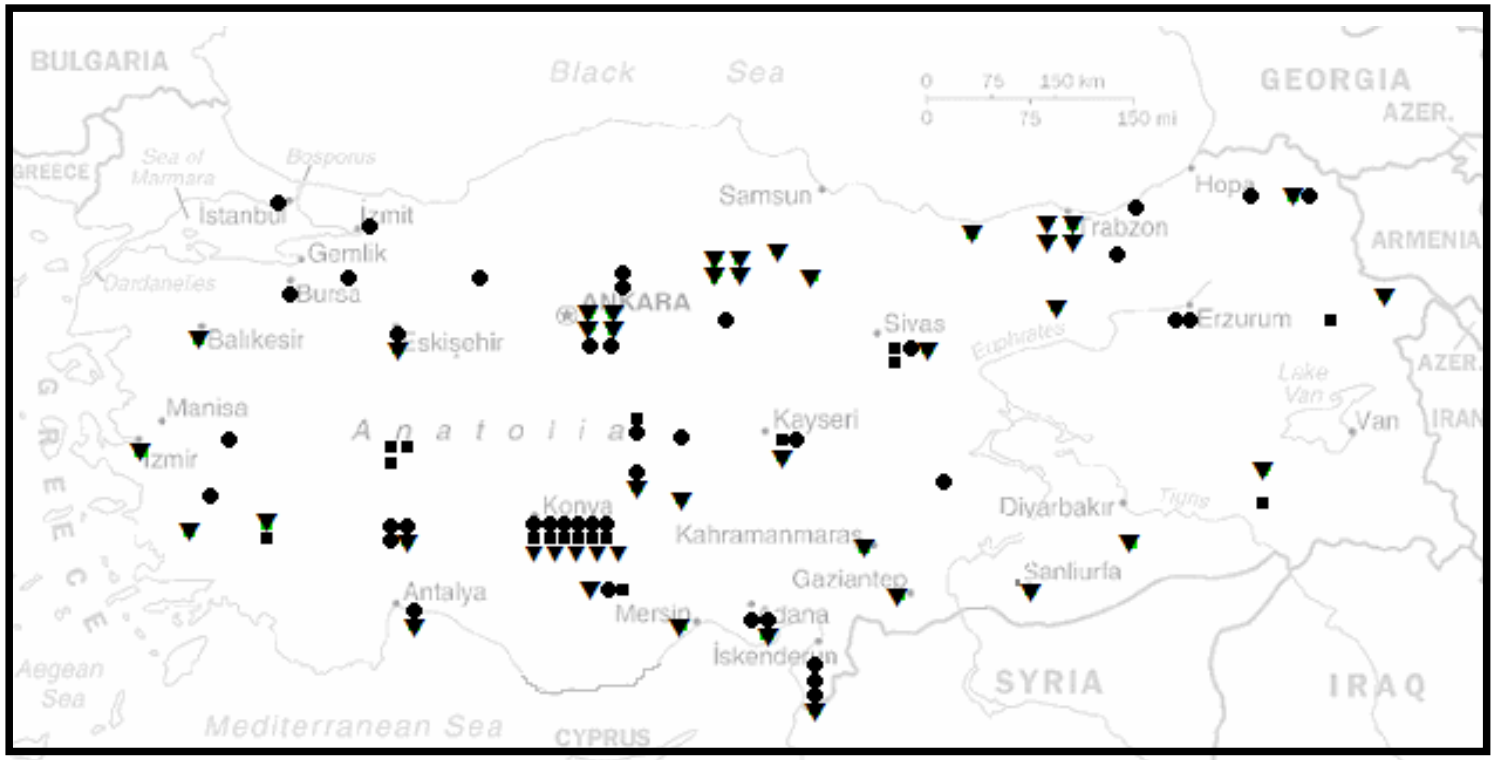


Figure 13. The distribution pattern of D1 insertion in Turkey.
Square : Homozygous individual for the D1 insertion (+/+)
Circle : Heterozygous individual (+/-)
Triangle : Homozygous individual for the lack of D1 insertion (-/-)

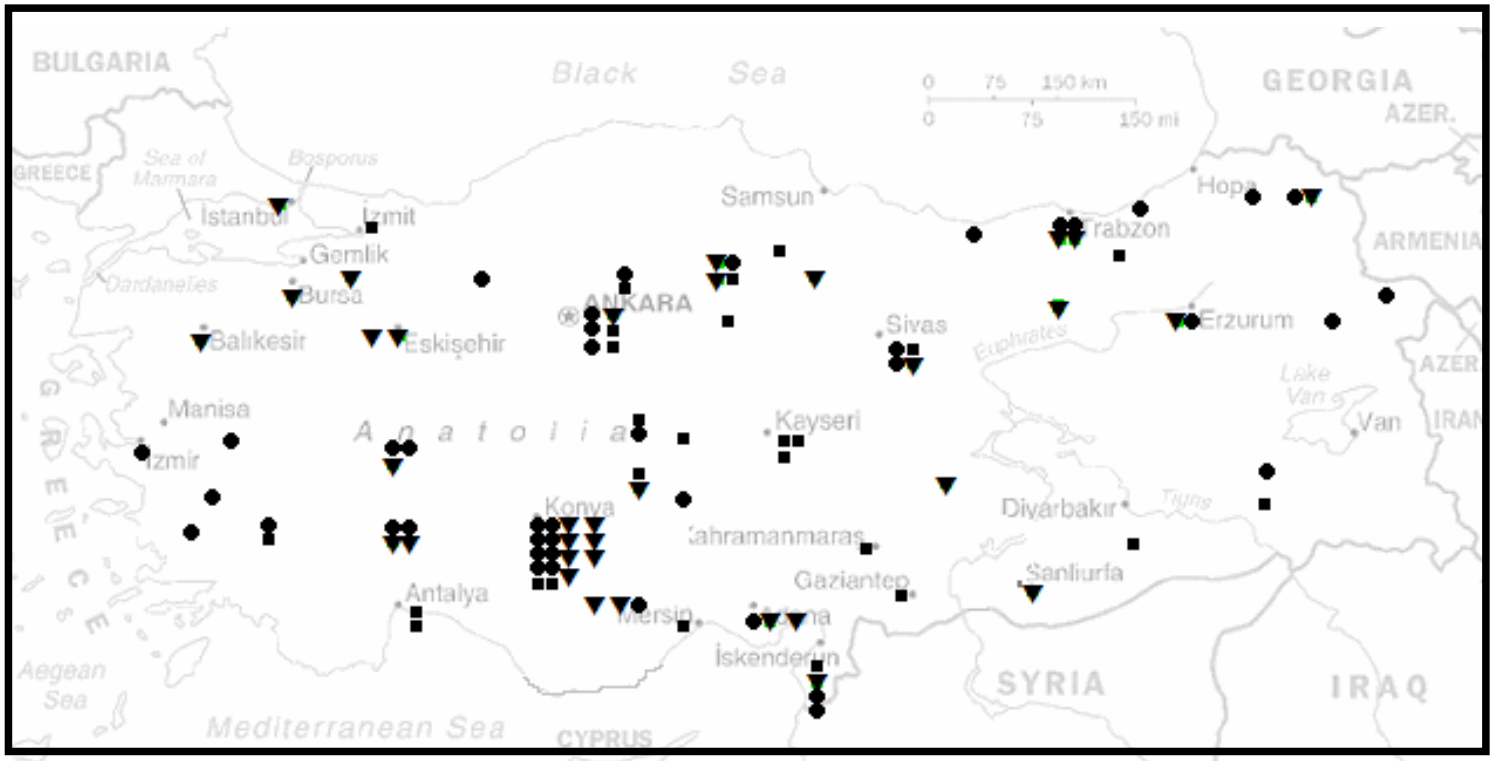


Figure 14. The distribution pattern of FXIIIB insertion in Turkey.
Square : Homozygous individual for the FXIIIB insertion (+/+)
Circle : Heterozygous individual (+/-)
Triangle : Homozygous individual for the lack of FXIIIB insertion (-/-)

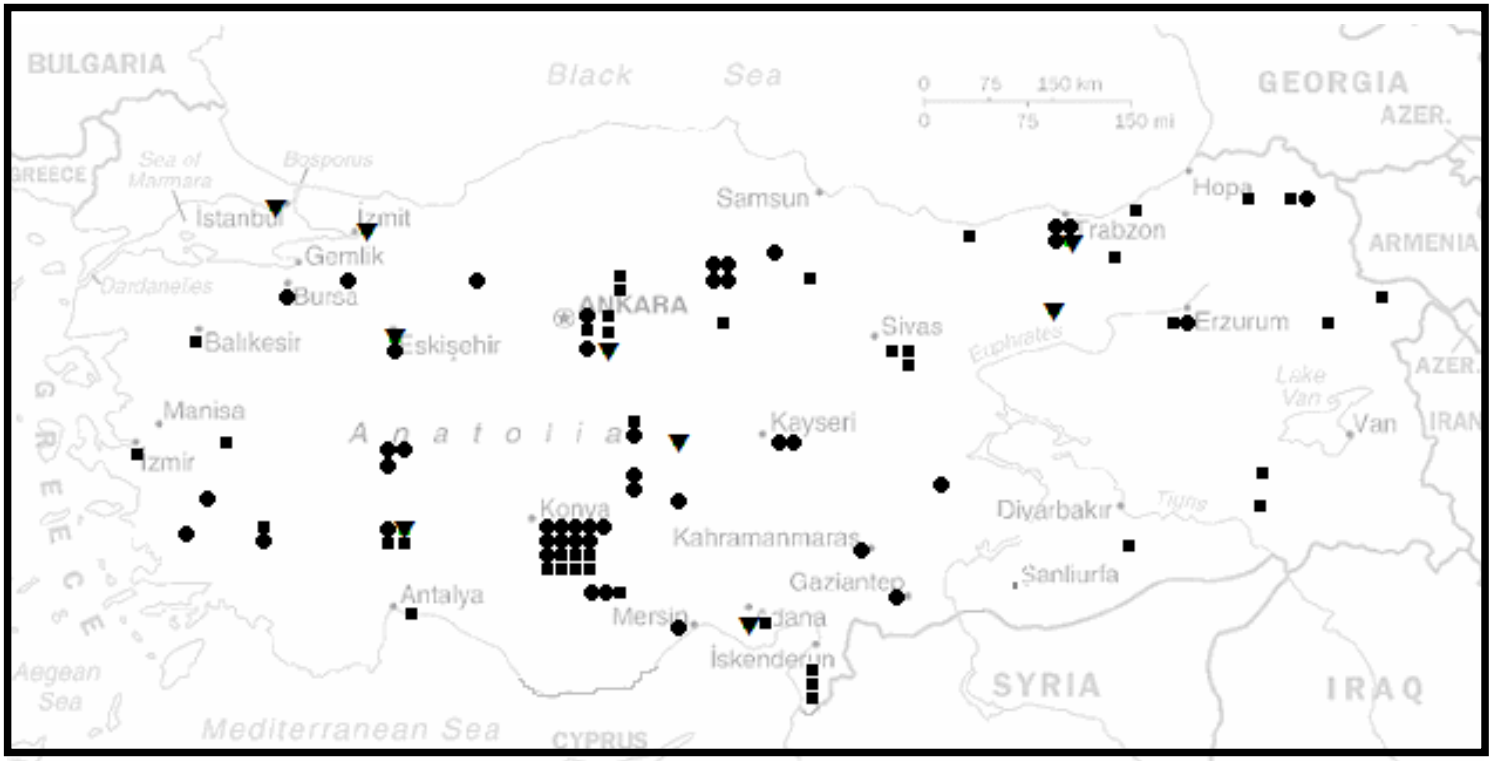


Figure 15. The distribution pattern of HS4.32 insertion in Turkey.
Square : Homozygous individual for the HS4.32 insertion (+/+)
Circle : Heterozygous individual (+/-)
Triangle : Homozygous individual for the lack of HS4.32 insertion (-/-)

