AN EXTENDED STUDY ON THE ALU INSERTION POLYMORPHISMS IN ANATOLIAN HUMAN POPULATION

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

BY

CERAN ŞEKERYAPAN

IN PARTIAL FULLFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
BIOLOGY

SEPTEMBER 2005

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. Semra Kocabıyık
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. Zeki Kaya	METU, Biology	_____
Prof. Dr. İnci Togan	METU, Biology	_____
Prof. Dr. Işık Bökesoy	Ankara Univ., Faculty of Medicine	_____
Prof. Dr. Serap Emre	Hacettepe Univ., Faculty of Medicine	_____
Ass. Prof. Dr. Elif Erson	METU, Biology	_____

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

Name, Last name :

Signature :

ABSTRACT

AN EXTENDED STUDY ON THE *ALU* INSERTION POLYMORPHISMS IN ANATOLIAN HUMAN POPULATION

ŞEKERYAPAN, Ceran
M. Sc., Department of Biological Sciences
Supervisor: Prof. Dr. İnci TOGAN

September 2005, 66 pages

In the present study, for estimating the Central Asia contribution to the Anatolia, nine *Alu* insertion polymorphisms (*ACE, PV92, FXIIIB, APO, A25, B65, TPA25, D1, HS4.32*) in 100 individuals from Anatolia were examined. *Alu* insertion frequency for these loci were calculated as 0,410; 0,220; 0,579; 0,963; 0,067; 0,667; 0,390; 0,427; and 0,637 respectively and they were found to be in Hardy-Weinberg equilibrium ($p<0,05$). Observed insertion frequencies of each loci were compared with those of the previous observations (Dinç, 2003; Comas et al., 2004) and it was found that the present study results were not different than those obtained by Comas et al. (2004). Thus, these two data were pooled (N = 143) and used to examine genetic relationships between populations from Eurasia and Africa.

Pairwise Fst statistics indicated that there is higher genetic similarity between Anatolia and all of the Balkans and some of the Caucasian populations. Neighbor Joining (NJ) tree based on Reynold's genetic distances and Principal Component Analysis (PCA) both grouped the Anatolian populations with Balkans and some of the Caucasian populations and show clear differentiation of Asian populations from the Anatolian population.

The relative genetic contribution of Central Asian genes to the current Anatolian gene pool was quantified using Admix analysis, considering for comparison populations of Balkans (Greek, Romania, Albania and Hungarian) and Central Asia (Uighur, Uzbeks, Tajicks, Kazaks, Kyrgyzes, Dungans). Estimates suggest roughly 28 % contribution from Asia to Anatolia in concordance with the previous estimation (Benedetto *et al*., 2001).

Keywords: *Alu* insertion polymorphism, human, Anatolia, Neighbor Joining tree, Principal Component Analysis, Admixture.

ÖZ

ANADOLU İNSAN TOPLULUĞUNDA ALU-ARA İLAVE
ÇEŞİTLİLİĞİ ÜZERİNE GENİŞLETİLMİŞ ÇALIŞMA

ŞEKERYAPAN, Ceran
Yüksek Lisans, Biyoloji Bölümü
Tez Yöneticisi: Prof. Dr. İnci TOGAN

Eylül 2005, 66 sayfa

Anadolu'ya Orta Asya katkısını hesaplamak amacıyla bu çalışmada dokuz *Alu* ara-ilave polimorfizmleri (*ACE, PV92, FXIIIB, APO, A25, B65, TPA25, D1, HS4.32*), Anadolu'dan 100 bireyde incelenmiştir. Bu lokusların ilave frekansları sırasıyla 0,410; 0,220; 0,579; 0,963; 0,067; 0,667; 0,390; 0,427; ve 0,637 olarak hesaplanmış ve lokuslar Hardy-Weinberg dengesinde bulunmustur ($p<0,05$). Her bir lokusun gozlemlenen ilave frekansları önceki gozlemler ile karşılaştırılmış (Dinç, 2003; Comas et al., 2004), ve bu calışmada elde edilen frekans değerleri Comas et al. (2004)'ın elde ettiklerinden istatistiksel olarak anlamlı derecede farklı olmadığı görülmüştür. Bu yüzden, iki data birleştirilmiş (N=143) ve Avrasya ve Afrikadaki populasyonlar ile arasındaki genetik ilişkiyi incelemede kullanılmıştır.

Fst istatistiği Anadolu ile tüm Balkan populasyonları arasında ve Anadolu ile bazı Kafkas populasyonları arasında yüksek genetik benzerliği işaret etmiştir. Temel öğeler ve Reynold'ın genetik uzaklığına dayanan komşu birleştirme ağacı analizleri Anadolu populasyonunu Balkanlar ve bazı Kafkas populasyonları ile gruplandırmış ve Asya populasyonlarının Anadolu populasyonundan açık farkını gostermiştir.

Balkan (Yunanistan, Romanya, Arnavutluk and Macaristan) ve Orta Asya (Uygur, Özbek, Tajik, Kazak, Kırgız, Dungan) populasyonları karsılaştırılarak Orta Asya'nın bugünkü Anadolu gen havuzuna genetik katkısı Admix analizi ile ölçülmüştür. Hesaplamalar önceki hesaplama (Benedetto et al., 2001) ile de uygunluk gostererek Asya'dan Anadolu'ya katkının 28 % olduğu göstermiştir.

Anahtar kelimeler: *Alu* ara-ilave polimorphizmleri, insane, Anadolu, komşu birleştirme ağacı, temel öğeler analizi, populasyon karışım analizi.

To my love, Raşit Özkan

ACKNOWLEDGEMENTS

First of all, I express my sincere appreciation to Prof. Dr. İnci Togan for her guidance, advice, criticism and insight throughout the research.

Secondly, I wish to express my gratitute to Ceren Caner Berkman for all her help.

And special thanks to my friends Sargun Tont and Çigdem Gökçek for their friendship and encouragement.

I'm grateful to my family for their effort to relieve my mind while my body under a stress.

Finally, thanks my love for the enormous motivation which enabled me to finish this work.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

The current analysis of genetic variation offers a way of viewing human population history concerning the origin of modern humans (in a wide range) and geographic scale (at a narrow range) (Comas *et al*., 1996).

Possibly the most interesting issue studied by geneticists has been the origin of anatomically modern humans for which there are two competing hypothesis. On the one side "multiregional hypothesis" states that modern *Homo sapiens* evolved across the world from regional archaic human population over the course of a million years in several different locations in the old world (Figure 1A) which were always linked by gene flow (Jorde *et al*., 1998).

Whereas "recent African origin" hypothesis derived from the analysis of several mitochondrial DNA (mtDNA) polymorphisms posits that transformation of archaic to anatomically modern forms of *Homo sapiens* appeared first in Africa, the time back to the most recent common ancestor (TMRCA) of modern human mtDNA was 190,000 years and that all present day humans are descendants of that African population (Cann *et al*., 1987) (Figure 1B). Nei & Roychoudhury (1993) suggest that

modern humans arose in Africa approximately 200,000 years ago and started to disperse throughout the old word 100,000 years ago.

Other population geneticists proposed another model which is intermediate between the strict "recent African origin model" and "multiregional evolution model" (Templeton, 2002). Templeton (2002) stated that a minimum of two out-of-Africa expansion events occurred instead of only one (Figure 2). Templeton analyzed human genetic trees for maternally inherited mitochondrial DNA, paternally inherited Y-chromosomal DNA, and eight other DNA regions, including two on the X chromosome, to reach his conclusions.

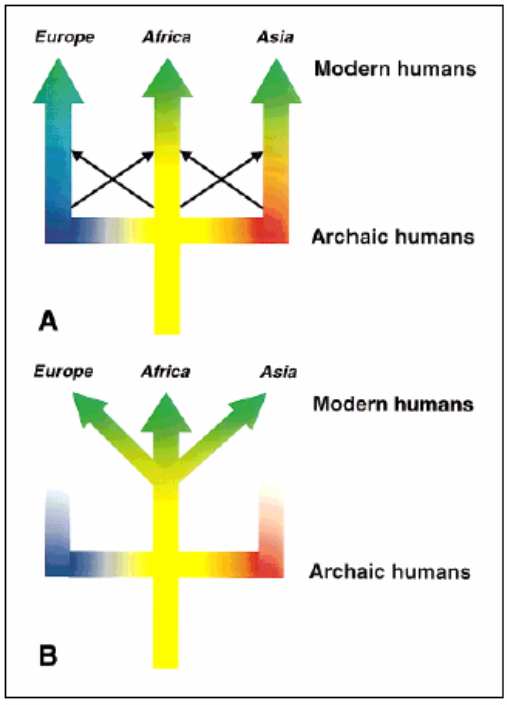

Figure 1. A: Illustration of the multiregional hypothesis. Gene flow, shown by black arrows, maintains genetic similarities between populations. B: Illustration of the African replacement hypothesis (Adopted from Jorde *et al*., 1998).

However, after a decade, subsequent data from the genomic DNA failed to support "recent African origin" model. A recent study claimed that the nuclear data regard population expansion associated with the spread of anatomically modern humans. It was suggested that present day modern humans are not exclusively derived from early modern Africans, but have a significant genetic inheritance from non-African archaic as well (Eswaran *et al*., 2005). They explore this proposed scenario through the Monte Carlo simulations. Simulations suggest that as much as 80% of nuclear loci have assimilated genetic material from non-African archaic humans.

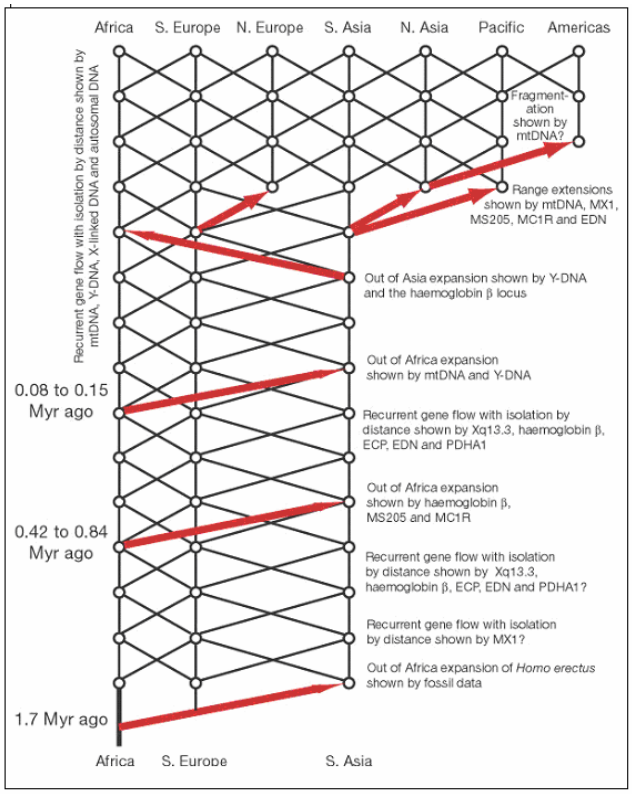

Figure 2. A model of recent human evolution. Major expansions of human populations are indicated by red arrows. Genetic descent is indicated by vertical lines, and gene flow by diagonal lines (Adopted from Templeton, 2002).

On the other hand, both the morphological analysis and the archaeological record indicated the early presence of modern populations in the Middle East and postulate an expansion from there into Europe in the upper Paleolithic (~ 45,000 years before the present (BP)) (Simoni *et al*., 2000) (Figure 3).

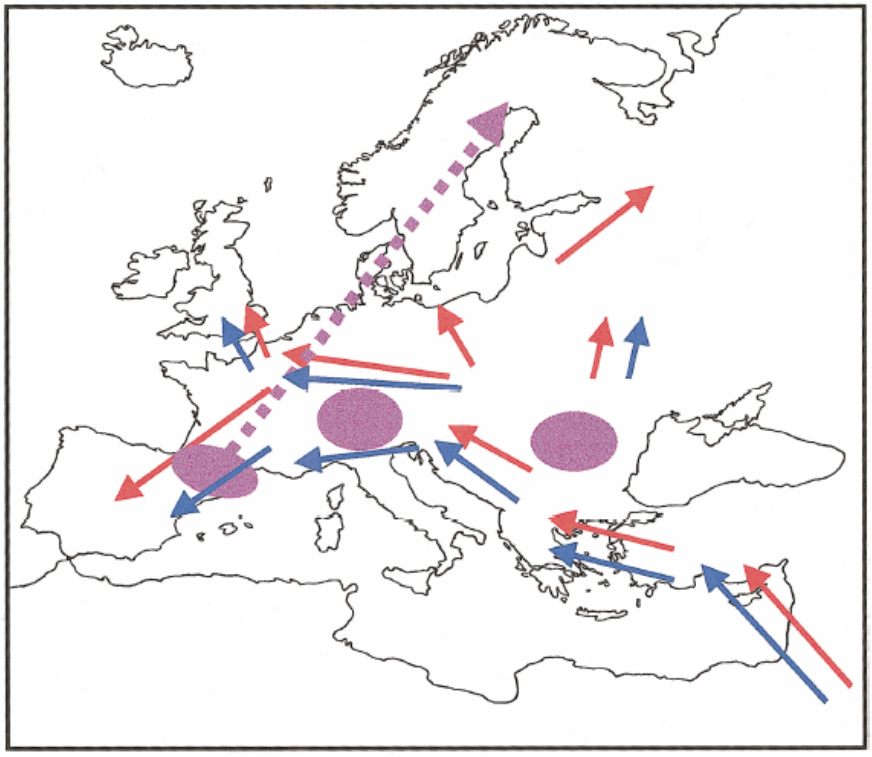

Figure 3. Expansion from Middle East into Europe. Scheme of the main dispersal processes supposed to have occurred during the Paleolithic first colonization of Europe (red arrows) and during the Neolithic demic diffusion (blue arrows). The approximate location of glacial refugia is represented (violet ovals), as is a major Mesolithic expansion from Iberia (dashed violet arrow) (Adopted from Simoni *et al*., 2000).

It was also discovered from the analysis of genetic data (two hyper variable sequence segments in the control region of mtDNA) that several major demographic expansions have influenced the Asian & Balkan genetic make-up as a result of the first introduction of modern humans into Europe. Demographic expansion occurred sequentially in the Middle East, through Turkey between 50,000 and 100,000 years ago in the upper Paleolithic (Calafell *et a*l., 1996).

Other more recent expansions and invasions known to have also affected Turkey is in Neolithic era (between 10,000-5,000 years BP). The Neolithic lifestyle originated in Western Asia, in the area known as "Fertile Crescent" and Anatolia was the bridge from which the new subsistence pattern spread to Europe. Farming spread in the Neolithic era from the Middle East and Anatolia through the Balkans and into Europe by "demic diffusion" model, where the new farming technology is brought into the territory by incomers. Both a Paleolithic expansion and the Neolithic demic diffusion of farmers could have determined a longitudinal cline of mtDNA diversity (Simoni *et al*., 2000).

Much more recently, in the 11th century A.D Turkic nomadic people from Central Asia occupied the grassland in the Asia Minor, imposing a language of the Turkic group. The linguistic and political consequences of these episodes are well documented (Renfrew, 1992) but it is not clear to what extent the Anatolian gene pool was effected by Oghuz invasion. The proportion of gene flow associated with the arrival of Central Asian Turkic speaking people to Anatolia were estimated as 30 % based upon mtDNA control region sequences and six microsatellite loci of the Y chromosome and one *Alu* insertion (YAP) (Benedetto *et al*., 2001).

In this context, study on the genetic diversity of Anatolian population can be used to estimate the Central Asia contribution to the Anatolia and expansion from Anatolia to the Europe.

Anatolia has been became a subject of many studies based on not only mitochondrial (Comas *et al*., 1996; Calafel *et al*., 1996; Benedetto *et al*., 2001) and nuclear genome

(Dinç, 2003, Comas *et al*., 2004) but also Y chromosome (Benedetto *et al*., 2001; Cinnioğlu *et al*., 2004).

The mitochondrial genome offers a very different perspective on human evolution. Because the mtDNA is inherited only through the maternal cytoplasm, variation in mtDNA sequence provides a record of the maternal lineage of our species. In contrast to the autosomal DNA, there is no recombination in mtDNA. Hence, the difference between any two mitochondrial sequences represents only the mutations that have taken place since each sequence was derived from a common ancestor. The noncoding mitochondrial control region is especially rich in polymorphisms, and the mutation rate is estimated to be approximately 2 to 3.10^{-7} per nucleotide per generation. Conversely, Y chromosome DNA provides a record of the paternal lineage.

Human nuclear genome has many different types of polymorphisms including classical systems (blood groups and protein polymorphisms), single nucleotide variants, and restriction fragment length polymorphisms (RFLPs). It was estimated that the mutation rate for single nucleotides is approximately 10^{-7} to 10^{-8} per generation. The slow mutation rate of these polymorphisms provides a means of visualizing the ancient history of our species, but these polymorphisms are likely to be relatively uninformative about recent history (Jorde *et al*., 1998).

The term “repetitive element” describes various DNA sequences that are present in multiple copies in the genomes in which they reside. Repetitive elements can be subdivided into those that are tandemly arrayed (for example, microsatellites, minisatellites and telomeres) or interspersed (for example mobile elements and processed pseudogenes) (Deininger & Batzer, 2002). Another class of nuclear polymorphisms consists of tandemly repeated DNA sequences in which the repeated DNA sequence is two to five base pairs in length. These are termed microsatellite systems, and the number of repeats in each system varies from one individual to the next. Their mutation rate is much higher than that of single nucleotides, approaching

10^{-3} per generation. Because of this high mutation rate, microsatellites have the potential to provide information about recent evolutionary events (Jorde *et al*., 1998).

Polymorphic mobile element insertions should more accurately reflect population relationships than many other genetic markers. Over million of year's primate genomes are prepared with a variety of repetitive sequences (elements). There are mobile sequences which are able to move from their parent genomic location to their target; either by a cut and paste mechanism involving a DNA or RNA intermediate. Table 1 summarizes mobile elements in the human genome (Deininger & Batzer, 2002).

Table 1. Summary of mobile elements in the human genome.

Element	Percent of total genome	Copy number
L1 (LINE)	16.9	0.5×10^6
Alu (SINE)	10.6	1.1×10^6
L2 (LINE)	3.2	0.3×10^6
MIR (SINE)	2.5	0.46×10^6
LTR elements	8.3	0.3×10^6
DNA elements	2.8	0.3×10^6
Processed pseudogenes	<1.0	$1–2 \times 10^4$
Total	~45	$\sim 3 \times 10^6$

(Adopted from Deininger & Batzer, 2002).

Over 40% of the human genome is derived from retroelements, with an additional few percent as DNA transpose elements (Deininger & Batzer, 2002). Among these LINE (L1), a type of long interspersed element and *Alu*, a type of short interspersed

element (SINE) are most common. LINE or SINE has some advantages over other nuclear based polymorphisms for in phylogenic studies. Firstly, the presence of an individual is thought to represent identity by descent (IBD), since the probability that two different young mobile elements integrating independently in the same chromosomal location is negligible (Deininger & Batzer, 2002). Polymorphic mobile element insertions should thus more accurately reflect population relationships than many other genetic markers (sequence data, restriction fragment length, polymorphisms [RFLP] and microsatalites) in which sharing of the same allele between two individuals may reflect identity by state only. *Alu* or L1 element insertion probably is a stable marker, because it is likely that insertion occurred once during human evolution with no subsequent deletion of the element (Perna *et al*., 1992).

Another advantage of these genetic markers is that the ancestral state of an insertion polymorphism is known to be the absence of the element at a particular genomic location (Batzer *et al*., 1994). Precise knowledge of the ancestral state of a genomic polymorphism allows us to draw trees of population relationships without making unnecessary assumptions (Deininger & Batzer 2002).

Although during the last million of years many *Alu* and L1 insertions have been fixed, still a number of human-specific insertions are dimorphic. An insertion may be present or absent on each of the paired chromosomes of different individuals. These polymorphisms, showing differences in allele and genotype frequencies between populations, help us reconstruct human prehistory (Stoneking *et al*; 1997, Comas *et al*., 2000; Nasidze *et al*., 2001; Watkins *et al*.2001; Romualdi *et* al., 2002).

In the present study the nine *Alu* insertion loci were examined as genetic markers. The *Alu* family of SINEs is one of the most successful mobile genetic elements to arise to a copy number of more than 500,000 within the human genome during approximately 65 million years of primate evolution (Batzer *et al*., 1996).

Alu sequences which are about 300 nucleotides in length are thought to be ancestrally derived from the 7SL RNA gene and are transcribed by RNA polymerase III. The RNA transcript is then reverse-transcribed and inserted into a new location in the genome, in a process termed retroposition. *Alus* have no protein coding capacity in that they are not able to code for their own reverse transcriptase. They are dependent on factors produced by Long Interspersed Elements (LINEs) for their amplification. LINEs have been shown to encode a functional reverse transcriptase that also has an endonuclease domain. The endonuclease activity shows a preferred cleavage at sites resembling TTT/AA and both L1 and *Alu* elements preferentially integrate at sequences related to this same consensus (El-Sawy and Deininger, 2005). It has been proposed that the T's at the 3' terminus of the nick prime reverse transcription from the polyadenylated RNA in a process termed "Target-Primed Reverse Transcription" (Cost *et al*., 2002) (Figure 4).

It may be also mentioned that *Alu* repeats have a relatively high GC content and, although dispersed mainly throughout the euchromatic regions of the genome, are preferentially located in the GC-rich and gene-rich regions. The tendency to be located in GC-rich (and accordingly, gene rich) regions suggest that *Alu* repeats are not just genome parasites, but are making a useful contribution to cells containing them (Strachan and Read, 2004).

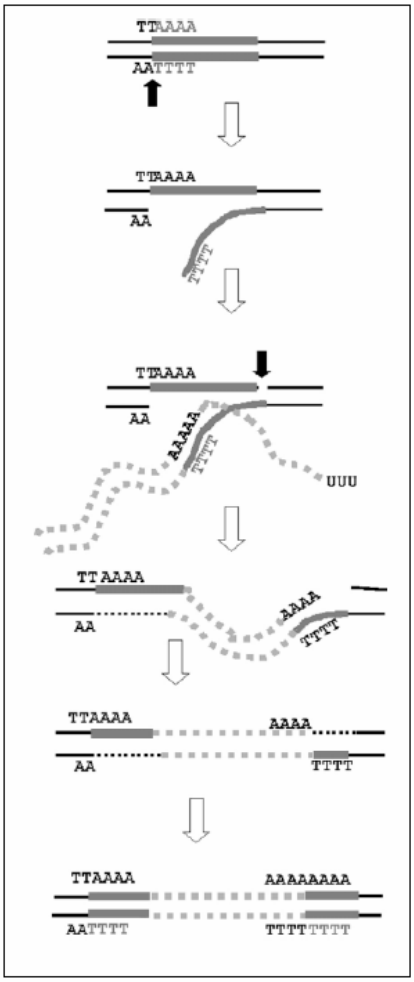

Figure 4. Mechanism of *Alu* retroposition. The solid bars represent the target site that is duplicated with insertion of the element. The endonuclease cleavage sequence (TTTT/AA) is shown and represents the insertion site. Nicking by LINE-1 endonuclease occurs (black arrow), followed by target primed reverse transcription of the *Alu* RNA (with its internally coded A stretch) with the T stretch at the 3' end of the nick, a process known as Target Primed Reverse Transcription (TPRT). The plus strand of the DNA is nicked (top black arrow) by an unknown mechanism. This is followed by integration of the *Alu* element and duplication of the target site, regenerating the original endonuclease cleavage site at the 5) end of the *Alu* (Adopted from Sawy & Deininger, 2005).