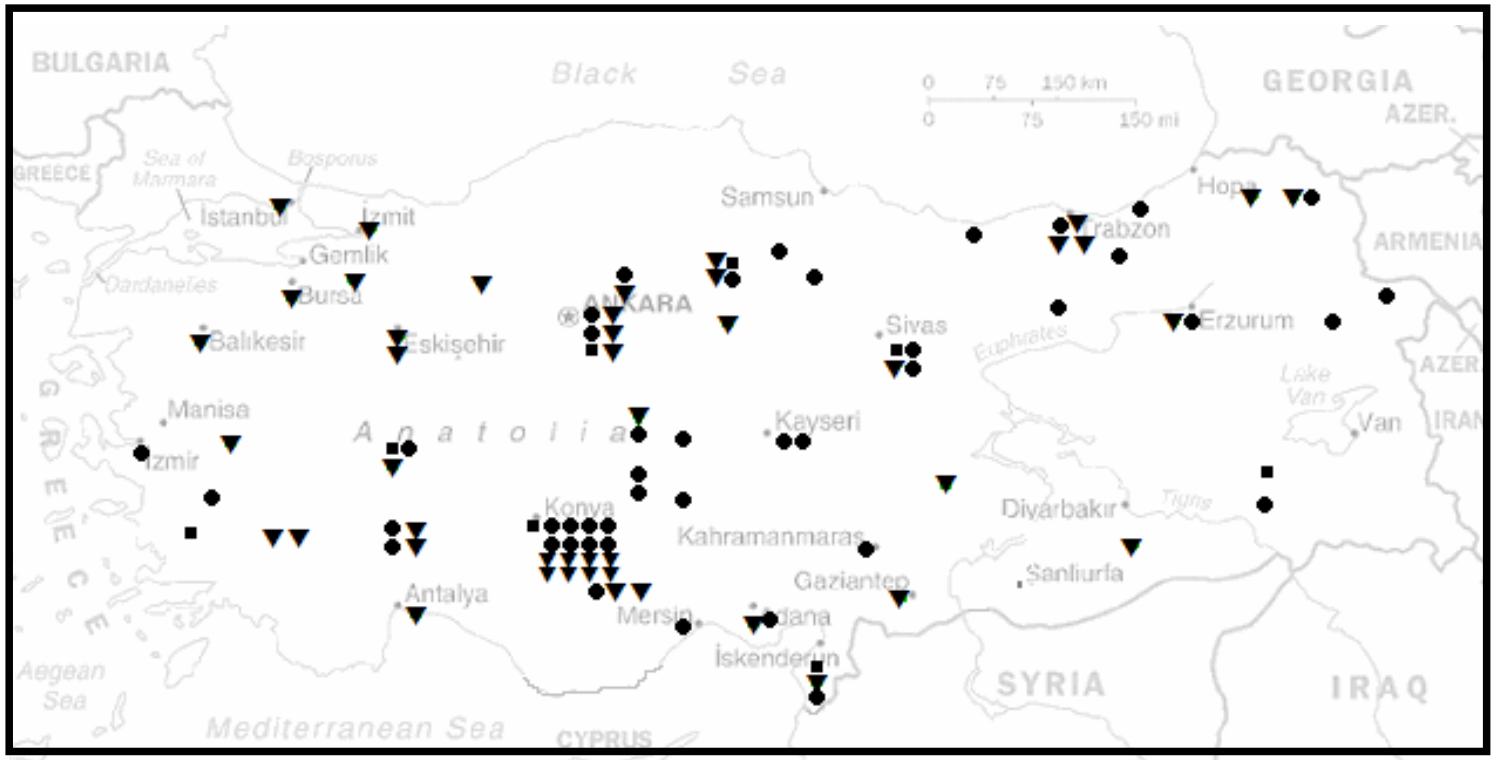


Figure 16. The distribution pattern of HS4.69 insertion in Turkey.
Square : Homozygous individual for the HS4.69 insertion (+/+)
Circle : Heterozygous individual (+/-)
Triangle : Homozygous individual for the lack of HS4.69 insertion (-/-)

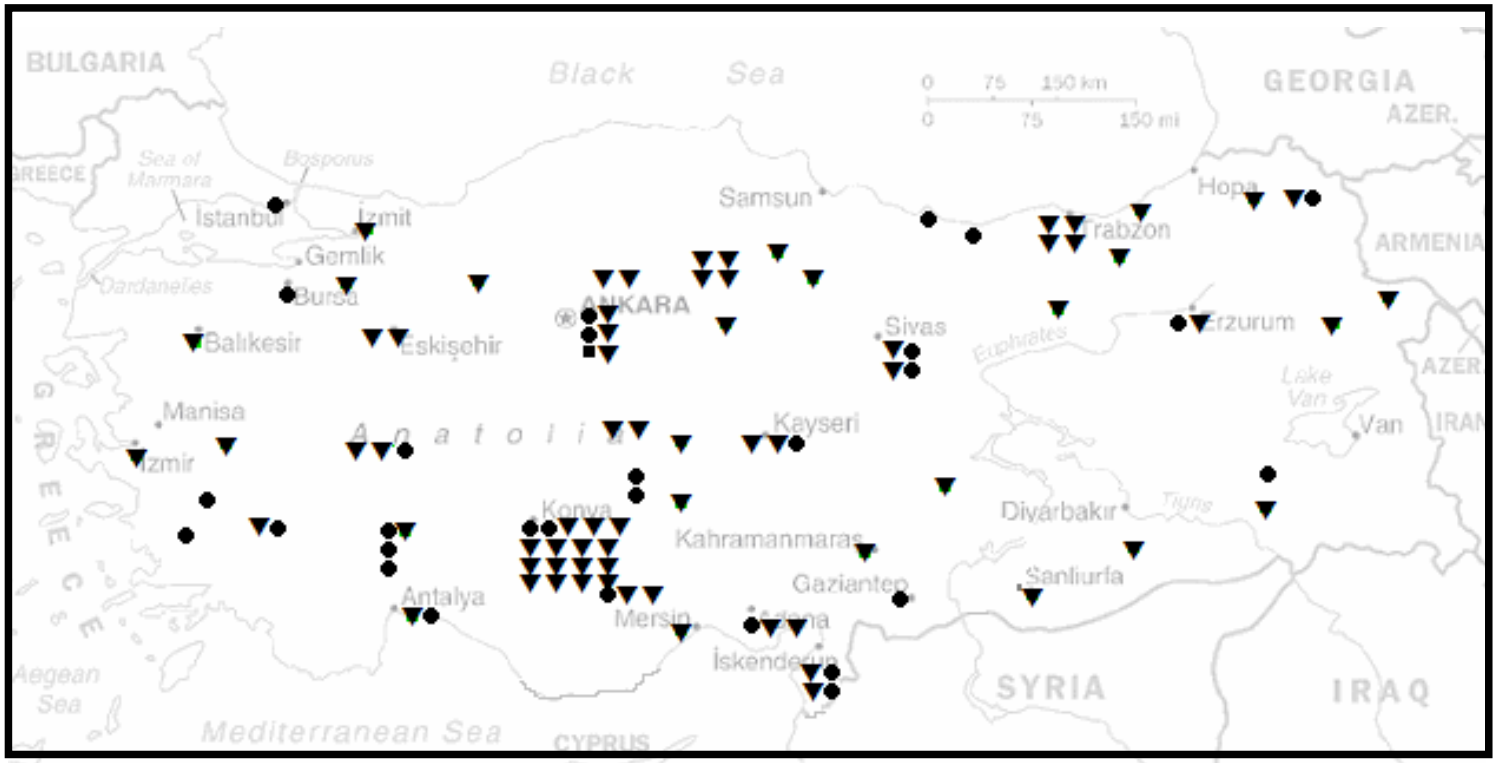


Figure 17. The distribution pattern of PV92 insertion in Turkey.
Square : Homozygous individual for the PV92 insertion (+/+)
Circle : Heterozygous individual (+/-)
Triangle : Homozygous individual for the lack of PV92 insertion (-/-)

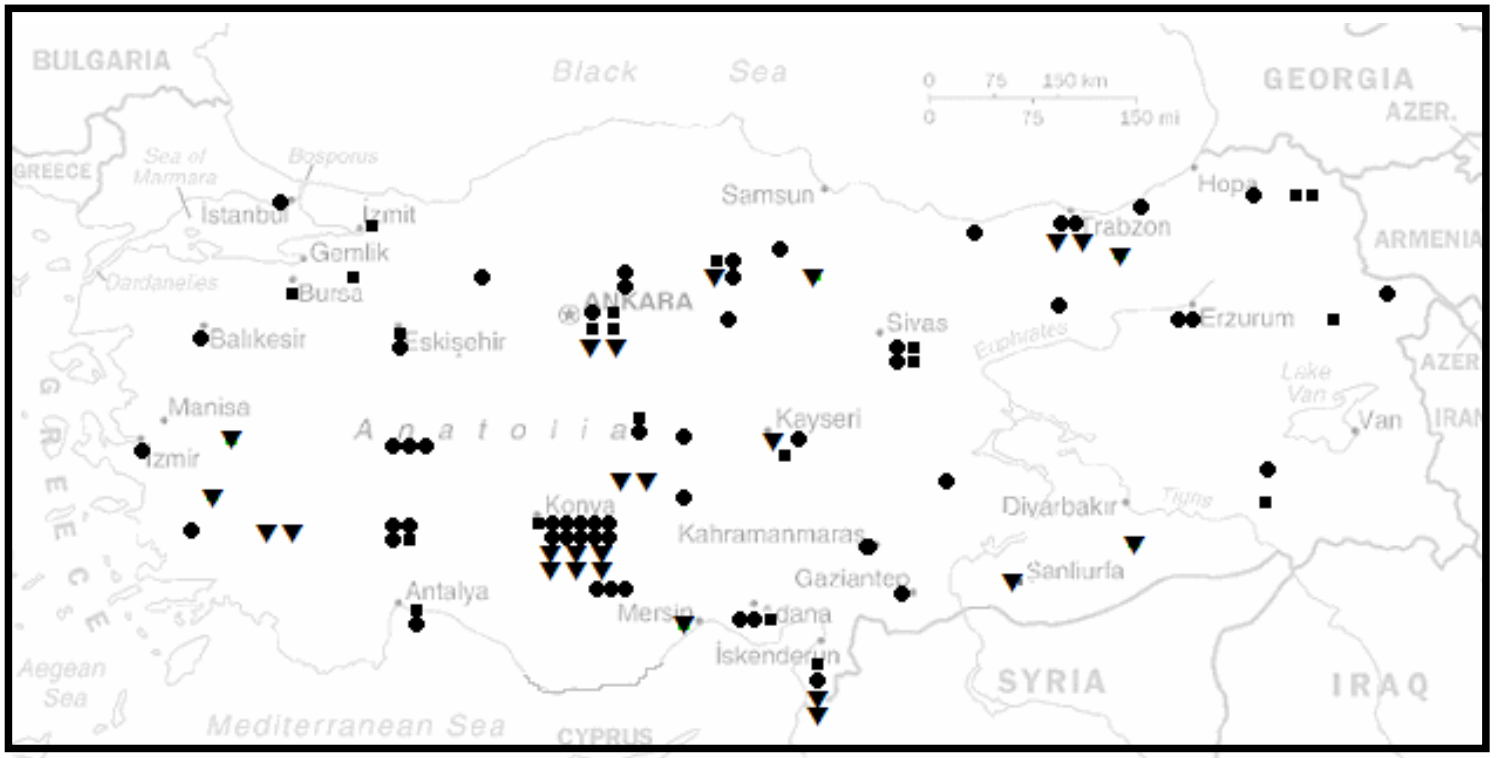


Figure 18. The distribution pattern of TPA25 insertion in Turkey.
Square : Homozygous individual for the TPA25 insertion (+/+)
Circle : Heterozygous individual (+/-)
Triangle : Homozygous individual for the lack of TPA25 insertion (-/-)

3.2 Present Anatolian Data Together with the Data of Previous Studies

Considerable data for the frequency of the ten *Alu* insertions in various populations are present (Stoneking *et al.,* 1997, Nasidze *et al.,* 2001, Romualdi *et al.,* 2002). Present data were compared with two different data sets; the first being the data of Stoneking *et al,* (1997) and Nasidze *et al.* (2001) with the frequencies of eight *Alu* insertion polymorphisms (ACE, APO, A25, B65, D1, FXIIIB, PV92 and TPA25) and the second being the data of Romualdi *et al.* (2002) with the frequencies of ten *Alu* insertion polymorphisms (ACE, APO, A25, B65, D1, FXIIIB, HS4.32, HS4.69, PV92 and TPA25) and both are presented in Appendix C and D, respectively.

3.2.1 Genetic Distances Between Populations

Genetic distances between all possible pairs of populations present in the first and the second data sets were calculated by using frequencies of *Alu* insertions given in Appendix C and D, and a distance matrix was obtained (data not shown). Only the genetic distances between the Anatolian Turks and the populations in the first and the second data sets are presented in Table 5 and 6, respectively. The lowest genetic distances observed in the first data set obtained by eight *Alu* insertions were between Anatolian Turks and Turkish Cypriots, Anatolian Turks and Swiss, Anatolian Turks and Georgians, Anatolian Turks and Azerbaijanians. On the other hand, the lowest genetic distances observed in the second data set obtained by ten *Alu* insertions were between Anatolian Turks and French, Anatolian Turks and Syrians, Anatolian Turks and

Georgian, Anatolian Turks and Bretons. The striking difference in the two data sets is that genetic distances calculated in the second data set are considerably higher than that of in the first data set. This can easily be noticed since the lowest genetic distances are 0.0007 and 0.0095 in the first and second data sets, respectively.

Table 5. Genetic distances between Anatolian Turks and other worldwide populations from the first data set with eight *Alu* insertion polymorphisms.

GENETIC DISTANCES			
Populations	**Anatolians**	**Populations**	**Anatolians**
Armenians	0.0543	!Kung	0.1138
Azerbaijanians	**0.0049**	Malaysian	0.0315
Cherkessians	0.0439	Mayan	0.0400
Darginians	0.0063	Moluccan	0.0573
Georgians	**0.0029**	Mvskoke	0.0457
Ingushians	0.1197	Nguni	0.0399
Alaska Natives	0.0179	Nigerian	0.1296
Australia	0.1847	Pakistan	0.0452
Bretons	0.0068	PNG-Coastal	0.0632
China	0.0665	PNG-Highland	0.1437
European American	0.0165	Pushtoon	0.0229
Filipino	0.0097	Pygmy-CAR	0.0412
French	0.0167	Pygmy-Zaire	0.1195
French Acadian	0.0104	Sotho	0.0283
Greek Cypriot	0.0135	**Swiss**	**0.0021**
Greenland Natives	0.0207	Taiwan	0.0109
India-Christian	0.0266	Tamill	0.0316
India-Hindu	0.0787	Tenggaras	0.0541
India-Muslim	0.0115	**Turkish Cypriot**	**0.0007**
Java	0.0645	UAE	0.0827

Table 6. Genetic distances between Anatolian Turks and other worldwide populations from the second data set with ten *Alu* insertion polymorphisms.