The analysis of human *Alu*-insertion polymorphisms has been used previously to adress questions about human origins and demography (Perna *et al*., 1992; Batzer *et al*., 1994, 1996; Stoneking *et al*., 1997; Comas *et al*., 2000; Watkins *et al*., 2001; Nasidze *et al*., 2001; Romualdi *et al*., 2002; Watkins *et al*., 2003; Comas *et al*., 2004). We present here an analysis of nine *Alu* insertion polymorphisms in Anatolian population. In one of the previous studies (Dinç, 2003) the same nine loci were examined on 100 Turkish individuals. In the present study an independent 100 individuals, all from Anatolia were studied. Comparative analyses were done between these two data from Anatolia and the ones from some other populations from Eurasia and Africa.

The purposes of this thesis are:

i- to determine the *Alu* insertion variability based on expanded data in the Anatolian human population,

ii- to reveal the genetic relatedness of Anatolians to some Eurasian populations,

iii- to provide an estimation of Central Asian contribution to the present Anatolian gene pool based on the *Alu* markers,

iv- to contribute to the knowledge on history of Anatolian gene pool.

CHAPTER 2

MATERIALS AND METHODS

2.1 MATERIALS

In this study, nine autosomal human specific "*Alu* insertion polymorphism" (*ACE, PV92, FXIIIB, APO, A25, B65, TPA25, D1, HS4.32*) loci were typed.

Hundred individuals who were residents of different Anatolian provinces were randomly sampled. Blood or buccal samples were obtained from unrelated volunteer donors whose families were from the same area for at least three generations. Informed consent was obtained from all of the individuals who were participating in the study. Geographic locations of the individuals analyzed were shown in Figure 5. and summarized in Table 2.

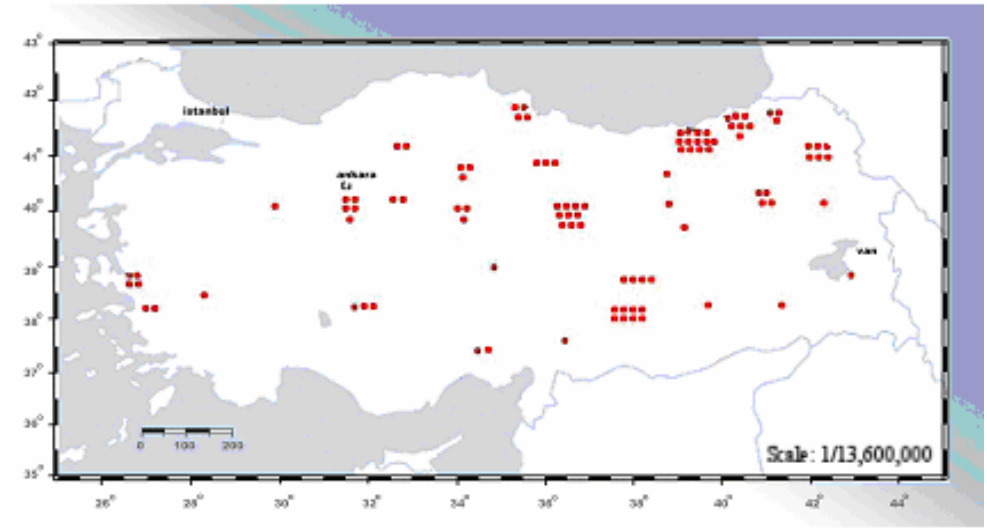

Figure 5. Geographic locations of the employed individuals. Dots indicate the birthplaces of individuals.

Table 2. Geographic locations and number of the employed individuals.

Location	Individuals analyzed	Location	Individuals analyzed
İzmir	4	Ağrı	1
Artvin	3	Erzurum	4
Gaziantep	1	Sivas	10
Rize	6	Kars	6
Trabzon	13	Yozgat	3
Adıyaman	8	Kırıkkale	2
Malatya	4	Çorum	3
Samsun	4	Kayseri	1
Adana	2	Van	1
Gümüşhane	1	Eskişehir	1
Çankırı	2	Denizli	1
Caykara	1	Siirt	1
Konya	3	Tunceli	1
Tokat	3	Diyarbakır	1
Aydın	2	Erzincan	1
Ankara	5	**Total**	**100**

2.2 METHODS

2.2.1 DNA Isolation

Human genomic DNA was isolated from 100 individuals, in the form of 75 buccal and 25 blood samples, by employing two different isolation methods depending on the type of the samples.

DNA isolation from human blood samples were in accordance with the classical phenol-chloroform-isoamylalcohol method. DNA was also isolated from the buccal samples this time by employing Dinç's method (Dinç, 2003).

2.2.1.1 DNA Isolation from Blood Samples

Ten ml of blood samples were collected in the tubes containing EDTA solution and stored at +4 ºC until ready to use. Lysis buffer had been added to these tubes until total solution was 50 ml. Lysis buffer was used to lyse red blood cells and to get the remaining nuclei of white blood cells. Afterwards, tubes were mixed for 10 minutes by inversion. Then the remaining nuclei of white blood cells were obtained as a pellet after the centrifugation at 3000 rpm for 10 minutes at +4 ºC. The supernatants were carefully discarded and the pellet was resuspended in 3 ml of Salt/EDTA buffer (NaCl and EDTA) by vortex. Later, 0.3 ml of 10 % SDS and 150 µl proteinase-K (10 mg/ml) were added and then the tubes were incubated at +55 ºC for 3 hours.

After the incubation, solution was cooled down to the room temperature. Then 3 ml phenol was added to separate DNA molecules from RNA molecules. This separation was due to their different solubility properties in this solvent. Then the samples were centrifuged at 3000 rpm for 10 minutes at +4 ºC. After that, two phases were obtained in our tubes. The supernatant was mixed with 3 ml phenol-chloroform-isoamylalcohol solution (25:24:1) and again centrifuged at 3000 rpm for 10 minutes at +4 ºC. Then the upper phase containing DNA molecules was carefully collected and transferred into a new glass tube by a transfer pipette. Finally, 1 ml 3 M Sodium Acetate (NaAc) and about 2 volumes of 99 % ice-cold ethanol were added to this

solution. Glass tubes were mixed gently by inversion. In the mean time, because the alcohol molecules compete with those of the water molecules surrounding the DNA, DNA precipitates. Precipitated DNA was transferred to ependorf tubes containing 500 µl of Tris EDTA (TE) buffer (pH: 7.5) and stored at –20 ºC.

2.2.1.2 DNA Isolation from Buccal Samples

A technique, having two overnight incubations at different temperatures was used to isolate the DNA from the buccal samples. Firstly, epithelial cells were collected from the inside of both cheeks with a sterile brush. Then the head of the brush was placed into a 1.5 ml ependorf tube containing 700 µl 2X Lysis buffer and 10 µl proteinase-K (20 mg/ml) was added and the cells were digested overnight at 37 ºC.

After the digestion, 500 µl of phenol-chloroform-isoamylalcohol solution (25:24:1) was added to the ependorf tubes. Then this suspension was centrifuged at 10,000 rpm for 2 minutes at room temperature (RT), and the upper aqueous phase had been transferred into a new tube, again 500 µl of phenol-chloroform-isoamylalcohol solution was added and suspension was centrifuged at 10,000 rpm for 2 minutes at room temperature. Then the second aqueous phase was transferred into a new tube and 500 µl of chloroform-isoamylalcohol solution (24:1) was added. Again the suspension was centrifuged at 10,000 rpm for 2 minutes at room temperature. After the third aqueous phase was collected into a new tube, 0.1 volume of 3 M Sodium Acetate (NaAc) and 0.6 volumes of isopropanol were added into the solution. The tube was incubated overnight at – 20 ºC.

After the incubation, solution was firstly centrifuged at 10,000 rpm for 10 minutes. Then, the pellet was rinsed with 500 µl 70 % ethanol and the DNA was precipitated by centrifuging at 10,000 rpm for 5 minutes at RT. The supernatant was discarded and the pellet was dried at RT and then resuspanded in 100 µl of Tris-EDTA (TE) buffer. DNA samples were stored at –20 ºC until used.

2.2.2 Amplification of DNA with Polymerase Chain Reaction (PCR)

Each of the isolated DNA was amplified by 9 different pairs of primers; each of the pair was complementary to a different *Alu* insertion loci. These loci correspond to nine human-specific *Alu* insertion polymorphisms (*ACE, APO, TPA25, FXIIIB, PV92, B65, A25, D1, HS4.32*). PCR amplification was carried out in 25 µl reaction mixture which is described below:

- 1X PCR Buffer (50mM KCl and 10mM Tris-HCl)
- 1 mM BSA (10mg/ml)
- 1.5 mM MgCl2 for *ACE, APO, FXIIIB, HS4.32, HS4.69, PV92, TPA25* and 4 mM MgCl2 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 DNAdH2OSequences of primers, chromosomal locations of the loci and the optimal annealing temperatures of the primers are shown in Table 3.

Amplification of DNA samples for *ACE, APO, PV92* and *A25* loci were performed as described previously (Dinç, 2003) shown in Table 4. However for other loci, amplification conditions were modified as indicated: 1 minutes at 94 ºC, 2 minutes at an appropriate annealing temperature, 2 minutes at 72 ºC during 35 cycles, with a final elongation step of 72 ºC for 10 minutes (Chbel et al., 2003). Also, a "Hot start" was applied in accordance with the protocol with an addition of a pre-PCR heat cycle (Ueda et al., 1996).

Table 3. Oligonucleotide primers, their annealing temperatures (Tm), and locations of polymorphic *Alu* loci; *TPA25, PV92, FXIIIB, D1, APO, ACE,* as described in Batzer *et al.* (1996); *B65, A25* Arcot *et al.*

Alu loci	FORWARD	REVERSE	Tm	ChromosomalLocation
TPA25	5'-GTAAGAGTTCCGTAACAGGACAGCT-3'	5'-CCCCACCCTAGGAGAACTTCTCTTT-3'	58	Intron 8 of PLAT.
PV92	5'-AACTGGGAAAATTTGAAGAGAAAGT-3'	5'-TGAGTTCTCAACTCCTGTGTGTTAG-3'	54	On chromosome 16.
FXIIIB	5'-TCAACTCCATGAGATTTTCAGAAGT-3'	5'-CTGGAAAAAATGTATTCAGGTGAGT-3'	56	Intron 10 of gene F13B.
D1	5'-TGCTGATGCCCAGGGTTAGTAAA-3'	5'-TTTCTGCTATGCTCTTCCCTCTC-3'	70	On chromosome 3.
APO	5'-AAGTGCTGTAGGCCATTTAGATTAG-3'	5'-AGTCTTCGATGACAGCGTATACAGA-3'	50	Adjacent to theAPOA1, APOC3, APOA4 cluster on chromosome 11.
ACE	5'-CTGGAGACCACTCCCATCCTTTCT-3'	5'-GATGTGGCCATCACATTCGTCAGAT-3'	58	Intron 16 of ACE gene.
B65	5'-ATATCCTAAAAGGGACACCA-3'	5'-AAAATTTATGGCATGCGTAT-3'	60	On chromosome 11.
A25	5'-CCACAAATAGGCTCATGTAGAAC-3'	5'-TATAATATGGCCTGGATTATACC-3'	63	On chromosome 8.
HS4.32	5'-GTTTATTGGGCTAACCTGGG-3'	5'-TGACCAGCTAACTTCTACTTTAACC	57	On chromosome 12

Table 4. PCR conditions for *ACE, APO, PV92* and *A25* loci.

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

(Adopted from Dinç, 2003)

2.2.3 Analysis of the PCR products

After obtaining the PCR products 9 µl of PCR products mixed with 7 µl of 6 X bromophenol blue dyes were separated on 2 % TBE agarose gel. The gel was run in 0, 5 X TBE buffer at 100 V until the bands reach end to the gel. Then it was stained with 0,5 µg/ml of ethidium bromide and the reaction products were directly visualized under the UV light. The sites of the PCR products are given in Table 5.

Nine autosomal human specific *Alu* insertion polymorphisms were analyzed by means of the agarose gel electrophoresis of the PCR products. Since all of the alleles were different approximately by 300 bp the method employed in the study allowed us to type the three different genotype of nine *Alu* loci easily.

Table 5. Sizes of the PCR products depending on the presence and absence of *Alu* insertion.

Alu loci	Insertion * (bp)	No insertion ** (bp)
ACE	490	190
APO	433	122
A25	552	268
B65	423	81
D1	622	333
FXIIIB	725	425
HS4.32	601	289
HS4.69	572	262
PV92	416	101
TPA25	424	125

* insertion positive (+); ** insertion negative (-) (Dinç, 2003).

2.2.4 Statistical Analysis of Data

2.2.4.1 Allele Frequencies, Heterozygosity and Evaluation of Hardy-Weinberg Equilibrium

Allele frequencies among nine of the *Alu* loci for 100 individuals, in Anatolian population were estimated with the calculation of the relative proportions of each *Alu* insertion polymorphism by means of the following formula;

Frequency of an allele (p) = [(2 x number of homozygote) + (number of heterozygote)] / 2 x total number of individuals

Heterozygosity is a measure of the diversity of a polymorphic locus; for a diploid locus, it gives the average probability that the alleles carried by an individual are different from each other (Jobling *et al*., 2004). The amount of heterozygosity, determined by the percentage of individuals that are heterozygous (carry different alleles) for a population was calculated with the following formula:

Observed Heterozygosity = number of heterozygotes / total number of individuals

Both the allele frequencies of each *Alu* loci and the amount of observed heterozygosities were calculated by Genetix 4.0 software (Belkhir *et al*., 1996-2001).

After finding the frequencies of each *Alu* insertion polymorphism, expected frequencies of the genotypes were calculated accordance to the Hardy-Weinberg Equilibrium (1908).

Thus, the expected frequencies of genotypes were calculated as follows:

Expected proportion of homozygote : p^2+q^2

Expected proportion of heterozygotes: $2pq = 1-(p^2+q^2)$

where $p+q = 1$

The expected genotype frequencies; expected heterozygotes calculated based on the Hardy-Weinberg Equilibrium and the observed number of each genotype for each of the *Alu* loci were compared to test whether observed genotype frequencies were sufficiently close to the expected ones. In order to determine if population is in the expected H-W genotype frequencies, Chi square (χ^2) test was used:

$$\chi^2 = \Sigma^k_{i=1} (O - E)^2 / E$$

where O and E are the observed and expected numbers of a particular genotype and k is the number of genotypic classes (Hendrick, 1985). Calculated value of chi square for each *Alu* insertion were evaluated based on chi square distribution with the degrees of freedom equal to one and $\alpha = 0.05$ (df=1, $\chi^2_{0.95}$=3.841).

2.2.4.2 Test for the equality of frequencies

To test the equality of two populations, z-test presented by Daniel (1999) was used. To test the null hypothesis (Ho) that the differences between two population proportions is zero, two sided test was applied.

2.2.4.3 The F_{ST} Statistic

Wright (1951, 1965b) developed an approach, consisting of three different F coefficients (F_{ST}, F_{IT}, F_{IS}), for describing the distribution of genetic variation in a subdivided population.

F_{ST} is a measure of the genetic differentiation of subpopulations. The average F_{ST} coefficient can be calculated for multiple loci as follows:

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

where H_S is the average expected heterozygosity within subpopulations over loci and H_T is the average expected heterozygosity in the total population over loci (Hendrick, 1993).

For some of the populations *Alu* polymorphism data were already available. F_{ST} related genetic distance between Anatolia and populations was estimated by means

of the GENETIX 4.0 software (Belkhir *et a*l., 1996-2001). Data were permuted for 1000 times in order to test the significance of the pair-wise F_{st} values.

In this study, Weir & Cockerham's approach was employed. Weir and Cockerham (1984) presented equations for estimating F statistics in a variety of situations, including the case of multiple alleles and loci.

2.2.4.4 Reynold's Genetic Distance and Neighbor Joining (NJ) Tree

Reynold's Genetic Distances (Reynolds *et al*.1983) were computed between Anatolian population and the populations studied by Stoneking *et al*. (1997), Comas *et al*. (200), Nasidze *et al*. (2001), Romualdi *et al*. (2002), Khitrinskaya *et al*. (2003), Dinç (2003) and Comas *et al*. (2004) by means of the GENDIST program in the PHYLIP package program (Felsenstein, 1993). Neighbor Joining Tree building method, one of the most popular distance-based methods, was used to define the similarity between populations. In the Neighbor Joining Tree building method, after the estimation of the genetic distance network the tree is constructed by using an algorithm based on some functional relationships among the distance values (Saitou & Nei, 1987). Neighbor joining trees were constructed by using the NEIGHBOR program and 1000 bootstrap replicates were generated by the SEQBOOT program and consensus tree was built with CONSENSE program as implemented in PHYLIP 3.6 (Felsenstein, 1993).

2.2.4.5 Principal Component Analysis (PCA) of gene frequencies

Principal Component Analysis (PCA) was performed to generate two or three-dimensional graphical representation of distances between populations using the raw data of allele frequencies by means of the NTSYS (Numerical Taxonomy and Multivariate Analysis System) (Rohlf, 1993).

Principal Component Analysis is a useful complement to tree construction when frequency data are being considered, as it, unlike a tree, does not impose a history of bifurcations on contemporary population structure. Rather, it shows genetic distance between populations as Euclidean distance in two or three dimensions (Sherry & Batzer, 1997). Individual axes, known as principle components (PCs), are extracted sequentially, with each PC encapsulating as much of the remaining variation as possible (Jobling *et al*., 2004).

2.2.4.6 Estimating Admixture

Genetic admixture is the process by which a hybrid population is formed from contributions by two or more parental or ancestral populations (Jobling et al., 2004). The process of mixing of the populations known as "admixture", specifically refers to the process in which populations that have remained isolated from one another for a long period because of geographical, ecological, or cultural barriers are occasionally brought into contact or it refers a special case of gene flow between populations (historical migration) (Bertorelle & Excoffier, 1998).

In the present study we estimate the relative genetic contribution of the parental: Balkans and Central Asian populations to the current Anatolian gene pool (admixed population). Assuming that all gene flow from Central Asia into Turkey occurred at one moment in time, the relationship between the allele frequencies in Turkey, Europe and Central Asia is:

$$q_T = (1 - mR)\, q_E + mRq_A$$

where mR is the only unknown term, and can be defined as the instantaneous gene-flow rate, the amount of immigration necessary for q_E to reach q_T in one generation's time.

CHAPTER 3

RESULTS

As an example of typical gel photography, the one belonging to Alu element insertion polymorphism is located in intron 8 of gene PLAT was shown in Figure 5. Three possible genotypes for the TPA25 Alu insertion polymorphism were indicated in Figure 2. For instance in lane C3 an individual homozygote for the presence of TPA25 (424 bp product only), in lane C9 an individual homozygote for the absence of TPA25 (125 bp product only), and in lane C12 a heterozygote (both the 424 bp and the 125 bp products) individual.

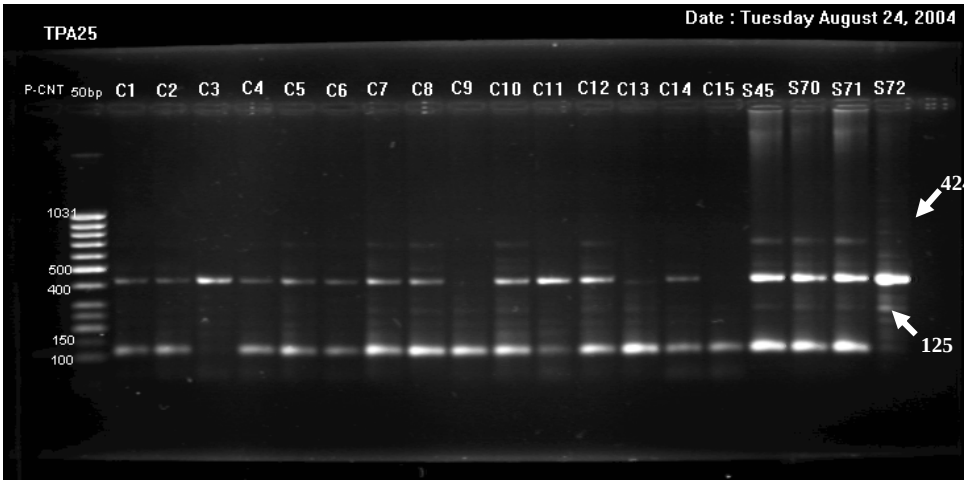

Figure 6. Photograph of a 2 % agarose gel containing the PCR products of TPA25.
P-CNT: PCR control. 50 bp: 50 bp DNA ladder, Fermentas.
Lanes: C3, S72: Homozygous individuals for the presence of TPA25 insertion (+/+)
C1, C2, C4, C5, C6, C7, C8, C10, C11, C12, C14, S45, S70, S71: Heterozygous individuals (+/-)
C9, C15, C13: Homozygous individuals for the absence of TPA25 insertion (-/-).

3.1 FREQUENCIES OF NINE *ALU* LOCI IN ANATOLIA

Nine *Alu* loci (*ACE, APO, PV92, TPA25, FXIIIB, B65, A25, D1, HS4.32*) were typed and allele frequencies were estimated, in 100 individuals from Anatolian population. Allele frequencies for nine *Alu* loci and sample sizes were employed for each of the loci shown in Table 6.

Table 6. *Alu* insertion frequencies in the Anatolian population.

Alu loci	ACE	APO	PV92	TPA25	FXIIIB	B65	A25	D1	HS4.32
Alu insertion frequencies	0.410	0.963	0.220	0.390	0.579	0.667	0.067	0.427	0.637
N*	100	95	84	100	89	84	97	82	73

*****Number of individuals tested for each locus.

All loci were polymorphic in the Anatolian population and it was found that there is a good fit between observed genotype frequencies and Hardy-Weinberg expected proportions, in all of the loci. Employing Chi square (χ^2) test, it was decided that the observed and expected frequencies are not different from each other. Hence the population was found to be in Hardy-Weinberg equilibrium ($p>0.05$) for all of the loci. The expected genotype numbers, expected heterozygosities and χ^2 values can be seen in Table 7.