GENETIC DISTANCES			
Populations	**Anatolians**	**Populations**	**Anatolians**
African American	0.0515	Ingushian	0.1469
Armenian	0.0612	Kabardinian	0.0740
Azerbaijani	0.0391	!Kung	0.1544
Bantu Speakers	0.1222	Maya	0.0844
Bretons	**0.0303**	Moluccas	0.0853
Cajun	0.0407	Mvskoke	0.0636
Cherkessian	0.0500	Nguni	0.1881
Darginian	0.0368	Nusa Tenggarans	0.0955
European American	0.0559	PNG Coastal	0.1245
French	**0.0095**	PNG Highland	0.1751
Georgian	**0.0260**	Swiss	0.0499
German	0.0395	**Syrians**	**0.0175**
Greek Cypriot	0.0549	Turk Cypriot	0.0399
Hispanic American	0.0412	Yanomamo	0.1284
Hungarian	0.0495		

3.2.2 Neighbor Joining (NJ) Tree

Neighbor joining (NJ) trees are constructed to examine population relationships. Therefore, Anatolian Turks were compared with worldwide populations in the first and second data sets by constructing NJ trees as shown in Figures 19 and 20, respectively. To root the tree, a hypothetical ancestral population, in which the frequency of the *Alu*

element at each locus was set to zero, was added to the analysis. In both of the trees, obvious groupings of Africans (with the ancestral population) and non-Africans were observed. Apparently, the *Alu* insertion frequencies in African populations have undergone the least amount of change from the ancestral state.

Each population or group was connected to the tree by a branch whose length is proportional to the genetic distance. The numbers at the branching nodes indicate the bootstrap numbers of 1000 bootstrap replicates. Bootstrap resampling is used to assess the strength of the support of the data for the branching structure of the tree. In both trees bootstrap numbers were low so only bootstrap values larger than 30 were indicated.

Neighbor joining trees in Figures 19 and 20 showed that generally populations which are geographically close to each other are also genetically close to each other. The tree in Figure 19 illustrated that non- Africans formed groupings such as; Australians, East and Southeast Asians and Americans. Some populations (European Americans, French, Bretons and French Acadians) formed European cluster and some other populations (Darginians and Ingushians) formed Caucasian cluster, while other European and Caucasian populations were intermingled. Anatolia was observed to be in this intermingled structure. A grouping for West Asian populations was also detected where only UAE and Pakistan was away from this grouping and this was also observed in the neighbor joining tree of Stoneking *et al.*, (1997).

In the tree in Figure 20, a better visualization of groupings of the worldwide populations was expected since two more *Alu* insertions were added to the analysis. However, this was not the case and clustering of populations were vaguer than the previous tree but with relatively higher bootstraps values. There were grouping of Americans; Maya and Yanomamo with the highest bootstrap value of 90.3 %. Small groupings of Caucasians (Cherkessians-Georgians and Darginians-Azerbaijanians), Europeans (Hungarians-Turkish Cypriots) and Australians (PNG Coastland-PNG Highland) were also observed.

Reasons for this surprising discrepancy between two trees can be that the first and second data sets are from different studies; therefore, different populations are compared with Anatolians in each case and sample sizes for each population are also different and relatively low (41 and 50) in each data set.

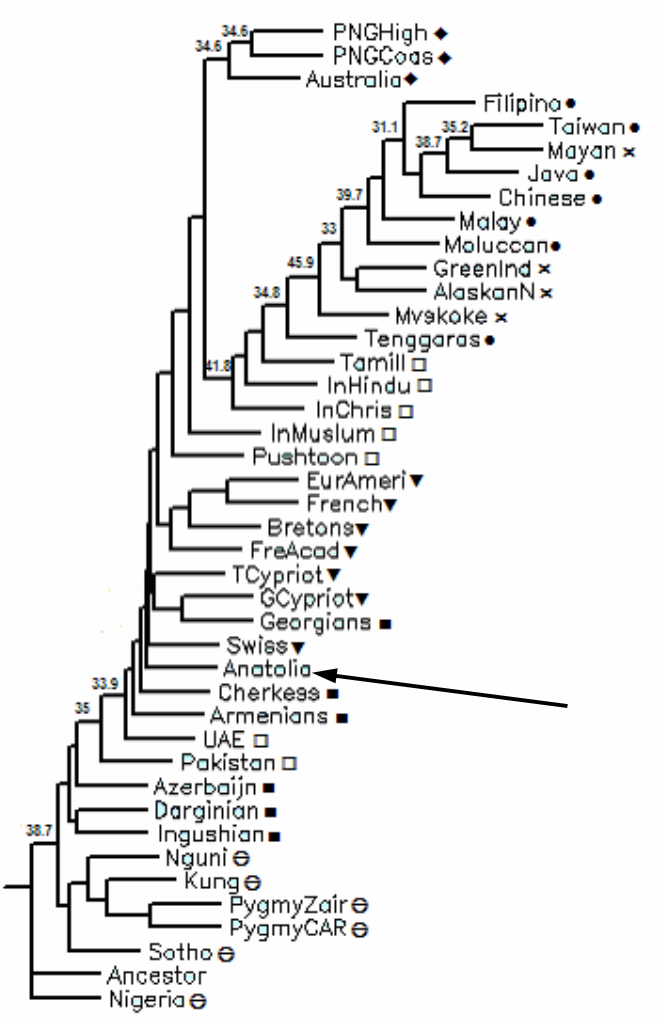


Figure 19. Neighbor joining tree constructed by using the first data set with 41 populations analyzed for eight *Alu* insertions. The arrow points to Anatolia and the symbols used on the tree indicate the following groups of populations:

⊖: Africans
◆: Australians
✕: Americans
■: Caucasians
●: East and Southeast Asians
▼: Europeans
☐: West Asians

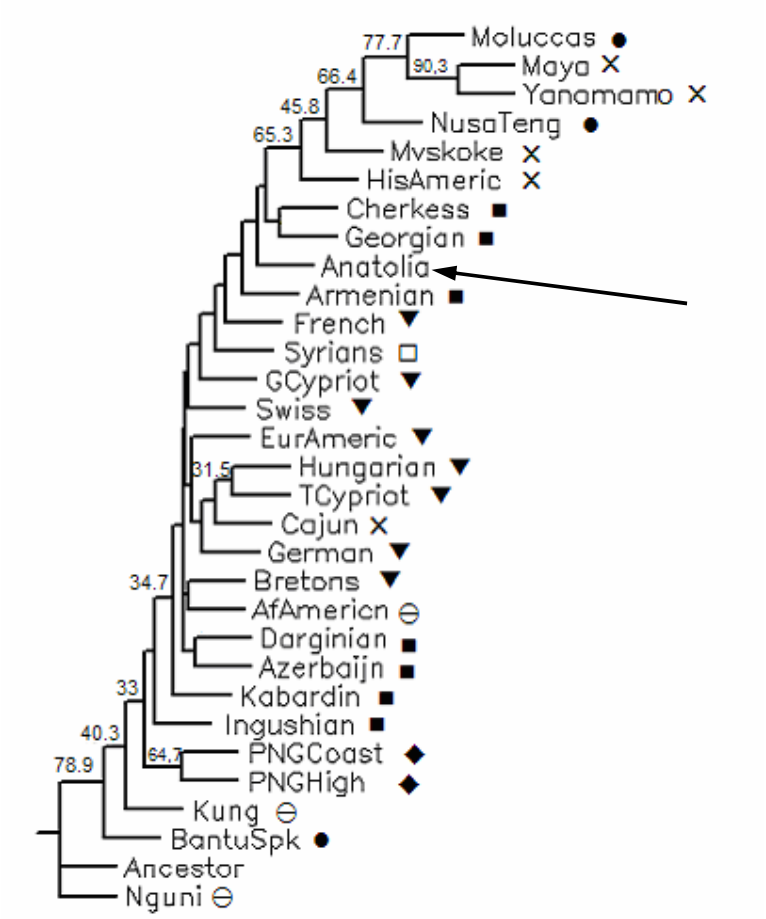


Figure 20. Neighbor joining tree constructed by using the second data set with 30 populations analyzed for ten *Alu* insertions. The arrow points to Anatolia and the symbols used on the tree indicate the following groups of populations:

⊖: Africans
✕: Americans
●: Southeast Asians
□: West Asians
◆: Australians
■: Caucasians
▼: Europeans

3.2.3 Principal Component Analysis (PCA)

With principal component analysis, relative positions of the populations based on *Alu* insertion frequencies were observed in the space generated by the 3 principal component axes. Figure 21 and 22 show the first two and three principal components of the PCA constructed with the first data set based on the frequencies of eight *Alu* insertion polymorphisms, respectively. Additionally, Figure 23 and 24 show the first two and three principal components of the PCA constructed with the second data set based on the frequencies of ten *Alu* insertion polymorphisms, respectively. Groups of populations were delimited arbitrarily for better visualization of the clusters of populations in the Figures 21 and 23.

The equations of the components for the first data set's (8 *Alu*) PCA are as follows;

PC1 = - 0.745 ACE - 0.269 APO + 0.502 A25 + 0.494 B65 - 0.006 D1 - 0.913 FXIIIB - 0.783 PV92 - 0.473 TPA25

PC2 = - 0.396 ACE + 0.593 APO + 0.175 A25 + 0.650 B65 + 0.846 D1 + 0.105 FXIIIB + 0.119 PV92 + 0.603 TPA25

PC3 = 0.308 ACE - 0.375 APO + 0.724 A25 + 0.178 B65 + 0.267 D1 + 0.145 FXIIIB + 0.407 PV92 - 0.323 TPA25

The first principal component (PC1) covers 35.12 % of the overall variation while 25.39 % and 14.65 % of the total variation were covered by the second (PC2) and the third (PC3) principal components, respectively. So, 75.16 % of the total variation is explained with three components of the first principal component analysis. It can be deduced from the weightings of the variables (in the equation of the first component) that; on the first axis FXIIIB, PV92 and ACE contributed the most and almost equally to the differentiation of the populations, whereas D1 and ACE contributed the least. As can be seen from Figures 21 and 22, the first axis separated African populations and they have occupied one end of the axis. East and Southeast Asians are the most distant ones to Africans and this result is parallel with the result of the neighbor joining tree constructed with the first data set. This clustering is basically due to the differences in the frequencies of FXIIIB, PV92 and ACE. Moreover, it is observed that in the second axis D1, B65 and TPA25 have the most contribution in differentiating Australians from the rest of the populations. Lastly, the third axis was weighed mostly by A25, and it covers a small portion (14.65 %) of the total variation and does not present distinct clusters.

The equations of the components for the second data set's (10 *Alu*) PCA are as follows;

PC1 = 0.817 ACE - 0.037 APO - 0.200 A25 - 0.405 B65 - 0.021 D1 + 0.856 FXIIIB - 0.571 HS4.32 + 0.117 HS4.69 + 0.890 PV92 + 0.514 TPA25

PC2 = - 0.225 ACE + 0.744 APO - 0.213 A25 + 0.577 B65 + 0.663 D1 + 0.253 FXIIIB + 0.551 HS4.32 + 0.379 HS4.69 + 0.073 PV92 + 0.765 TPA25

PC3 = 0.368 ACE - 0.150 APO - 0.637 A25 - 0.291 B65 - 0.316 D1 - 0.288 FXIIIB + 0.309 HS4.32 + 0.827 HS4.69 - 0.323 PV92 + 0.016 TPA25

The first principal component (PC1) accounts for 29.54 % of the total variation, while the second (PC2) and the third (PC3) principal components account for 25.89 % and 15.69 % of the overall variation, respectively. Therefore, 71.12 % of the entire variation is explained with three components of the second principal component analysis. It can be inferred from the weightings of the variables (in the equation of the first component) that; on the first axis PV92, FXIIIB and ACE contributed the most and almost equally to the differentiation of the populations, whereas D1 and APO contributed the least. As can be seen from Figures 23 and 24, the first axis separated Americans, which occupied one end of the axis. In addition, it is observed that in the second axis TPA25, D1 and APO have the most contribution in differentiating Australians from the rest of the populations. Lastly, the third axis was weighed mostly by HS4.69 and A25 and since it covers a small portion (15.69 %) of the total variation, it does not show definite clusters. In none of the axes were B65 and HS4.32 the variables with the highest power of differentiation of the populations.