Each of the individuals' genotypes and their geographic locations were indicated on the maps (Figures 7-14). It was found that the locus called *APO* had the highest insertion frequency compared to other markers in Anatolian population. Homozygous individuals for *APO* insertion were observed nearly in every sampled region (Figure 10). Conversely, *A25* loci had the lowest insertion frequency. Unlike *APO,* an *A25* locus was observed to have homozygous individuals for the absence of insertion, in all over the Anatolia (Figure 11).

Table 7. Statistics of nine *Alu* loci (*ACE, APO, PV92, TPA25, FXIIIB, B65, A25, D1, HS4.32*) in Anatolia.

Alu Insertion	N	Frequency of the insertion	Genotype	Observed number	Expected number	X^{2*}	Observed Heterozygosity	Expected Heterozygosity
ACE	100	0.4100	+/+	14	17	0.0133	0.5400	0.4838
			-/-	32	35			
			+/-	54	48			
APO	95	0.9632	+/+	88	88	0.0014	0.0737	0.0709
			-/-	0	0			
			+/-	7	7			
PV92	84	0.2202	+/+	6	4	0.0178	0.2976	0.3435
			-/-	53	51			
			+/-	25	29			
TPA25	100	0.3900	+/+	11	15	0.0313	0.5600	0.4758
			-/-	33	37			
			+/-	56	48			
FXIIIB	89	0.5787	+/+	30	30	0.00008	0.4831	0.4876
			-/-	16	16			
			+/-	43	43			
B65	84	0.6667	+/+	35	37	0.0155	0.5000	0.4444
			-/-	7	9			
			+/-	42	37			
A25	97	0.0670	+/+	0	0	0.0105	0.1340	0.1250
			-/-	84	84			
			+/-	13	12			
D1	82	0.4268	+/+	12	15	0.0192	0.5610	0.4893
			-/-	25	27			
			+/-	46	40			
HS4.32	73	0.6370	+/+	35	30	0.0932	0.3287	0.4625
			-/-	15	10			
			+/-	24	34			
Average Heterozygosities							0.3864	0.3759

*X^2 values < 3.84 ($X^2_{1,0.95} = 3.84$) were considered to be nonsignificant.

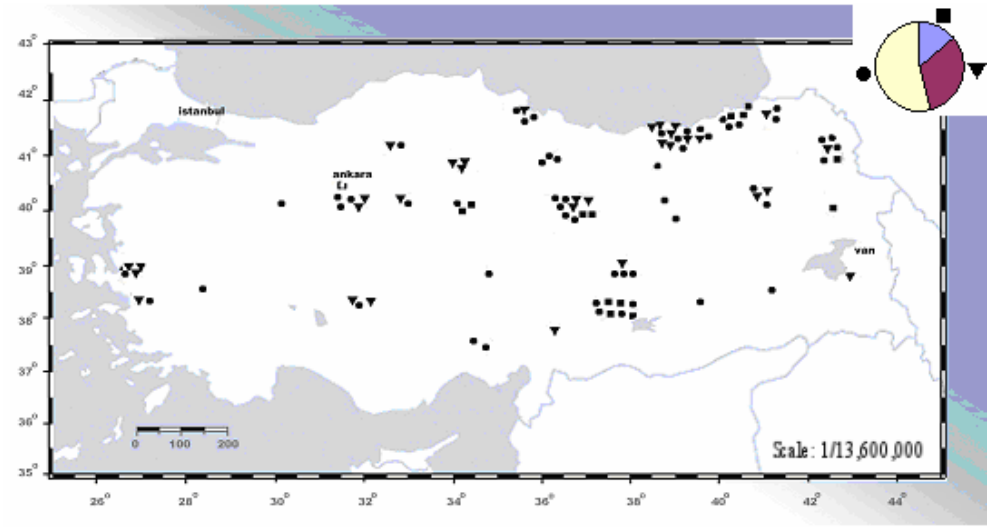

Figure 7. The distribution pattern of *ACE* insertion in 100 individuals from Anatolia.

■ **:** Homozygous individual for the *ACE* insertion (+/+),
● : Heterozygous individual (+/-),
▼ : Homozygous individual for the lack of *ACE* insertion (-/-).

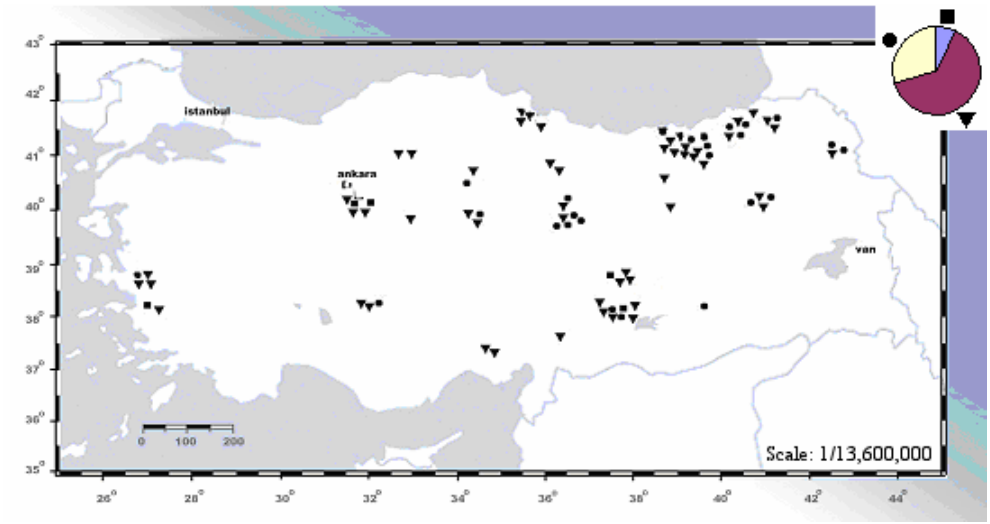

Figure 8. The distribution pattern of *PV92* insertion in 84 individuals from Anatolia.

■ **:** Homozygous individual for the *PV92* insertion (+/+),
● : Heterozygous individual (+/-),
▼ : Homozygous individual for the lack of *PV92* insertion (-/-).

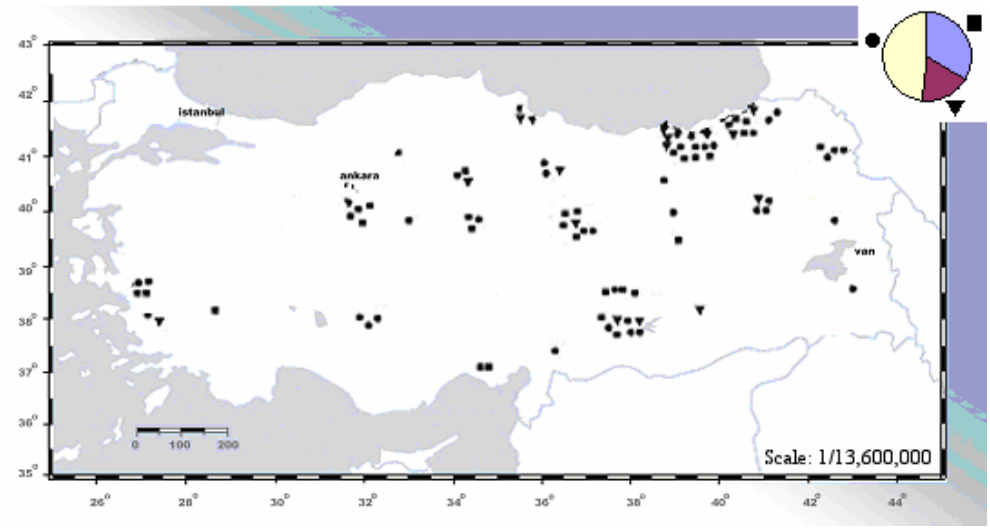

Figure 9. The distribution pattern of *FXIIIB* insertion in 89 individuals from Anatolia.

■ **:** Homozygous individual for the *FXIIIB* insertion (+/+),
● : Heterozygous individual (+/-),
▼ : Homozygous individual for the lack of *FXIIIB* insertion (-/-).

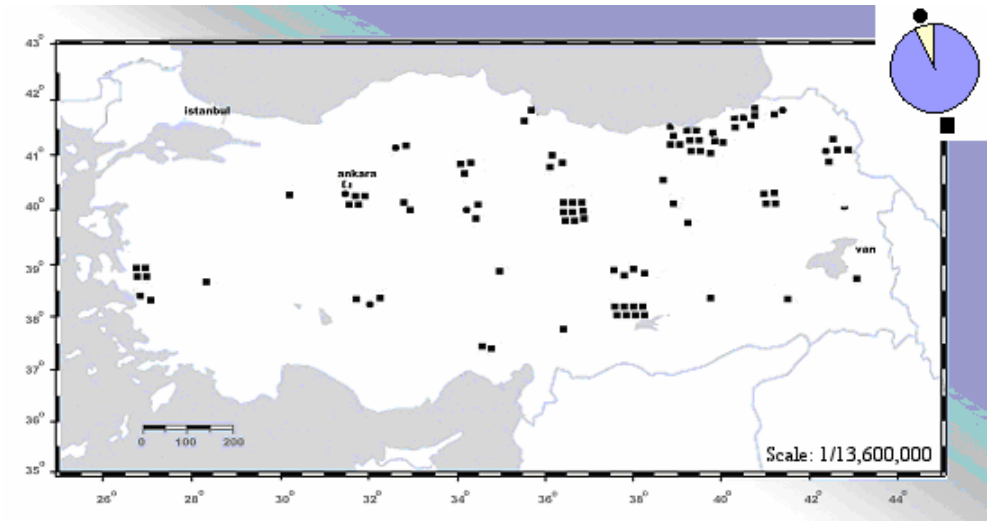

Figure 10. The distribution pattern of *APO* insertion in 95 individuals from Anatolia.

■ **:** Homozygous individual for the *APO* insertion (+/+),
● : Heterozygous individual (+/-),
▼ : Homozygous individual for the lack of *APO* insertion (-/-).

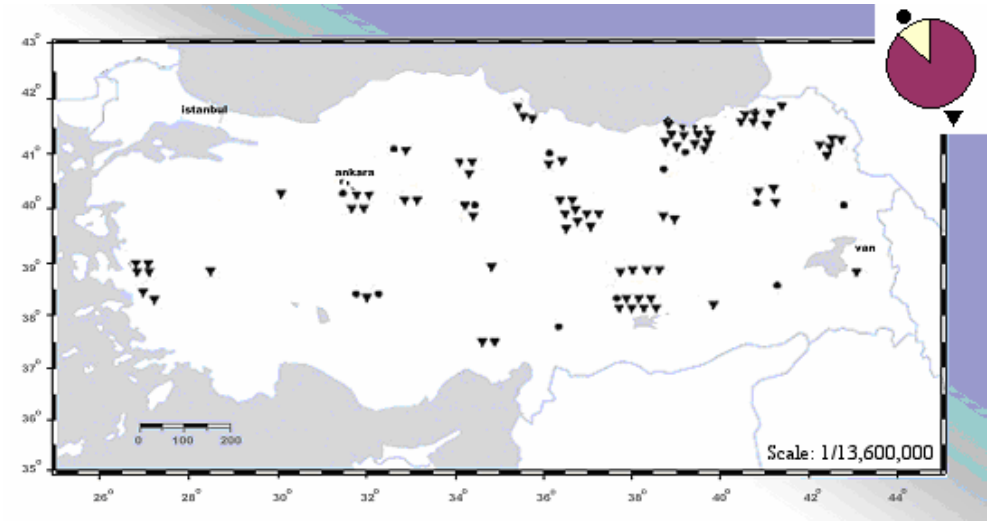

Figure 11. The distribution pattern of *A25* insertion in 97 individuals from Anatolia.