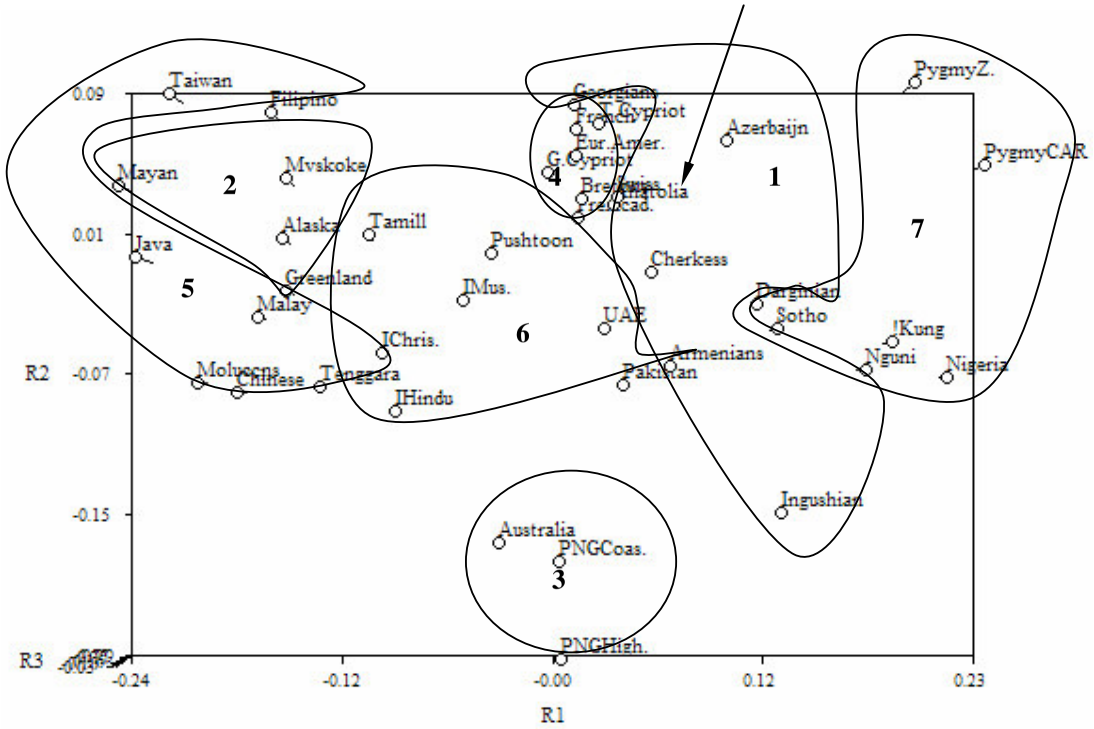


Figure 21. Plot of the first two principal components of the allele frequencies at the eight polymorphic *Alu* insertion loci in 41 populations, where the numbers indicate the name of the following groups: 1- Caucasians, 2- Americans, 3- Australians, 4- Europeans, 5- East and Southeast Asians, 6- West Asians and 7- Africans. The arrow points to Anatolia.

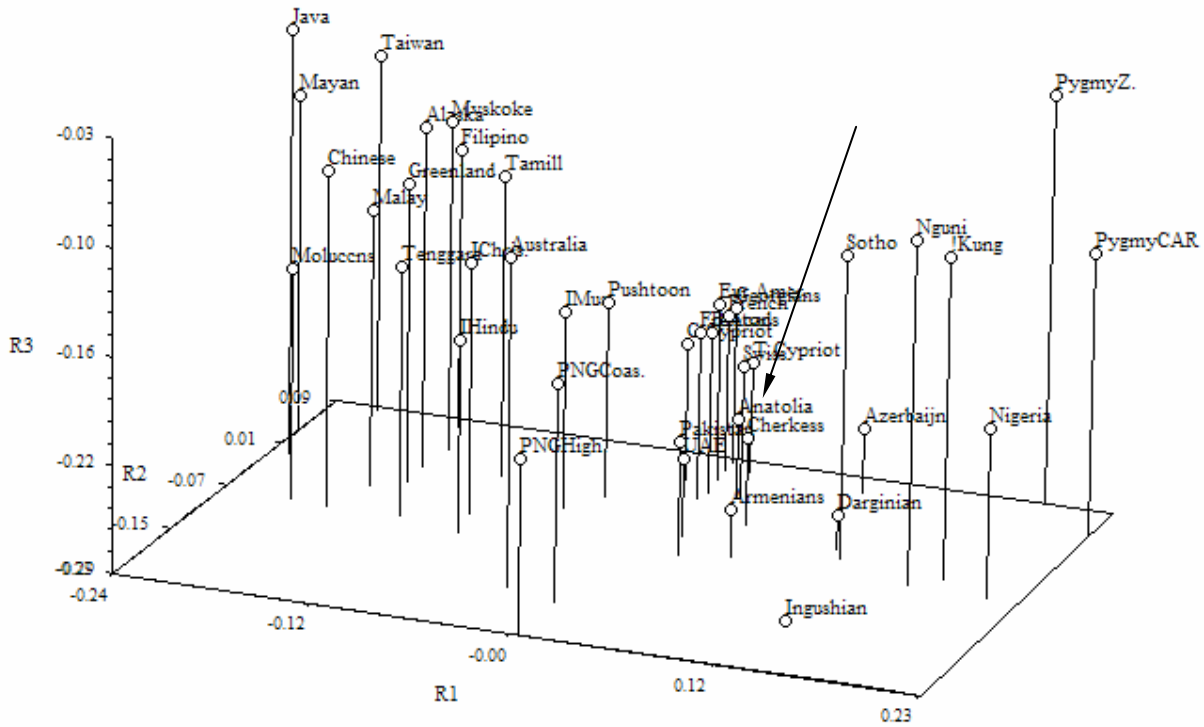


Figure 22. Plot of the three principal components of the allele frequencies at the eight polymorphic *Alu* insertion loci in 41 populations. The arrow points to Anatolia.

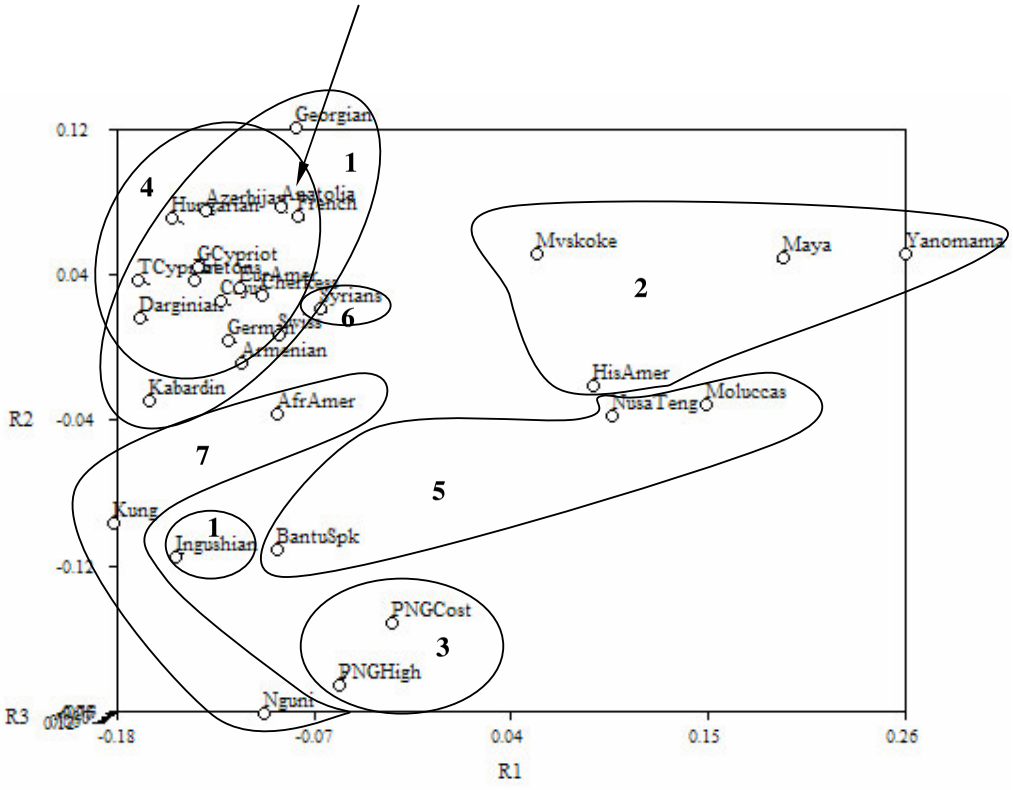


Figure 23. Plot of the first two principal components of the allele frequencies at the ten polymorphic *Alu* insertion loci in 30 populations, where the numbers indicate the name of the following groups: 1- Caucasians, 2- Americans, 3- Australians, 4- Europeans, 5- East and Southeast Asians, 6- West Asians and 7-Africans. The arrow points to Anatolia.

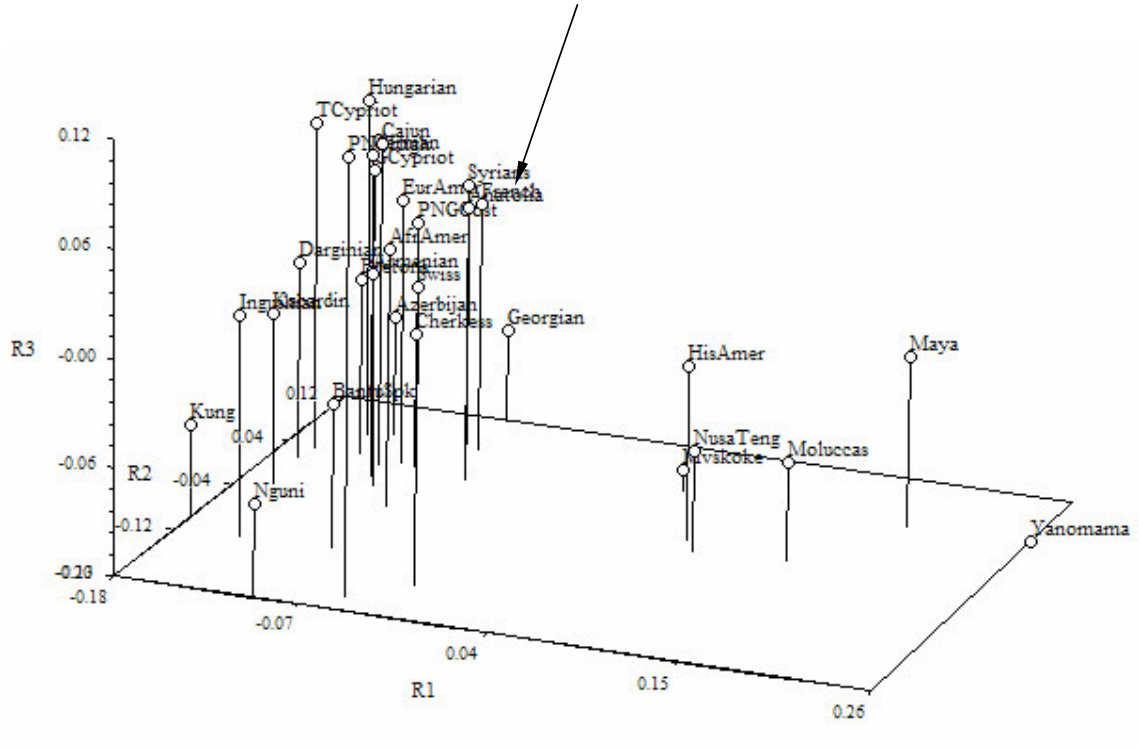


Figure 24. Plot of the three principal components of the allele frequencies at the ten polymorphic *Alu* insertion loci in 30 populations. The arrow points to Anatolia.

3.2.4 Discriminant Function Analysis

Canonical variates analysis (CVA), which is the multigroup discriminant analysis, was applied to the first data set with such groups of worldwide populations; 1- Caucasians, 2- Americans, 3- Australians, 4- Europeans, 5- East and Southeast Asians, 6- West Asians and 7- Africans as described in Appendix C. Figure 25 presents the plot with two axes, showing the relations of the groups of populations and variables functioning in the differentiation of the groups. Each dashed line represents one variable: 1- ACE, 2- APO, 3- PV92, 4- TPA25, 5- D1, 6- FXIIIB, 7- B65 and 8- A25. Groups of populations formed clusters such as: Caucasians-Europeans (1-4), West Asians-Australians (6-3), Americans-East and Southeast Asians (2-5). Separation of each cluster was done by the variables shown as dashed lines between the groups of populations (Figure 25).

The first and the second axes explained 75.63 % and 18.05 % of the total variation, respectively. Where, the third axis (not shown) explained only 3.92 % of the total variation. Therefore; with a plot with three axes, 97.60 % of the total variation between the groups of the populations was explained. Discriminant weightings of each variable for each axis were also attained by the result of the canonical variates analysis. These weight values indicate the power of discrimination of each variable, which are given below with the equations;

C.V. 1 = - 0.070 ACE - 0.002 APO + 0.002 A25 + 0 .051 B65 - 0.014 D1 -0.036 FXIIIB - 0.517 PV92 - 0.047 TPA25

C.V. 2 = 0.739 ACE + 0.083 APO - 0.066 A25 + 0.686 B65 + 0.034 D1 - 0.026 FXIIIB - 0.093 PV92 + 0.035 TPA25

C.V. 3 = 0.079 ACE + 0.020 APO - 0.037 A25 + 0.586 B65 - 0.023 D1 - 0.033 FXIIIB - 0.011 PV92 - 0.047 TPA25

The most variation was explained by PV92 and ACE, in the first and the second canonical variates, respectively. B65 explained the most of the variation in the third canonical variate.

This analysis was also applied to the second data set with such groups of worldwide populations; 1- Caucasians, 2- Americans, 3- Australians, 4- Europeans, 5- East and Southeast Asians and 6- Africans as described in Appendix D. Figure 26 gives the diagram with the positions and interactions of the population groups and the variables functioning in the differentiation of the groups. Each dashed line represents one variable: 1- TPA25, 2- ACE, 3- APO, 4- FXIIIB, 5- PV92, 6- D1, 7- B65, 8- A25, 9- HS4.32 and 10- HS4.69. Groups of populations formed only one cluster; Africans-East and Southeast Asians (6-5). Separation of the cluster with other groups of populations was done by the variables shown as dashed lines between the groups of populations (Figure 26).

The first and the second axes explained 44.21 % and 38.99 % of the total variation, respectively. Where, the third axis (not shown) explained only 8.58 % of the total variation. Therefore, with a diagram with three axes, 91.78 % of the total variation between the groups of the populations was explained. Discriminat weightings, which

designate the power of discrimination of each variable, for each of ten variables on each axis, are such as:

C.V. 1 = - 0.033 ACE + 0.001 APO + 0.091 A25 - 0.027 B65 + 0.951 D1 -0.867 FXIIIB + 0.019 HS4.32 - 0.053 HS4.69 - 0.041 PV92 + 0.001 TPA25

C.V. 2 = - 0.088 ACE - 0.097 APO - 0.077 A25 + 0.691 B65 - 0.041 D1 + 0.019 FXIIIB + 0 .871 HS4.32 - 0.099 HS4.69 - 0.056 PV92 + 0.018 TPA25

C.V. 3 = - 0.079 ACE - 0.074 APO + 0.001 A25 - 0.159 B65 - 0.016 D1 + 0.013 FXIIIB -0.024 HS4.32 + 0.017 HS4.69 + 0.034 PV92 + 0.062 TPA25

The most variation was explained by D1 and HS4.32, in the first and the second canonical variates, respectively. B65 explained most of the variation in the third canonical variate.

3.2.5 The Fst Analysis

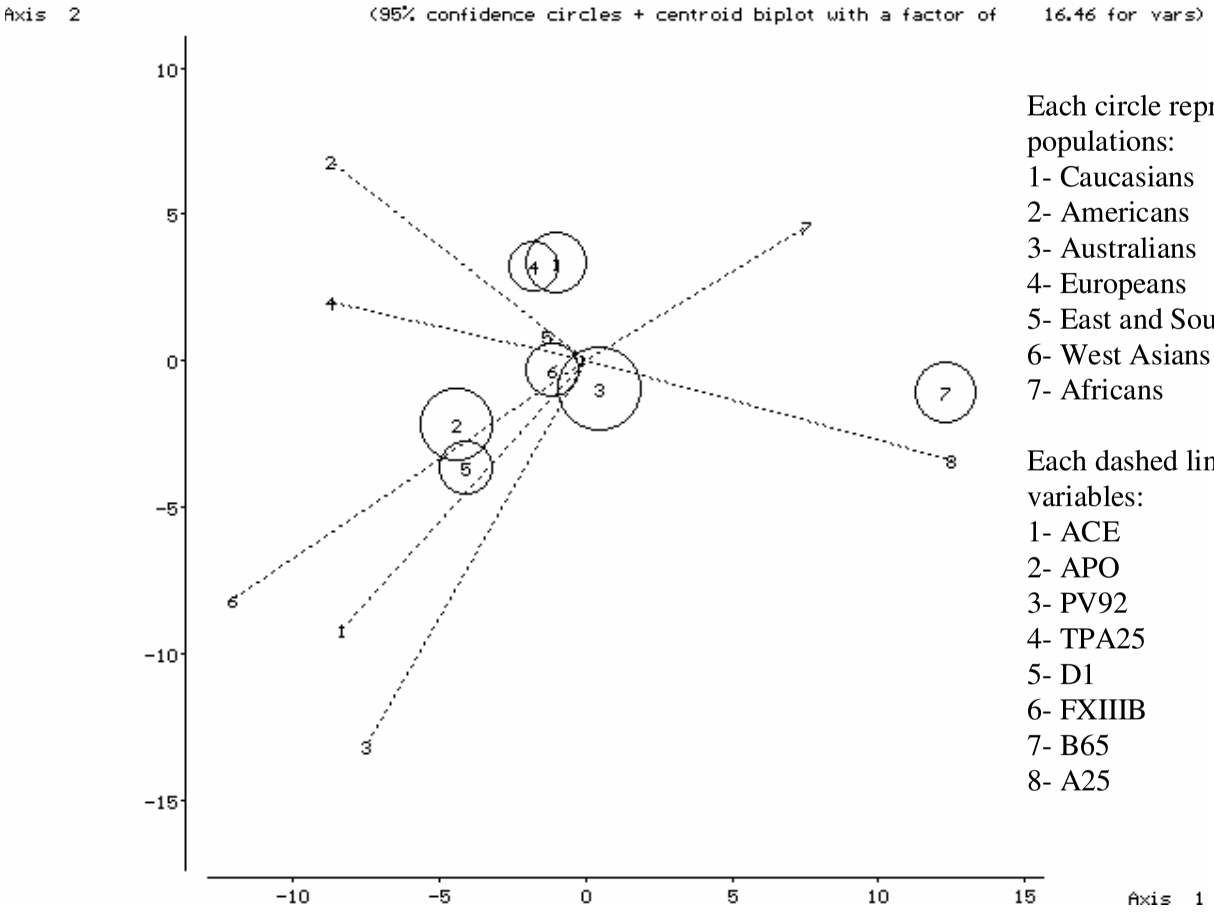
The degree of relationship between Anatolian Turks and previously studied populations from the first data set was determined. Those populations which are geographically close to Anatolia were considered among the available data and those which are genetically close to Anatolia were determined by calculating Fst genetic distances based on eight *Alu* insertion polymorphisms (Table 9). The range of the pair-wise Fst values was -0.00279 – 0.14855. Pair-wise Fst values between Anatolians-Swiss and Anatolians-French Acadian are not statistically significant indicating that there is no

significant genetic distance difference between Swiss and French Acadian and Anatolians. The significance level for each comparison was given in the Table 7.

Table 7. Fst genetic distances between Anatolians and populations from the first data set based on eight *Alu* insertion frequencies.

POPULATIONS	Fst VALUES
Anatolians – Swiss	-0.00279 (NS)
Anatolians - French Acadian	0.00660 (NS)
Anatolians – Bretons	0.01194*
Anatolians - French	0.01238*
Anatolians - European American	0.01395*
Anatolians - Turkish Cypriot	0.01521*
Anatolians - Greek Cypriot	0.01850**
Anatolians - Cherkessians	0.01994*
Anatolians - Georgians	0.02427**
Anatolians - Pushtoon	0.02865**
Anatolians - Armenians	0.03012**
Anatolians - UAE	0.03644***
Anatolians - Darginians	0.04565**
Anatolians - Pakistan	0.05503***
Anatolians - Azerbaijanians	0.07629***
Anatolians - Ingushians	0.14855***

(NS: non-significant, *P<0.05, **P<0.01, ***P<0.001)



Each circle represents the following groups of populations:
1- Caucasians
2- Americans
3- Australians
4- Europeans
5- East and Southeast Asians
6- West Asians
7- Africans

Each dashed line represents the following variables:
1- ACE
2- APO
3- PV92
4- TPA25
5- D1
6- FXIIIB
7- B65
8- A25

Figure 25. Plot of canonical function analysis applied to the first data set with 7 groups of populations.

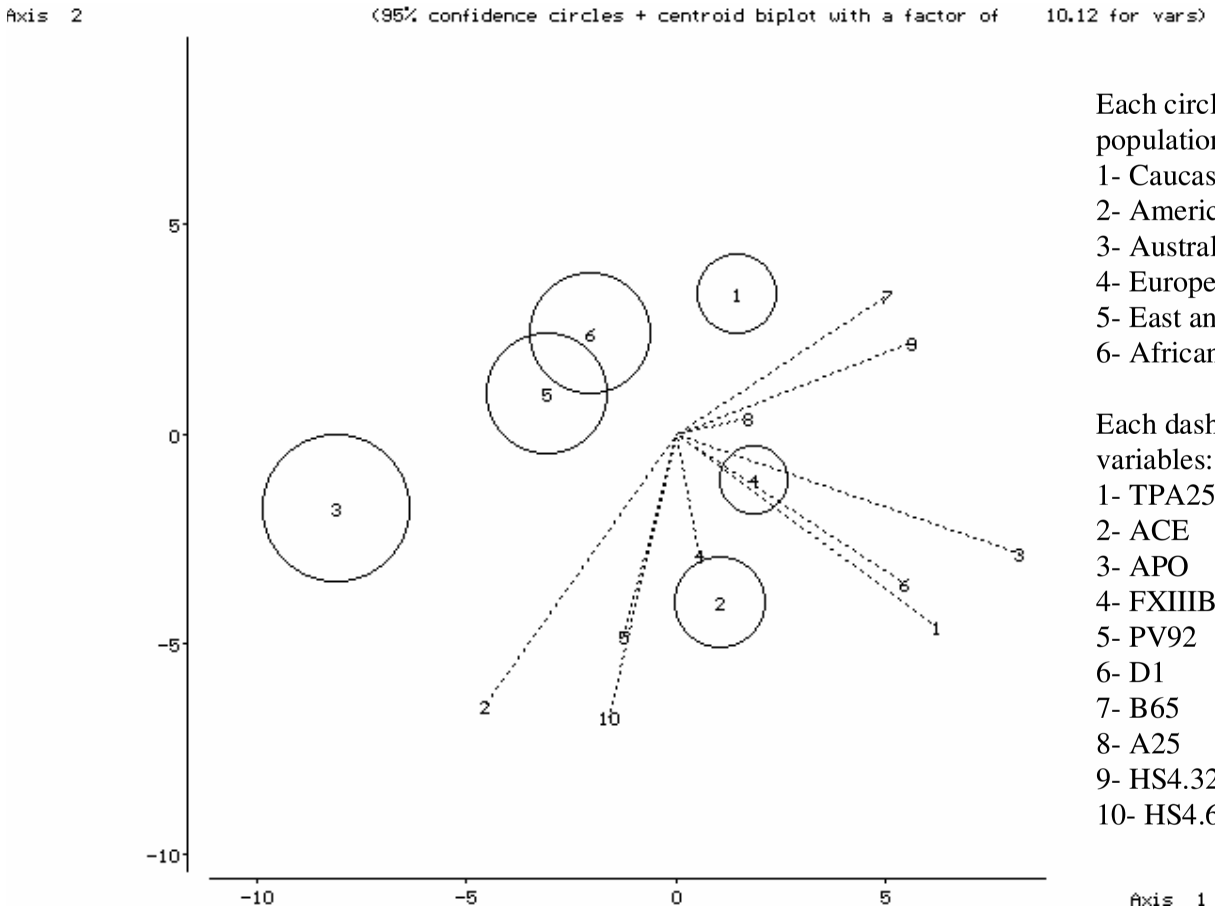


Figure 26. Plot of canonical function analysis applied to the second data set with 6 groups of populations.

3.2.6 The Plot of Heterozygosity *vs.* Distance from Centroid

According to Harpending and Ward (1982), in a structured population with discrete subpopulations, which exchange genes at a regular rate, a simple linear relationship is expected between the mean heterozygosity of each subpopulation and the genetic distance from the centroid (the arithmetic mean of the allele frequencies). In this study, the heterozygosity *vs.* distance from centroid graph was plotted for 41 populations including Anatolia (Figure 27). Major deviations from the expected relationship were only observed in African populations due to their greater heterozygosity values than predicted by the model. This deviation was attributed to the larger effective population sizes than non-African populations, which was also pointed out in Stoneking *et al.* (1997). Minor deviations from the expected relationship were observed in UAE, PNG Highland and Caucasian populations; especially Ingushians were the most deviant ones. This result is parallel with the results of Nasidze *et al.,* (2001) which stated that the average heterozygosity for each population was lower in Ingushians than in the other populations in Caucasus populations. The rest of the populations together with the Anatolian Turks represented a good fit between the observed relationship and that predicted by the model. Together with Anatolians, India-Hindu population is also on the line showing the exact expected relationship predicted by the model.

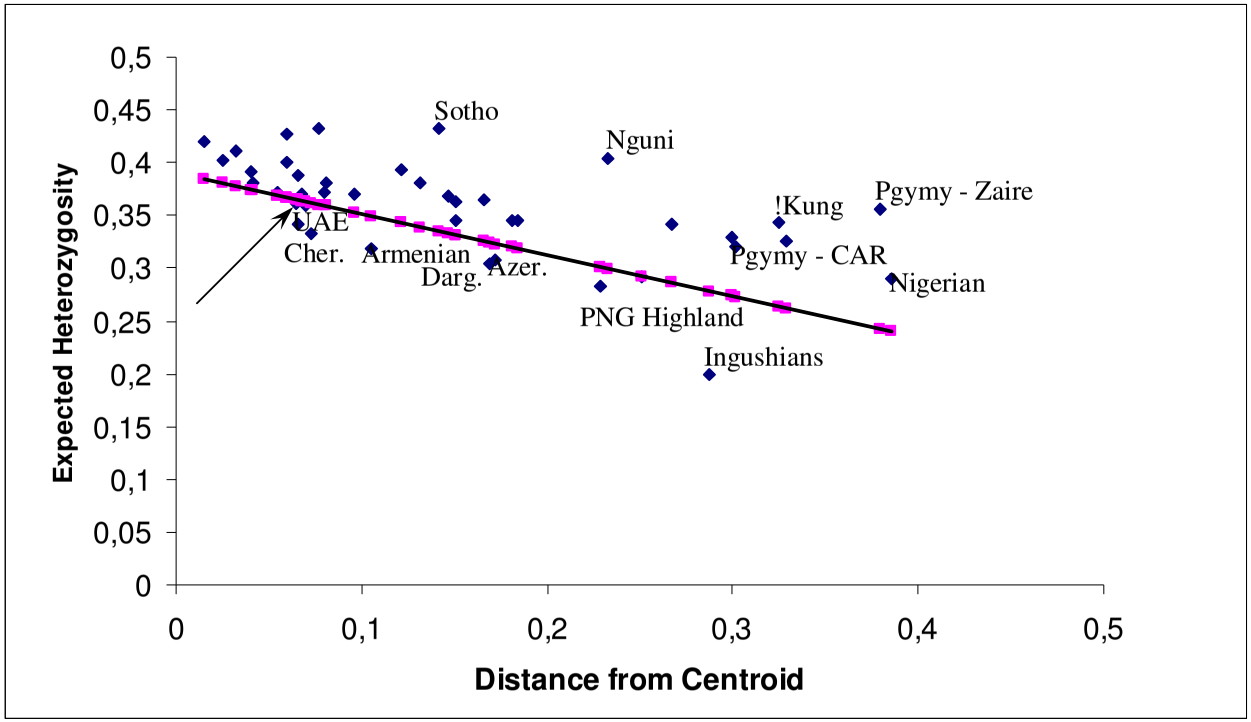


Figure 27. The plot of heterozygosity *vs*. distance from centroid. The squares forming the line show the expected relationship predicted by the model of Harpending and Ward (1982), according to the formula $h_i = H(1 - r_i)$, where r_i is the distance from the centroid, and h_i and H are the heterozygosities of population i and the total population, respectively. The diamonds are the observed values for 41 worldwide populations, where the arrow points to the Anatolian Turkish population.