■ **:** Homozygous individual for the *A25* insertion (+/+),
● : Heterozygous individual (+/-),
▼ : Homozygous individual for the lack of *A25* insertion (-/-).

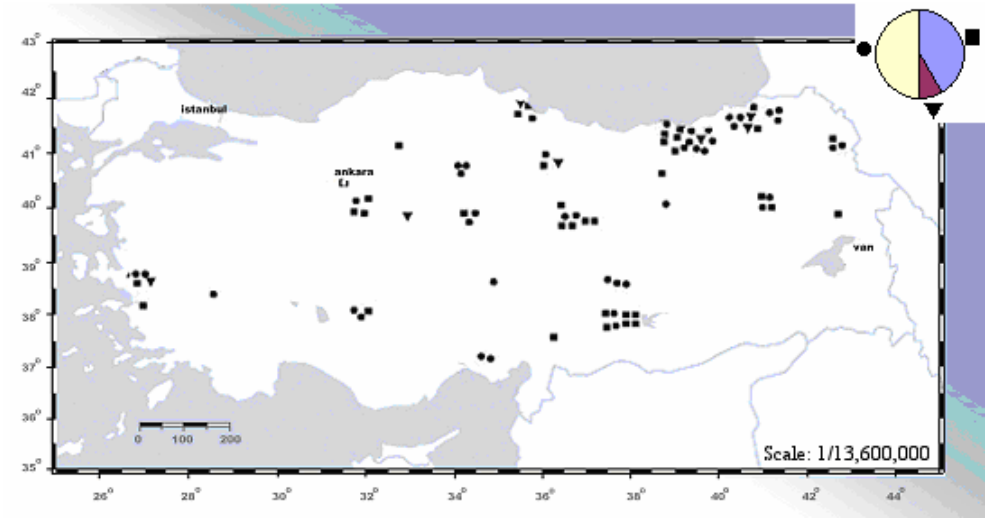

Figure 12. The distribution pattern of *B65* insertion in 84 individuals from Anatolia.

■ **:** Homozygous individual for the *B65* insertion (+/+),
● : Heterozygous individual (+/-),
▼ : Homozygous individual for the lack of *B65* insertion (-/-).

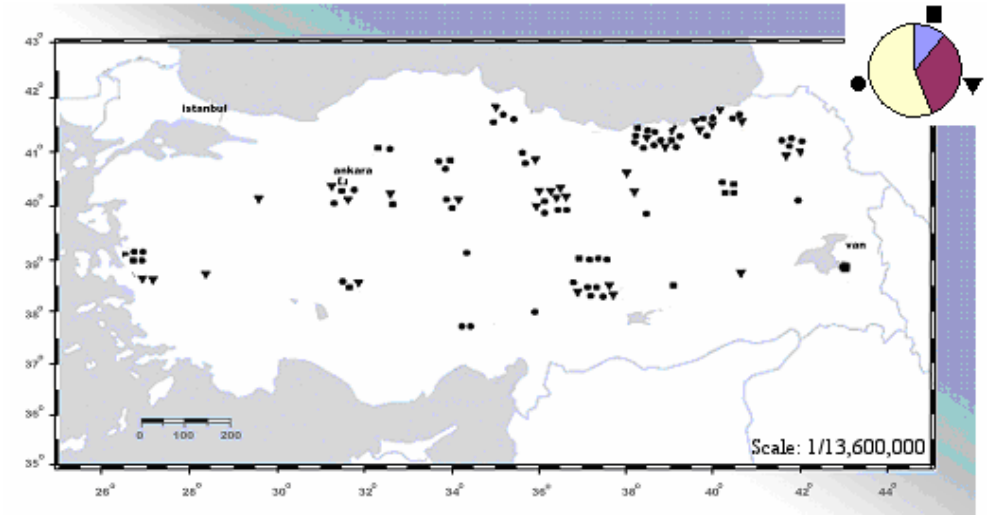

Figure 13. The distribution pattern of *TPA25* insertion in 100 individuals from Anatolia.

■ **:** Homozygous individual for the *TPA25* insertion (+/+),
● : Heterozygous individual (+/-),
▼ : Homozygous individual for the lack of *TPA25* insertion (-/-).

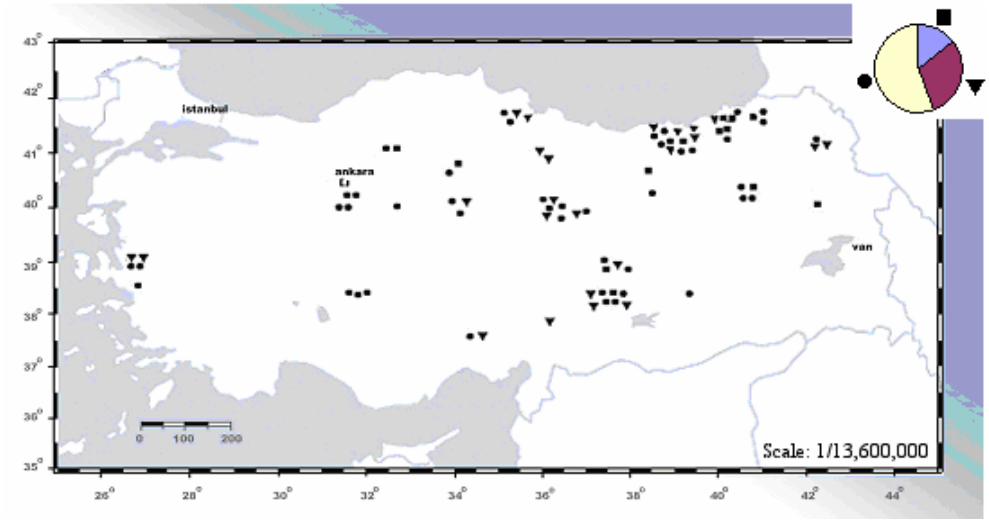

Figure 14. The distribution pattern of *D1* insertion in 82 individuals from Anatolia.

■ **:** Homozygous individual for the *D1* insertion (+/+),
● : Heterozygous individual (+/-),
▼ : Homozygous individual for the lack of *D1* insertion (-/-).

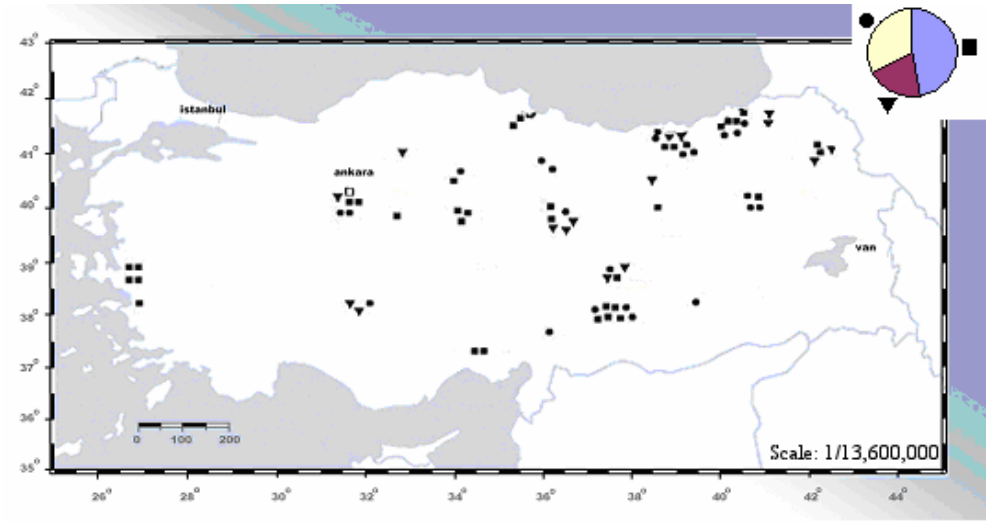

Figure 15. The distribution pattern of *HS4.32* insertion in 73 individuals from Anatolia.

■ **:** Homozygous individual for the *HS4.32* insertion (+/+),

● : Heterozygous individual (+/-),

▼ : Homozygous individual for the lack of *HS4.32* insertion (-/-).

3.2 PRESENT ANATOLIAN DATA TOGETHER WITH THE DATA FROM DIFFERENT HUMAN POPULATIONS

To compare the genetic similarity of the Anatolian population with these of the other populations it was focused on seven of the nine *Alu* loci (*ACE, APO, A25, B65, FXIIIB, PV92, and TPA25*) which were jointly employed by Stoneking *et al*. (1997), Comas *et al.* (2000), Nasidze *et al*. (2001), Romualdi *et al*. (2002), Khitrinskaya *et al*. (2003) and Dinç (2003), Comas *et al*. (2004). Insertion frequencies for seven *Alu* loci in many populations were presented together with the results of the present data in Appendix B.

3.2.1 The Fst Statistics

To determine the degree of similarity between Anatolian population and some of the populations from Europe, Balkans, and Asia and Near East, Fst related genetic distances based on seven *Alu* loci were computed (Table 8). In the table pair-wise Fst values ranged from -0.00171 to 0.22918. The pair-wise Fst values between Anatolian population typed in this study and Anatolian population studied by Comas et al., (2004), Anatolian population and Cherkessians, Anatolian population and Georgians, Anatolian population and Albanians, Anatolian population and Macedonians, Anatolian population and Macedonian-Stip–Aromuns, Anatolian population and Albanian-Aromuns were not significantly different from each other based on the seven *Alu* loci. However, there was significant Fst based genetic distance between Anatolian population and with those of the Central Asian populations. Since the Anatolian population sample of the recent study (A1) was not significantly different than that of Comas et al., (2004)'s population (A13) they were pooled and referred as Anatolian population (A13).

3.2.2 The Reynolds' Genetic Distances between populations

Using the seven *Alu* insertion frequencies, Reynold's genetic distances (Reynolds *et al*.1983) between the populations of our data set are computed by means of the Gendist Program in the PHYLIP Program package (Felsenstein, 1993). Genetic distances between Anatolian population and the other worldwide populations are shown in Table 9. The lowest genetic distance was between Anatolian population studied in our study and Anatolian population studied by Comas et al., (2004). The smallest distance between the Anatolian population (present study) and among the ones compared was as follow: Anatolian population and Cherkessian, Anatolian population and Georgians, Anatolian population and Albania Aromuns, Anatolian population and Macedonia Krusevo Aromuns. Pairs of Anatolian population and some of the Balkans populations (Albania Aromuns, Macedonia –Krusevo Aromuns,

Macedonians, Macedonians-Stip Aromuns, Romania Aromuns, and Albanians) had lower genetic distance than the pairs of Anatolian population and the other populations such as Central Asians.

3.2.3 Neighbor Joining (NJ) Tree

To visualize the genetic similarity of the Anatolian population with those of some other Eurasian and African populations, unrooted Neighbor Joining Tree (Saitou Nei, 1987) was constructed and shown in Figure 16.

From the tree it was observed that the populations from Europe, Middle East, Caucasians, Balkan and Basque formed a mixed group on the tree. Anatolian populations (A1, A2, and A3) were also in this group. However Central Asian populations were clearly in another group. They were observed to form a cluster with other Asian populations on the tree.

Thousand bootstrap replications were performed. The bootstrap values ranged from 8.3 % to 63.0 %, they were all low, only bootstrap values larger than 50 were depicted at the branching nodes on the tree.