CHAPTER 4

DISCUSSION

4.1 An Overview and Comparison of the Results of this Study

In the present study, ten autosomal *Alu* insertion polymorphisms (ACE, APO, A25, B65, D1, FXIIIB, HS4.32, HS4.69, PV92 and TPA25) were typed in the Anatolian Turkish population. All of them were found to be polymorphic in Anatolia and this feature made them informative sites for evolutionary studies on the Anatolian Turkish population.

Insertion frequency of each *Alu* element was calculated and it was concluded that all the systems were in the Hardy-Weinberg equilibrium ($p>0.05$). However, in a previous study, which was the only study on four *Alu* insertion polymorphisms (ACE, A25, PV92 and APO) in Anatolia, it was concluded that all systems were deviated from Hardy-Weinberg equilibrium ($p<0.05$-$p<0.001$) due to heterozygote deficiency (Gerçeker, 1998). This result was probably obtained due to mistyping of heterozygous individuals. In the present study, this mistyping problem was resolved by applying hot start to the genomic DNA before the polymerase chain reaction amplification. Accuracy

in genotyping was tested by repeating the amplification process and then comparing the results of each amplification. Ueda *et al.* (1996) confirmed that genotyping of angiotension I-converting enzyme (ACE) gene polymorphism could give misleading results by using standard PCR conditions, in which heterozygotes are mistyped as homozygote deletions. As a result, this study established that hot start eliminates mistyping of heterozygous individuals for ACE insertion by improving the stringency of primer annealing.

In order to see the status of the frequency values of these ten *Alu* insertions in the world, the results of the present study and the data from previous studies (Stoneking *et al.*, 1997, Nasidze *et al.*, 2001 and Romualdi *et al.,* 2002) were combined and given in Appendices C and D as two data sets. When the frequency values of these *Alu* insertions in Anatolians were examined, it became evident that these values are in the frequency range of Anatolia's neighbors Europeans and Caucasians. The exceptions were APO, A25, B65, PV92 in the first data set and APO, FXIIIB, HS4.69 in the second. This situation is presented in Table 8.

Alu insertion loci are biallelic; hence they have a maximum possible heterozygosity of 0.5. In the present study, the observed average heterozygosity for ten *Alu* insertions in Anatolia was calculated as 0.3637 and the expected average heterozygosity was 0.3752. Since heterozygosity is a measure of genetic diversity in a population, it can be concluded that *Alu* insertion variation is considerable in Anatolia. Moreover, there is a good fit between the expected and observed average heterozygosity values, which was also proven by the presence of the Hardy-Weinberg equilibrium in

all of the systems analyzed. Observed average heterozygosity of Anatolians for eight *Alu* insertions (ACE, APO, A25, B65, D1, FXIIIB, PV92 and TPA25) was also calculated to be 0.338. This value was compared with the heterozygosities of Caucasians and Europeans; which are 0.311 and 0.396, respectively (Nasidze *et al.*, 2001 and Stoneking *et al.*, 1997). Anatolians are intermediary between Caucasians and Europeans with respect to both heterozygosity and geographic location and since in peopling of Europe there were movements through Anatolia to Europe, higher heterozygosities in Anatolia than that of Europe was expected. The very same observation based on mtDNA was also observed (Comas *et al.*, 1996).

Table 8. The frequency values of *Alu* insertions in Anatolians and their range in Caucasians and Europeans for the first and second data sets.

		The First Data Set		**The Second Data Set**	
***Alu* Insertion**	**Anatolians**	**Europeans**	**Caucasians**	**Europeans**	**Caucasians**
ACE	0.33	0.33 - 0.51	0.17 - 0.48	0.28 - 0.50	0.17 - 0.48
APO	0.95	0.90 - 0.99	0.86 - 0.94	0.87 - 0.97	0.86 - 0.94
A25	0.07	0.09 - 0.20	0.00 - 0.09	0.00 - 0.17	0.00 - 0.11
B65	0.49	0.53 - 0.65	0.21 - 0.73	0.35 - 0.58	0.21 - 0.73
D1	0.37	0.27 - 0.44	0.00 - 0.42	0.15 - 0.47	0.00 - 0.42
FXIIIB	0.46	0.39 - 0.62	0.00 - 0.61	0.00 - 0.29	0.00 - 0.61
HS4.32	0.67	-	-	0.25 - 0.79	0.40 - 0.81
HS4.69	0.30	-	-	0.21 - 0.38	0.00 - 0.03
PV92	0.15	0.18 - 0.33	0.01 - 0.38	0.10 - 0.27	0.01 - 0.38
TPA25	0.48	0.43 - 0.58	0.22 - 0.51	0.32 - 0.60	0.22 - 0.51

Nei's D_A genetic distances (Tables 5 and 6) in two different data sets and the two neighbor joining trees (Figures 19 and 20) constructed by using these genetic distances displayed surprising discrepancies. Genetic distance values in the second data set were found to be larger and this discrepancy may be due to two additional (HS4.32 and HS4.69) *Alu* insertions studied in the second data set. The neighbor joining tree constructed with the first data set by using eight *Alu* insertion polymorphisms showed the population clusters more clearly than the second one. The striking differences mentioned above can be due to different samplings in different studies and different populations with different population sizes in each data set. On the other hand, increasing sample sizes of the populations may give better results because, average sample sizes for populations in the first and the second data sets were approximately 41 and 50, respectively. Therefore, small sample sizes compared to the sample size of Anatolians studied in this study (approximately 100) may be responsible for this result. Another reason may be that the optimum number of *Alu* insertion polymorphisms that should be examined in each population for more reliable results has not been determined yet.

Additionally, in both neighbor joining trees, low bootstrap numbers were observed that might be due to small sample sizes of the populations used in the analysis. Therefore, only bootstrap values larger than 30 out of 100 were denoted at the branching nodes of the trees.

In order to further assess population relationships in a three dimensional space, the principal component analysis was applied to the first and second data sets. The

principal component analysis showed a better clustering of human populations according to their geographical locations than the neighbor joining tree. However, PCA yields less detailed genetic relatedness of the populations than neighbor joining tree, because NJ trees try to join the two closest populations, where as PCA makes clusters of populations. If the number of variables is n, then the number of the principal components is n-1. Therefore, in this study with eight and ten variables, in the first and second PCA, respectively, three components do not account for a high proportion of the total genetic variation present between the populations. This is why the three principal components explain only 75.16 % and 71.12 % of the total variation in the first and second PCA, respectively. In both of the plots of principal component analysis, the Anatolian Turkish population was obviously placed in the European cluster as illustrated in Figures 21, 22 and 23, 24.

Another very important property of principal component analysis is that it manifests the most effective variables in differentiating the populations and explaining their variation. The results of the two principal component analyses performed for two different data sets were parallel. The analyses implied that FXIIIB, PV92 and ACE *Alu* insertion polymorphisms were the most successful *Alu* insertions in differentiating the populations. Therefore, this result will help in the choice of *Alu* insertions for further evolutionary studies. When this result is compared with Gerçeker (1998), which ranked four *Alu* insertions according to their power of differentiation as: ACE, PV92, A25 and APO, it is observed that the importance of ACE and PV92 should not be underestimated. FXIIIB was the *Alu* insertion with the highest discrimination power in the data of Stoneking *et al.*, (1997), which also correlated with the results of this study.

Discriminant function analysis is another analysis that helps to determine the power of discrimination of a variable. However, this analysis is performed with data separated into groups prior to the analysis and it maximizes the difference between groups. Thus, a higher percentage of variation is explained by discriminant function analysis (97.60% and 91.78%) compared to principal component analysis (75.16% and 71.12%). In this study, Anatolia was placed in the European group due to the results of the NJ tree and PCA, which concluded that Anatolia was genetically close to Europe and found in the European cluster. Multigroup discriminant analysis known as canonical variates analysis was applied to the two data sets and PV92 and D1 was the variables explaining the most variation in the first and the second data sets, respectively.

Canonical function analysis has another result showing the positions and relations of the groups of the populations. These plots for the first and second data sets are given in Figures 25 and 26, respectively. The plot of the canonical function analysis in Figure 25 shows close relationship between the Caucasians and Europeans which was also stated by Nasidze *et al.* (2001). However, in the second plot shown in Figure 26, the Caucasians and Europeans seemed to be distant from each other and this is a surprising result. The variables explaining the separation of the groups of populations can be determined in the plots (Figures 25 and 26) by analyzing the dashed lines. This result can later be used in the choice of variables to be studied. For instance, in Figure 25, variable 8 (A25) functioned well for the separation of group 7 (Africans) from the rest of the groups of populations. In Figure 26, variable 3 (APO) had an important role in differentiating group 2 (Americans) from the rest of the groups of populations. By examining the directions of the dashed lines, a rough estimation for the parallelism of

variables can be made. In Figure 25, since the dashed lines standing for FXIIIB (6), ACE (1) and PV92 (3) are parallel to each other, it can be said that these variables do the same differentiation among the groups of populations. Therefore, it is unnecessary to use three of them together in a study. In Figure 26, a strong parallelism was observed between the variables PV92 (5) and HS4.69 (10), between D1 (6) and TPA25 (1) and also between B65 (7) and HS4.32 (9).

The Fst analysis was performed to determine the genetic distance between Anatolians and geographically close populations. In addition to the Nei's D_A genetic distance, Fst analysis was also applied in this study because in this analysis the significance of the genetic distance values can be obtained with the help of permutations. A lower Fst value means a lower genetic distance and thus a higher genetic similarity between two populations. As shown in Table 7, non-significant Fst values were observed in the pairs of Anatolians-Swiss and Anatolians-French Acadian. Since; Swiss and French Acadian are in the European cluster, this result obviously implies that Anatolian Turks are genetically very close to Europeans.

Table 7 surprisingly shows that Anatolian Turks have significant differences between Azerbaijanians (Fst = 0.07629***), which are both Altaic-speaking populations, and Armenians (0.03012**), which are geographically very close to Anatolia. However, Nasidze *et al.,* 2001 stated that Armenians and Azerbaijanians, being small and isolated populations are genetically similar although they are Indo-European and Altaic-speaking populations, respectively. Therefore, the genetic

differences between Anatolians-Azerbaijanians and Anatolians-Armenians cannot be attributed to the isolation of the populations.

On the plot of heterozygosity *vs.* distance from centroid, it was observed that the Anatolia and India-Hindu populations had exactly the expected heterozygosity values as expected by the model of Harpending and Ward (1982). This result is fairly different from the result of Gerçeker (1998), which had lower heterozygosity values in Anatolia than expected by the model. Major deviations were observed in African populations and minor deviations in Caucasians and Australians.

In addition to the studies (Stoneking *et al,* 1997, Nasidze *et al,* 2001 and Romualdi *et al.*, 2002) whose data were used for comparison of the data of Anatolians, Comas *et al.*, (2000) is another study that used *Alu* insertion polymorphisms in Northwest Africa and Iberian Peninsula. The data of Comas *et al.*, (2000) was not included in the comparisons of Anatolia data because Northwest Africa and Iberian Peninsula are quite far from Anatolia geographically.

In conclusion, with the results of the first PCA and Fst analysis, it became clear that the Anatolian Turkish population is genetically closer to Europeans than its eastern neighbors; the Caucasians. This result is reasonable since Calafell *et al.*, (1996) and Comas *et al.*, (1996) proposed that the mean pairwise differences in two hypervariable sequence segments in the control region of mtDNA suggested that a demographic expansion occurred sequentially in the Middle East through Turkey to the rest of Europe in times ranging between 50,000-100,000 and 35,000-100,000 years ago, respectively.

4.2 Conclusion

This study included the examination of ten *Alu* insertion polymorphisms in the Anatolian Turkish population. All systems were in Hardy Weinberg equilibrium and almost all of the frequencies for the insertions of the *Alu* elements were in the range of Causcasians and Europeans.

The Anatolian Turkish population was found to be clustered with the European populations due to the analyses performed by using the allele frequencies. Moreover, this clustering and the genetic similarity of the populations were both proved by the Fst analysis. However, Central Asian populations should also be included in a further study to be able to have an improved conclusion.

In addition, many more loci in populations with higher sample sizes must be examined to increase the statistical confidence of the results. Furthermore, polymorphic LINEs can be studied together with *Alu* insertion polymorphisms due to their common important and valuable features for human population genetic studies (Sheen *et al.*, 2000).

The results of this study can be combined with the results of other studies performed on Anatolian Turks with different polymorphic genetic markers. Therefore, this will help the genetic characterization of Turkish populations.

REFERENCES

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

Batzer, M.A. and Deninger, P.L., (2002). *Alu* repeats and human genomic diversity. *Nature Reviews*, 3: 370-380.

Batzer, M.A., Arcot, S.S., Phinney, J.W., Alegria-Hartman, M., Kass, D.H., Milligan, S.M., Kimpton, C., Gill, P., Hochmeister, M., Ioannou, P.A., Herrera, R.J., Boudreau, D.A., Scheer, W.D., Keats, B.J.B., Deininger, P.L. and Stoneking, M., (1996a). Genetic variation of recent *Alu* insertions in human populations. *Journal of Molecular Evolution*, 42: 22-29.

Batzer, M.A., Deninger, P.L., Hellmann-Blumberg, U., Jurka, J., Labuda, D., Rubin, C.M., Schmid, C.W., Zietkiewicz, E and Zuckerkandl, E., (1996b). Standardized nomenclature for *Alu* repeats. *Journal of Molecular Biology*, 42: 3-6.

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

Batzer, M.A., Kilroy, G.E., Richard, P.E., Shaikh, T.H., Desselle, T.D., Hoppens, C.L. and Deininger, P.L., (1990). Structure and variability of recently inserted *Alu* family members. *Nucleic Acids Research*, 18(23): 6793-6798.

Batzer, M.A., Rubin, C.M., Hellmann-Blumberg, U., Alegria-Hartman, M., Leeflang, E.P., Stern, J.D., Bazan, H.A., Shaikh, T.H., Deininger, P.L. and Schmid, C.W., (1995). Dispersion and insertion polymorphism in two small subfamilies of recently amplified human *Alu* repeats. *Journal of Molecular Biology*, 247: 418-427.

Batzer, M.A., Stoneking, M., Alegria-Hartman, M., Bazan, H., Kass, D.H., Shaikh, T.H., Novick, G.E., Ioannou, P.A., Scheer, W.D., Herrera, R.J. and Deininger, P.L., (1994). African origin of human-specific polymorphic *Alu* insertions. *Proceedings of the National Academy of Sciences*, 91: 12288-12292.

Belkhir, K., Borse, P., Chikhi, L., Goudet, J., Bonchomme, F., (1996-2001). Genetix 4.0 windows™ Software for population genetics. Laboratoire Genome et Populations, University of Montpellier 2, Montpellier, France. http://www.univ-montp2.fr/~genetix/gnetix.htm

Benedette, G.D., Ergüven, A., Stenico, M., Castri, L., Bertorelle, G., Togan, İ. and Barbujani, G., (2001). DNA diversity and population admixture in Anatolia. *American Journal of Physical Anthropolgy*, 115: 144-156.

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

Cann, R.L., Stoneking, M. and Wilson, A.C., (1987). Mitochondrial DNA and human evolution. *Nature*, 325: 31-36.

Carroll, M.L., Roy-Engel, A.M., Nguyen, S.V., Salem, A.H., Vogel, E., Vincent, B., Myers, J., Ahmad, Z., Nguyen, L., Sammarco, M., Watkins, W.S., Henke, J., Makalowski, W., Jorde, L.B., Deininger, P.L. and Batzer, M., (2001). Large-Scale Analysis of the *Alu* Ya5 and Yb8 Subfamilies and their contribution to human genomic diversity. *Journal of Molecular Biology*, 311: 17-40.

Cavalli-Sforza, L.L., (1998). The DNA revolution in population genetics. *Trends in Genetics*, 14(2): 60-65.

Comas, D., Calafell, F., Benchemsi, N., Helal, A., Lefrane, G., Stoneking, M., Batzer, M.A., Bertranpetit, J. and 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., Calafell, F., Mateu, E., Perez-Lezaun, A. and Bertranpetit, J., (1996). Geographic variation in human mitochondrial DNA control region sequence: the population history of Turkey and its relationships to European populations. *Molecular Biology and Evolution*, 13: 1067-1077.

Daniel, W.W., (1999). Biostatistics: A foundation for analysis in the health sciences. John Willey and Sons, Inc., USA, (7th Ed.).

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

Donaldson, C.J., Crapanzano, J.P., Watson, J.C., Levine, E.A. and Batzer, M.A., (2002). PROGINS *Alu* insertion and human genomic diversity. *Mutation Research*, 501: 137-141.

Dytham, C., (2003). Choosing and Using Statistics: A Biologist's Guide. Blackwell Publishing Company, (2nd Ed.).

Edwards, M. and Gibbs, R., (1992). A human dimorphism resulting from loss of an *Alu*. *Genomics*, 14: 590-597.

Ergüven, A., (1997). The genetic structure of Anatolian Turkish population based on fifteen loci. PhD Thesis, METU.

Felsenstein, J., (1993). PHYLIP (Phylogeny Inference Package), version 3.6c. Department of Genetics, University of Washington, Seattle. http://evolution.genetics.washington.edu/phylip.html

Gerçeker, F.Ö., (1998). *Alu* insertion polymorphisms in Anatolia. MSc. Thesis, METU.

Harpending, H.C. and Ward, R.H., (1982). Chemical systematics and human populations. In: Nitecki M. (ed) Biochemical aspects of evolutionary biology. University of Chicago Press, Chicago, pp 213-256; cited in: Batzer, M.A., Arcot, S.S., Phinney, J.W., Alegria-Hartman, M., Kass, D.H., Milligan, S.M., Kimpton, C., Gill, P., Hochmeister, M., Ioannou, P.A., Herrera, R.J., Boudreau, D.A., Scheer, W.D., Keats, B.J.B., Deininger, P.L. and Stoneking, M., (1996a). Genetic variation of recent *Alu* insertions in human populations. *Journal of Molecular Evolution*, 42: 22-29.

Harpending, H.C., Sherrry, S.T., Rogers, A.R. and Stoneking, M., (1993). The genetic structure of ancient human populations. *Current Anthropology*, 34(4): 483-496.

Hedrick, P.W., (2000). Genetics of Populations. Jones and Bartlett Publishers, Inc., UK, (2nd Ed.).

İçener, S., (2001). 9-bp deletion, *Alu* insertion and microsatellite polymorphisms in Anatolia and Central Asian Turks. MSc. Thesis, METU.

Iwasa, M., Wiegand, P., Rand, S., Schürenkamp, M., Atasoy, S. and Brinkmann, B., (1997). Genetic variation at five STR loci in subpopulations living in Turkey. *International Journal of Legal Medicine*, 110(3): 170-172.

Jorde, L.B., Watkins, W.S., Bamshad, M.J., Dixon, M.E., Ricker, C.E., Seielstad, M.T. and 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.

Jurka, J., (1993). A new subfamily of recently retroposed human *Alu* repeats. *Nucleic Acids Research*, 21(9): 2252.

Klug, W.S. and Cummings, M.R., (1997). Concepts of Genetics. Prentice-Hall, Inc., Simon and Schuster/A Viacom Company, Upper Saddle River, New Jersey, USA, (5th Ed.).

Knight, A., Batzer, M.A., Stoneking, M., Tiwari, H.K., Scheer, W.D., Herrera, R.J. and Deininger, P.L. (1996). DNA sequences of *Alu* elements indicate a recent replacement of the human autosomal genetic complement. *Proceedings of the National Academy of Sciences*, 93: 4360-4364.

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

Nei, M. and Kumar, S., (2000). Molecular evolution and phylogenetics. Oxford University Press.

Nei, M. and Roychoudhury, A.K., (1993). Evolutionary relationships of human populations on a global scale. *Molecular Biology and Evolution*, 10(5): 927-943.

Novick, G.E., Batzer, M.A., Deininger, P.L. and Herrera, R.J., (1996). The mobile genetic element *Alu* in the human genome. *BioScience*, 46: 32-40.

Önde, S. and Kence, A., (1995). Geographic variation analysis of the ABO and RH systems in Turkey. *Gene Geography*, 9: 211-220.

Podani, J., (1993). SYNTAX-pc: Computer programs for multivariate data analysis in ecology and systematics.

Renfrew, C., (1992). Archaeology, genetics and linguistic diversity. *Man*, 27: 445-478.

Roberts, J.M., (1993). History of the world. Oxford University Press, Inc., New York, USA.

Rohlf, F.J., (1993). NTSYS-pc: Numerical Taxonomy and Multivariate Analysis System, version 2.10q., App. Biostatistics, Inc.

Romualdi, C., Balding, D., Nasidze, I.S., Risch, G., Robichaux, M., Sherry, S.T., Stoneking, M., Batzer, M.A. and Barbujani, G., (2002). Patterns of human diversity, within and among continents, inferred from biallelic DNA polymorphisms. *Genome Research*, 2: 602-612.

Roy-Engel, A.M., Carrol, M.L., Vogel, E., A.H., Garber R.K., Nguyen, S.V., Salem, A.H., Batzer, M.A. and Deininger, P.L., (2001). *Alu* insertion polymorphisms for the study of human genomic diversity. *Genetics Society of America*, 159: 279-290.

Saatçioğlu, A., (1979). An analysis of the ABO gene frequencies in Turkey. *Human Heredity*, 36: 218-221.

Sheen, F., Sherry, S.T., Risch, G.M., Robichaux, M., Nasidze, I., Stoneking, M., Batzer, M., and Swergold, G.D., (2000). Reading between the LINEs: human genomic variation induced by LINE-1 retrotransposition. *Genome Research*, 10: 1496-1508.

Shen, M.R., Batzer, M.A. and Deininger, P.L., (1991). Evolution of the master *Alu* gene(s). *Journal of Molecular Evolution*, 33: 311-320.

Stoneking, M., Deininger, P.L. and Batzer, M.A. (2001). *Alu* insertion polymorphisms in humans: a review; in Donnelly, P. and Foley, R. (eds.). Genes, Fossils and Behaviour, IOS Press, Amsterdam, The Netherlands, 111-121.

Stoneking, M., Fontius, J.J., Clifford, S.L., Soodyall, H., Arcot, S.S., Saha, N., Jenkins, T., Tahir, M.A., Deininger, P.L. and Batzer, M.A., (1997). *Alu* insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Research*, 7: 1061-1071.

Takeshita, H., Meyer, E. and Brinkmann, B., (1997). The STR loci HumTPO and HumLPL: population geneic data in eight populations. *International Journal of Legal Medicine*, 110(4): 331-333.

Templeton, A.R., (2002). Out of Africa again and again. *Nature*, 416: 45-51.

Tishkoff, S.A., Pakstis, A.J., Stoneking, M., Kidd, J.R., Destro-Bisol, G., Sanjantila, A., Lu, R.-b., Deinard, A.S., Sirugo, G., Kenkins, T., Kidd, K.K. and Clark, A.G., (2000). Short tandem-repeat polymorphism/*Alu* halotype variation at the PLAT locus: implications for modern human origins. *American Journal of Human Genetics*, 67: 901-925.

Togan, İ. and Ergüven, A., (1994). Red cell enzyme and blood group polymorphisms in Turkey. *Gene Geography*, 8: 75-80.

Togan, İ., Yılmaz, A., Ergüven, A., Dereli, İ. and Özbaş, F., (1996). Glyoxalase 1 (GLO1) polymorphism in eight geographical regions of Turkey. *Anthropol. Anz.*, 54(2): 121-124.

Ueda, S., Heeley, R.P., Lees, K.R., Elliot, H.L. and Connell, J.M.C., (1996). Mistyping of the human angiotensin-converting enzyme gene polymorphism: frequency, causes and possible methods to avoid errors in typing. *Journal of Molecular Endocrinology*, 17: 27-30.

Ullu, E. and Tschudi, C., (1984). *Alu* sequences are processed 7SL RNA genes. *Nature*, 302: 171-172.

Vigilant, L., Stoneking, M., Harpending, H., Hawkes, K. and Wilson, A.C., (1991). African populations and the evolution of human mitochondrial DNA. *Science*, 253: 1503-1507.

Vural, B., Atlioğlu, E., Kolusayin, Ö., Togan, İ., Büyükdevrim, S. and Özçelik, T., (1998). Turkish population data on the HLA-DQα, LDLR, GYPA, HBGG, D7S8 and GC Loci. *International Journal of Leagal Medicine*, 111(1): 43-45.

Watkins, W.S., Ricker, C.E., Bamshad, M.J., Carroll, M.L., Nguyen, S.V., Batzer, M.A., Harpending, H.C., Rogers, A.R. and Jorde, L.B., (2001). Patterns of ancestral human diversity: An analysis of *Alu*-insertion and restriction-site polymorphisms. *American Journal of Human Genetics*, 68: 738-752.

Weir, B.S. and Cockerham, C.C., (1984). Estimating F statistics for the analysis of population structure. *Evolution*, 38: 1358-1370.

Wright, S., (1951). The genetic structure of populations. *Ann. Eugen*. 15**:** 313-354.

Zerjal, T., Dashnyam, B., Pandya, A., Kayser, M., Roewer, L., Santos, F.R., Schiefenhovel, W., Fretwell, N., Jobling, M.A., Harihara, S., Shimizu, K., Semjidmaa, D., Sajantila, A., Salo, P., Crawford, M.H., Ginter, E.K., Evgrafov, O.V. and Tyler-Smith, C., (1997). Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. *American Journal of Human Genetics*, 60(5): 1174-1183.

APPENDIX A

The province and the gender of the samples.