Table 8. Fst values between Anatolian population and some other populations from Eurasia and Africa. Anatolians (A1): Anatolian population studied in the present study; Anatolian 3 (A3): Anatolian population studied by Comas *et al.*, (2004); Anatolian 2 (A2): Anatolian population studied by Dinç (2004).

Populations- A1/ other populations	Fst-values	
Anatolians-Anatolians 3	-0.00171	NS
Anatolians-Cherkessians	0.00009	NS
Anatolians- Georgians	0.00167	NS
Anatolians- Macedonia Stip Aromuns	0.00625	NS
Anatolians- Albanians	0.00644	NS
Anatolians- Macedonia	0.00753	NS
Anatolians- Albania Aromuns	0.00764	NS
Anatolians- Macedonia Krusevo Aromuns	0.01126	*
Anatolians- Romanian Aromuns	0.01319	*
Anatolians- Greeks	0.01687	*
Anatolians-Anatolians 2	0.0208	**
Anatolians- Basques	0.02148	**
Anatolians- Andalusians	0.02426	**
Anatolians- Romanians	0.02641	**
Anatolians- India Muslim	0.02992	**
Anatolians- Pustoon	0.03022	**
Anatolians- Catalans	0.03072	**
Anatolians- Turkish Cypriot	0.031	**
Anatolians- Uzbeks	0.03769	***
Anatolians- UAE	0.04191	**
Anatolians-Tajiks	0.05136	***
Anatolians- Armenians	0.05334	***
Anatolians- French	0.0556	***
Anatolians-Swiss	0.07271	***
Anatolians- Kyrgyzes	0.07873	***
Anatolians-Greek Cypriot	0.07947	***
Anatolians- India Hindu	0.08096	***
Anatolians- Bretons	0.0819	***
Anatolians- Syria	0.08599	***
Anatolians- Tamil	0.08723	***
Anatolians- Pakistan	0.0959	***

Table 8. (continued)

Anatolians- Azerbaijanians	0.10455	***
Anatolians- Kazakhs	0.10888	***
Anatolians- European American	0.11119	***
Anatolians- India-Christian	0.11342	***
Anatolians- Darginians	0.12115	***
Anatolians- German	0.1467	***
Anatolians- Dungas	0.1506	***
Anatolians- Ingushians	0.19682	***
Anatolians- China	0.19789	***
Anatolians- Taiwan	0.22918	***

NS : Nonsignificant
* $p<0.05$; ** $p<0.01$; *** $p<0.001$

Table 9. Genetic Distances between Anatolian Population (A1) and some other Eurasian and African populations. A1: Anatolian population studied in the present study; A2 and A3 are as described in Table 8.

Pairwise Genetic Distances			
Population	A1	Population	A1
A3	0.005897	Swiss	0.068271
Cherkess	0.007291	Dungans	0.070715
Georgians	0.011153	Greek Cypriot	0.071645
Albania Aromuns	0.014116	Tenggaras	0.07404
MacKrus-Aromuns	0.016857	Tamill	0.077179
Maceds	0.01837	Malay	0.077278
MacStip-Aromuns	0.018464	Syrians	0.077937
Romania Aromuns	0.020124	Bretons	0.082003
Albanians	0.021526	Kazakhs	0.084087
A2	0.024956	Chinese	0.085981
Basque	0.025719	Pakistan	0.089047
Tajiks	0.032961	Taiwan	0.101209
Uzbeks	0.03519	India-Christian	0.101653
Greeks	0.035383	Moluccan	0.106183
India-Muslum	0.036106	European American	0.107272
Pushtoon	0.03683	Azerbaijn	0.11244
Romanians	0.037716	German	0.129346
Andalusia	0.038453	Hungarian	0.131005
Armenians	0.038676	Turkish Cypriot	0.131133
UAE	0.039548	Darginian	0.141827
India-Hindu	0.047496	Bantu	0.142834
Catalans	0.048363	Nguni	0.166013
Filipino	0.056626	Java	0.174161
Kyrgyzes	0.058202	Ingushian	0.197343
French	0.064897	!Kung	0.20379

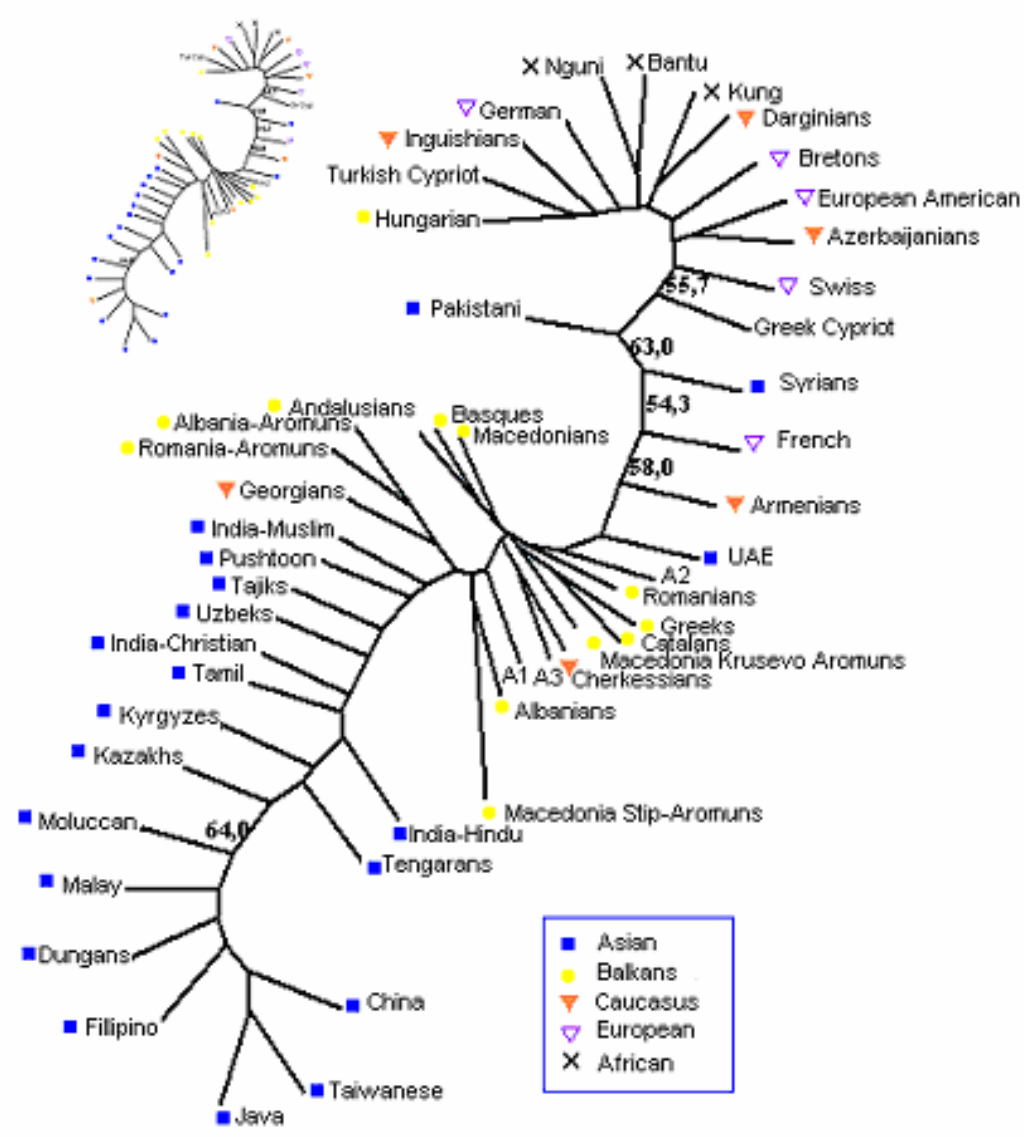

Figure 16. Neighbor joining tree constructed by using Anatolian population (A1). A1: Anatolian population studied in the present study; A2 and A3 are as described in Table 8.

3.3 Anatolian Data Together with the Previous Data from Anatolia

Alu insertion frequencies for that loci; ACE*, APO, PV92, TPA25, FXIIIB, B65, A25, D1, HS4.32* were already estimated by Dinç (2003) based on 100 individuals from all over Turkey. Comas et al. (2004) typed 43 individuals from the north of Aegean region in Turkey. Geographic locations of the individuals employed in those studies were shown in Figure 17 and 18. The insertion frequencies of nine *Alu* loci (ACE*, APO, A25, B65, D1, FXIIIB, HS4.32, PV92,* and *TPA25*) for three different samples of Anatolia can be seen in the Table 10.

Table 10. *Alu* insertion frequencies both in the present and previous Anatolian Data.

	N[a]	ACE	APO	PV92	TPA25	FXIIIB	B65	A25	D1	HS4.32
Present data	100*	0.410	0.963	0.220	0.390	0.579	0.667	0.067	0.427	0.637
Dinç (2003)	102*	0.333	0.951	0.152	0.480	0.461	0.485	0.069	0.371	0.668
Comas *et al.* (2004)	43	0.441	0.963	0.256	0.427	0.473	0.706	0.048	0.284	------

[a] (N) Number of individuals. * Average number of individuals typed for each locus.

When the significance of difference between two population proportions were examined (Daniel, 1999). It was observed that *PV92, TPA25, FXIIIB* and *B65* loci of the present data differed from those of Dinç (2003)'s with the p values of 0.04,

0.03, 0.009 and 0.0001 respectively. However, there were no significant differences between the data of present study and those of Comas *et al*., (2004). Therefore, from now on, the two data: the present one and Comas et al. (2004) are pooled and called A13, but not pooled with that of Dinç (2003)'s (A2).

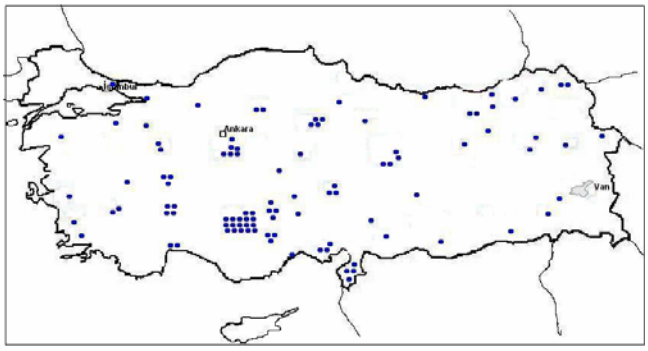

Figure 17. Geographic locations of the individuals employed by Dinç (2003). Dots indicate individuals analyzed by Dinç (2003) for 10 *Alu* loci.

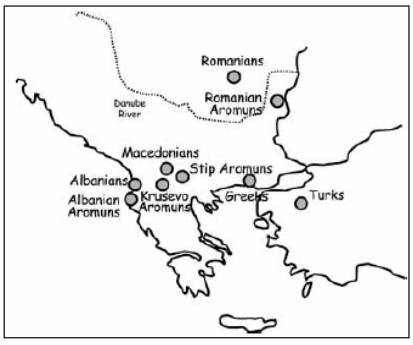

Figure 18. Geographic locations of the populations analyzed by Comas et al., (2004).

3.3.1 Analyses for the pooled Anatolian population

For the pooled data (Data from present study and data for Turks from Comas et al., (2004)'s study) (A13) locations of the samples are shown in Figure 18.