SAMPLE NO	PROVINCE	GENDER
1	ANKARA	F
2	SİVAS	F
3	ANTALYA	F
4	ANKARA	M
5	DENİZLİ	F
6	ANKARA	M
7	ISPARTA	F
8	BURSA	M
9	ADANA	F
10	KAYSERİ	M
11	ANTALYA	M
12	HATAY	F
13	HATAY	M
14	HATAY	M
15	ARTVİN	F
16	SİVAS	F
17	GİRESUN	M
18	MARDİN	M
19	BİLECİK	M
20	ERZURUM	M
21	ÇORUM	F
22	ERZURUM	M
23	NİĞDE	M
24	ANKARA	M
25	AMASYA	F

(continued)

26	KIRŞEHİR	M
27	UŞAK	F
28	AKSARAY	F
29	KONYA	M
30	ANKARA	F
31	TRABZON	F
32	MERSİN	M
33	ARDAHAN	M
34	SİVAS	M
35	SİVAS	F
36	DENİZLİ	F
37	TRABZON	M
38	ORDU	F
39	URFA	M
40	KAHRAMANMARAŞ	M
41	SİİRT	M
42	RİZE	F
43	YOZGAT	M
44	BALIKESİR	F
45	GAZİANTEP	F
46	ISPARTA	M
47	ÇORUM	F
48	ÇORUM	M
49	TRABZON	F
50	İSTANBUL	M
51	ARDAHAN	M
52	ESKİŞEHİR	M
53	ADANA	M
54	KONYA	M
55	ANKARA	M

(continued)

56	KARAMAN	M
57	ERZİNCAN	M
58	KOCAELİ	M
59	MALATYA	F
60	KARAMAN	M
61	NEVŞEHİR/ÜRGÜP	F
62	TRABZON	F
63	ESKİŞEHİR	F
64	AFYON	M
65	AKSARAY	M
66	KONYA	M
67	ADANA	M
68	BOLU	F
69	BİTLİS	F
70	ÇORUM	M
71	MİLAS	M
72	HATAY	M
73	KONYA	M
74	KONYA	M
75	KIRŞEHİR	M
76	KARAMAN	F
77	KONYA	M
78	AFYON	M
79	ISPARTA	M
80	KONYA	M
81	KONYA	M
82	KONYA	M
83	AĞRI	M
84	KONYA	M
85	KONYA	M

(continued)

86	KONYA	M
87	IĞDIR	M
88	KONYA	M
89	KAYSERİ	M
90	ÇANKIRI	F
91	KONYA	M
92	TOKAT	F
93	KONYA	M
94	KONYA	M
95	KONYA	F
96	İZMİR	M
97	KAYSERİ	M
98	AFYON	M
99	ÇANKIRI	M
100	ISPARTA	F
101	BAYBURT	M
102	AYDIN	M

APPENDIX B

1) QUESTIONNAIRE:

Your Name, Surname:

Your Sex:

Your Birth Place and Year:

Your mother's birth place:
Your father's birth place:

Your grandmother's (mother's) birth place:
Your grandfather's (mother's) birth place:

Your grandmother's (father's) birth place:
Your grandfather's (mother's) birth place:

Do you have any genetically disorder(s)?

Cancer:	Yes	No
You	☐	☐
Your mother	☐	☐
Your Father	☐	☐
Your brothers or sisters	☐	☐

Address and telephone number:

2) CONSENT:

Attached questionnaire is for the Master thesis named "*Alu* Insertion Polymorphisms in Anatolian Turks" by the graduate student Havva Dinç'in the Department of Biological Sciences of Middle East Technical University.

In this study it is aimed to determine the genetic structure and the evolutionary history of the Anatolia and to contribute knowledge on these subjects.

I know that my DNA will be used in a Master thesis in the Department of Biological Sciences of Middle East Technical University and I accept this usage.

Signature:

Name and Surname:

APPENDIX C

Frequency data of eight Alu insertion polymorphisms (ACE, APO, A25, B65, D1, FXIIIB, PV92 and TPA25) from Stoneking *et al.*, 1997, Nasidze *et al.*, 2001 together with the present data. (N: Number of the individuals studied. *: The results of the present study.)

		Frequencies							
Populations	**N**	**ACE**	**APO**	**A25**	**B65**	**D1**	**FXIIIB**	**PV92**	**TPA25**
AFRICANS									
!Kung	40	0.290	0.880	0.610	0.500	0.160	0.170	0.200	0.170
Nigerian	11	0.270	0.500	0.220	0.830	0.000	0.080	0.090	0.410
Nguni	43	0.400	0.600	0.410	0.600	0.270	0.120	0.240	0.210
Pygmy-CAR	17	0.120	0.740	0.350	0.780	0.470	0.000	0.260	0.210
Pymgy-Zaire	17	0.320	0.850	0.530	0.820	0.590	0.030	0.350	0.240
Sotho- Tswana	48	0.380	0.680	0.390	0.480	0.310	0.180	0.290	0.330
AMERICANS									
Alaska Natives	41	0.580	0.920	0.150	0.450	0.420	0.920	0.620	0.300
Greenland Natives	41	0.550	0.940	0.170	0.190	0.450	0.790	0.610	0.330
Mayan	51	0.680	0.940	0.210	0.270	0.450	0.900	0.790	0.650
Mvskoke	50	0.700	0.960	0.210	0.480	0.460	0.760	0.570	0.490
EAST AND SOUTHEAST ASIANS									
China	49	0.670	0.820	0.100	0.350	0.170	0.710	0.860	0.350
Filipino	47	0.530	0.980	0.140	0.570	0.360	0.720	0.800	0.630
Java	32	0.860	0.780	0.060	0.580	0.420	0.920	0.840	0.390
Malaysian	47	0.640	0.760	0.020	0.420	0.270	0.730	0.720	0.500
Moluccan	48	0.670	0.760	0.000	0.260	0.190	0.780	0.690	0.560
Taiwan	46	0.500	0.930	0.220	0.540	0.380	0.970	0.900	0.640
Tenggaras	90	0.650	0.780	0.050	0.400	0.190	0.810	0.500	0.380
WEST ASIANS									
India-Christan	27	0.600	0.670	0.140	0.310	0.280	0.610	0.480	0.570
India- Hindu	28	0.520	0.850	0.050	0.350	0.100	0.660	0.520	0.340
India- Muslim	26	0.520	0.860	0.120	0.400	0.320	0.660	0.300	0.410
Pakistan	42	0.440	0.720	0.070	0.370	0.170	0.230	0.300	0.510
Pushtoon	50	0.520	0.860	0.180	0.490	0.270	0.570	0.330	0.550
Tamill	47	0.690	0.810	0.170	0.550	0.340	0.610	0.560	0.560
UAE	42	0.330	0.970	0.120	0.410	0.080	0.390	0.300	0.440

(continued)

AUSTRALIA									
Australia	69	0.910	0.870	0.350	0.390	0.040	0.650	0.150	0.130
PNG-Coastal	48	0.660	0.660	0.020	0.270	0.170	0.300	0.360	0.160
PNG- Highland	68	0.740	0.680	0.040	0.180	0.010	0.300	0.240	0.160
CAUCASIANS									
Armenians	40	0.477	0.871	0.058	0.453	0.151	0.343	0.013	0.430
Azerbaijanians	34	0.216	0.943	0.000	0.697	0.333	0.100	0.382	0.513
Cherkessians	40	0.390	0.932	0.045	0.651	0.167	0.439	0.167	0.386
Darginians	16	0.167	0.864	0.028	0.321	0.346	0.143	0.167	0.361
Georgians	67	0.354	0.934	0.088	0.727	0.418	0.610	0.250	0.493
Ingushians	24	0.340	0.941	0.067	0.210	0.000	0.000	0.129	0.224
EUROPEANS									
Anatolians*	**100**	**0.333**	**0.951**	**0.069**	**0.485**	**0.371**	**0.460**	**0.152**	**0.480**
Bretons	54	0.480	0.900	0.160	0.560	0.390	0.400	0.270	0.560
European-American	57	0.510	0.940	0.200	0.560	0.440	0.470	0.180	0.560
French	53	0.480	0.990	0.160	0.570	0.460	0.420	0.230	0.560
French Acadian	46	0.510	0.920	0.120	0.530	0.420	0.480	0.180	0.430
Greek Cypriot	48	0.390	0.950	0.120	0.650	0.270	0.620	0.250	0.530
Turkish Cypriot	33	0.330	0.980	0.090	0.640	0.350	0.390	0.330	0.580
Swiss	43	0.370	0.940	0.120	0.580	0.340	0.480	0.200	0.450

Frequency data of ten *Alu* insertion polymorphisms (ACE, APO, A25, B65, D1, FXIIIB, HS4.32, HS4.69, PV92 and TPA25) from Romualdi *et al.,* 2002 together with the present data. (N: Number of the individuals studied. *: The results of the present study.)

		Frequencies									
Populations	**N**	**ACE**	**APO**	**A25**	**B65**	**D1**	**FXIIIB**	**HS4.32**	**HS4.69**	**PV92**	**TPA25**
AFRICANS											
African American	68	0.44	0.59	0.07	0.36	0.47	0.07	0.46	0.20	0.20	0.43
!Kung	38	0.13	0.99	0.23	0.48	0.00	0.07	0.24	0.10	0.17	0.14
Nguni	38	0.36	0.60	0.27	0.34	0.00	0.09	0.00	0.04	0.18	0.20
AMERICANS											
Cajun	61	0.46	0.96	0.13	0.43	0.28	0.06	0.64	0.40	0.21	0.43
Hispnic American	68	0.53	0.97	0.15	0.02	0.35	0.50	0.25	0.25	0.51	0.56
Maya	27	0.67	0.96	0.00	0.29	0.35	0.88	0.27	0.36	0.70	0.64
Mvskoke	30	0.57	1.00	0.14	0.55	0.46	0.56	0.30	0.00	0.53	0.53
Yanomamo	23	0.75	1.00	0.00	0.33	0.33	1.00	0.24	0.00	0.96	0.69

(continued)

SOUTHEAST ASIANS											
Bantu Speakers	40	0.33	0.73	0.08	0.46	0.21	0.00	0.04	0.14	0.28	0.30
Moluccas	45	0.56	0.75	0.00	0.25	0.20	0.78	0.23	0.09	0.69	0.55
Nusa Tenggarans	87	0.61	0.78	0.05	0.40	0.17	0.81	0.33	0.08	0.51	0.39
WEST ASIANS											
Syrians	68	0.40	0.93	0.00	0.31	0.29	0.28	0.52	0.32	0.10	0.51
AUSTRALIA											
PNG Coastal	47	0.66	0.66	0.02	0.20	0.08	0.18	0.31	0.38	0.29	0.17
PNG Highland	52	0.75	0.66	0.00	0.13	0.01	0.01	0.34	0.39	0.18	0.14
CAUCASIANS											
Armenian	37	0.48	0.87	0.06	0.45	0.15	0.34	0.63	0.00	0.01	0.43
Azerbaijani	35	0.22	0.94	0.00	0.70	0.33	0.10	0.64	0.00	0.38	0.51
Cherkessian	40	0.39	0.93	0.05	0.65	0.17	0.44	0.61	0.00	0.17	0.39
Darginian	16	0.17	0.86	0.03	0.32	0.35	0.14	0.77	0.00	0.17	0.36

(continued)

Ingushian	25	0.34	0.94	0.07	0.21	0.00	0.00	0.40	0.03	0.13	0.22
Kabardinian	28	0.27	0.93	0.11	0.43	0.13	0.14	0.60	0.02	0.15	0.29
Georgian	67	0.35	0.93	0.09	0.73	0.42	0.61	0.81	0.00	0.25	0.49
EUROPEANS											
Anatolians*	**100**	**0.33**	**0.95**	**0.07**	**0.49**	**0.37**	**0.46**	**0.67**	**0.30**	**0.15**	**0.48**
Bretons	66	0.28	0.90	0.17	0.51	0.47	0.16	0.61	0.21	0.27	0.32
European American	69	0.50	0.93	0.14	0.58	0.21	0.06	0.60	0.28	0.23	0.54
French	68	0.41	0.95	0.09	0.49	0.39	0.29	0.55	0.31	0.16	0.60
Greek Cypriot	47	0.33	0.95	0.00	0.53	0.15	0.18	0.69	0.25	0.15	0.51
German	67	0.46	0.87	0.14	0.35	0.31	0.05	0.55	0.36	0.10	0.51
Hungarian	67	0.38	0.97	0.04	0.45	0.46	0.00	0.67	0.38	0.12	0.50
Turk Cypriot	56	0.36	0.96	0.08	0.39	0.31	0.01	0.79	0.30	0.15	0.40
Swiss	66	0.39	0.94	0.15	0.57	0.34	0.16	0.25	0.32	0.16	0.50

APPENDIX E

CHEMICAL SOLUTIONS USED IN THIS STUDY

1) 10 X Lysis Buffer:
   770 mM NH_4Cl
   46 mM $KHCO_3$
   10 mM EDTA

2) Salt-EDTA Buffer
   75 mM NaCl
   25 mM EDTA

3) Sodium Dodecyl Sulfate
   10 % (w/v) SDS

4) Proteinase-K
   10 mg/ml (w/v) Proteinase K

5) Sodium Acetate (NaAc)
   3 M NaAc

6) Tris-EDTA (TE) Buffer (pH: 7.5)

10 mM Tris

1 mM EDTA

7) Sodium Hyroxide (NaoH) Solution

50 mM NaOH

8) Tris-HCl Solution (pH:8)

1M Tris-HCl

9) 5X Tris Borate EDTA (TBE) Buffer (pH: 8.0)

0.45 M Tris (Base)

0.45 M Boric Acid