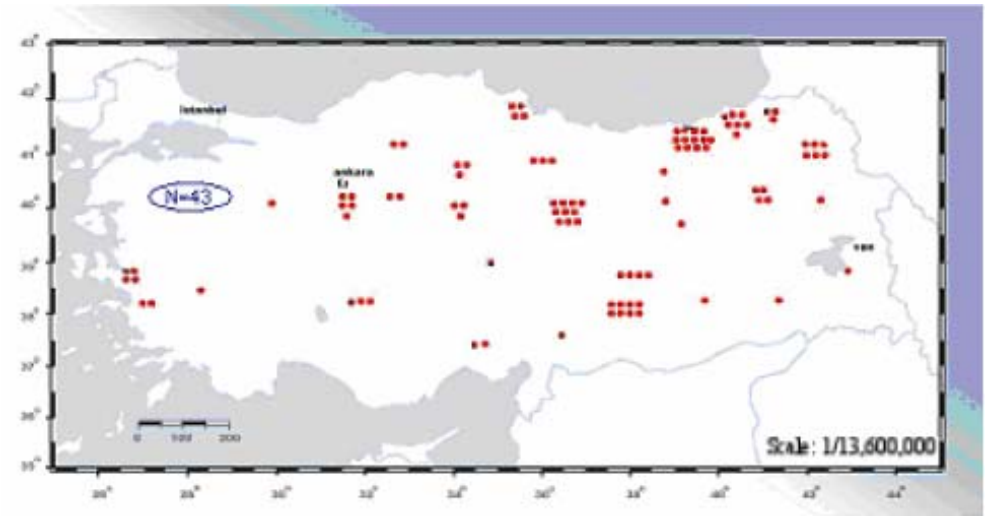

Figure 19. Geographic locations of the employed individuals (A13). N is number of individuals analyzed by Comas et al., (2004) and red dots denote the locations of the individuals typed in this study.

3.3.1.1 Distribution of nine *Alu* loci in Anatolia

Insertion frequencies for the pooled data (N=143) were calculated and shown in Table 11.

Table 11. *Alu* insertion frequencies in the A13 Anatolian population.

	ACE	APO	PV92	TPA25	FXIIIB	B65	A25	D1
Total data	0.419	0.963	0.232	0.401	0.544	0.681	0.061	0.378
N*	143	138	127	143	132	127	140	125

* N Number of individuals.

3.3.2 Present and Previous Anatolian (Comas et al., 2004) Data together with the Data from different human populations

Seven of the *Alu* loci (ACE, APO, A25, B65, FXIIIB, PV92, and TPA25) in A13 were compared with those of the Stoneking et al. (1997)'s, Comas et al. (2000)'s, Romualdi et al. (2002)'s, Nasidze et al. (2001)'s, Khitrinskaya et al. (2003)'s, Dinç (2003)'s and Comas et al. (2004)'s. These loci were the only ones commonly examined in all those studies.

3.3.2.1 The Fst Statistic

Fst related genetic distances based on seven *Alu* loci between Anatolian population (A13) and the populations from Eurasia were estimated. Pair-wise Fst values are shown in Table 12. They ranged between -0.00159 to 0.24337. The pair-wise Fst values between A13 and Cherkessians; Georgians; Macedonia; and Albanians are not significantly different from the each other. However, there were significant Fst values between A13 and all Central Asian populations; A13 was significantly different from Caucasian populations except Cherkessians and Georgians.

3.3.2.2 Reynold's Genetic Distances Between A13 and other Eurasian and African populations

Reynold's Genetic distances between A13 and other Eurasian populations were computed. Genetic distances between Anatolian population (A13) and others are shown in Table 13.

The results are the same as 3.5.1. Genetic distances between Anatolian population and Cherkess; Georgians; Macedonia-Krusevo Aromuns; Macedonians; Albania Aromuns; and Macedonians-Stip Aromuns were low. Again similar to the 3.5.1, genetic distance between the pairs of Anatolian population and the populations in Central Asia were high.

3.3.2.3 Neighbor Joining (NJ)Tree

Based on the recalculated genetic distances in accordance with the pooled Anatolian data (A13), a new neighbor joining tree was constructed (Figure 20). Similar topology to that of the previous tree (Figure 16) was obtained. On one end of the tree African populations are grouped. The opposite end of the tree is solely composed of the Asians. Balkan, some of the Caucasian, Middle Eastern, European populations are intermingled on the tree where Anatolian population (A13) was in this group. Central Asian populations were different than the populations of the first group and were together with other Asian populations

Thousand bootstrap replications were performed. The bootstrap values on the tree ranged from 3.5 % to 64.7 % and only the bootstrap values larger than 50 were depicted at the branching nodes on the tree.

Table 12. Fst values between Anatolian populations (A13) and some other populations from Eurasia and Africa. Anatolians: indicates A13; Anatolian 2: Anatolian population studied by Dinç (2004).

Populations- A13/ other populations	Fst-values	
Anatolians-Cherkessians	-0.00041	NS
Anatolians- Georgians	0.00235	NS
Anatolians- Albanians	0.00588	NS
Anatolians- Macedonia	0.00647	NS
Anatolians- Macedonia Stip Aromuns	0.00731	*
Anatolians- Albania Aromuns	0.00963	*
Anatolians- Romanian Aromuns	0.01255	*
Anatolians- Macedonia Krusevo Aromuns	0.01317	*
Anatolians- Greeks	0.0163	*
Anatolians- Basques	0.02072	**
Anatolians- Andalusians	0.02193	**
Anatolians-Anatolians 2	0.02228	***
Anatolians- Romanians	0.02295	**
Anatolians- Catalans	0.03146	***
Anatolians- Pustoon	0.03152	**
Anatolians- Turkish Cypriot	0.0346	***
Anatolians- Uzbeks	0.03508	***
Anatolians- India Muslim	0.03524	**
Anatolians- UAE	0.04169	***
Anatolians-Tajiks	0.05075	***
Anatolians- French	0.05155	***
Anatolians- Armenians	0.05383	***
Anatolians-Swiss	0.06639	***
Anatolians-Greek Cypriot	0.07282	***
Anatolians- Bretons	0.07825	***
Anatolians- Kyrgyzes	0.08112	***
Anatolians- Syria	0.08352	***
Anatolians- India Hindu	0.08468	***
Anatolians- Tamil	0.08617	***
Anatolians- Azerbaijanians	0.09235	***
Anatolians- Pakistan	0.09299	***
Anatolians- European American	0.10093	***
Anatolians- Kazakhs	0.11227	***

Table 12. (continued)

Anatolians- India-Christian	0.11619	***
Anatolians- Darginians	0.11846	***
Anatolians- German	0.14038	***
Anatolians- Dungas	0.14882	***
Anatolians- Ingushians	0.19191	***
Anatolians- China	0.19812	***
Anatolians- Taiwan	0.22833	***

NS : Nonsignificant
* $p<0.05$; ** $p<0.01$; *** $p<0.001$

Table 13. Genetic Distances between Anatolian Population (A13) and some other Eurasian and African populations. A2 are as described in Table 12.

Pairwise Genetic Distances		Population	A13
Population	A13	Kyrgyzes	0.061036
Cherkess	0.005057	Greek Cypriot	0.063968
Georgians	0.011059	Dungans	0.070692
Macedonians	0.014862	Syrians	0.07367
MacKrus-Aromuns	0.015824	Tamill	0.074108
Albania Aromuns	0.016336	Bretons	0.077134
MacStip- Aromuns	0.017021	Malay	0.078836
Albanians	0.019441	Tenggaras	0.079097
Romania Aromuns	0.019601	Pakistan	0.083246
Basque	0.021803	Kazakhs	0.087612
A2	0.024062	Chinese	0.089259
Uzbeks	0.032102	European American	0.095066
Romanians	0.032154	InChris	0.101245
Tajiks	0.033303	Azerbaijn	0.102905
Andalusia	0.033627	Taiwan	0.106342
Greeks	0.033666	Moluccan	0.109969
Armenians	0.034662	German	0.120091
Pushtoon	0.035284	Hungarian	0.120463
UAE	0.038336	Turkish Cypriot	0.122659
India- Muslum	0.039116	Bantu	0.135559
Catalans	0.046561	Darginian	0.139012
India-Hindu	0.051794	Nguni	0.162389
Filipino	0.056351	Java	0.176814
French	0.057939	Ingushian	0.192695
Swiss	0.059314	!Kung	0.20117

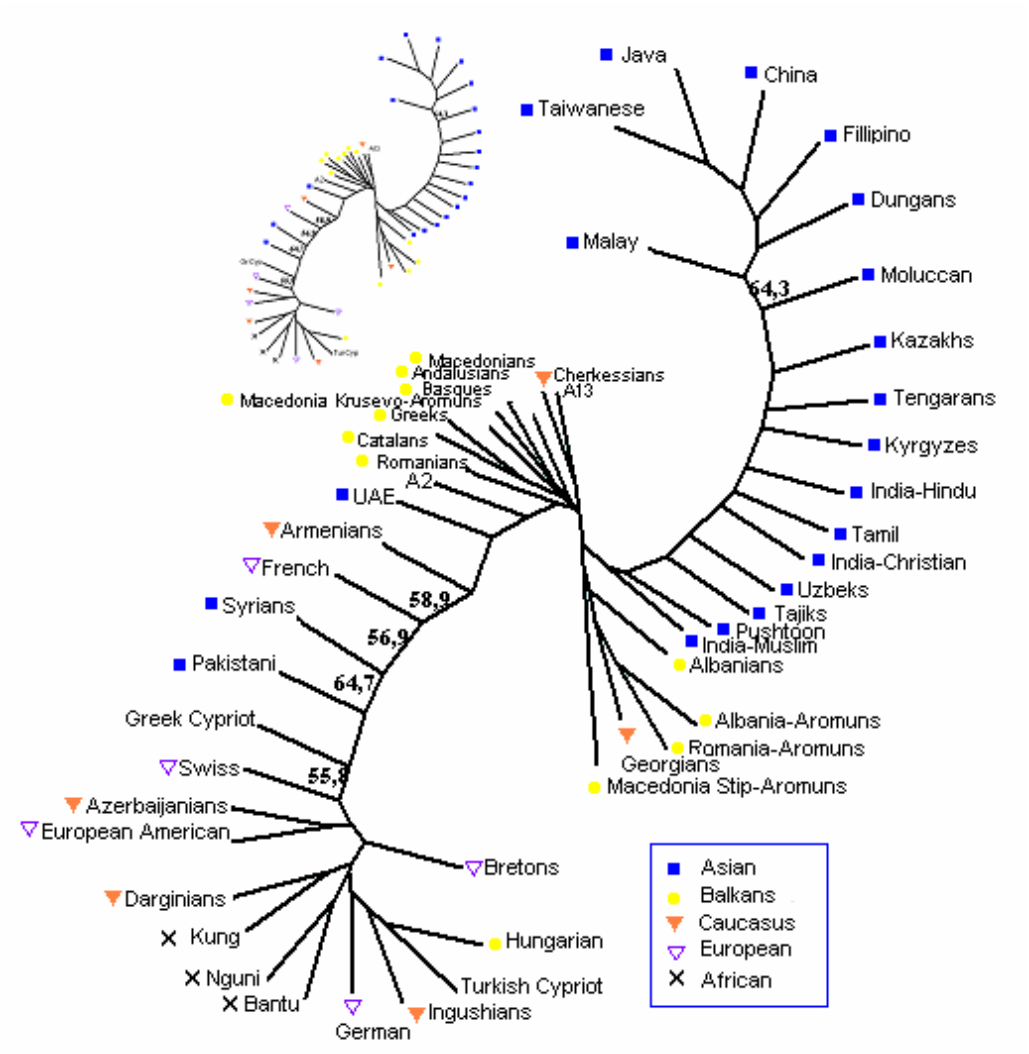

Figure 20. Neighbor joining tree constructed by using Anatolian population (A13). A2 are as described in Table 11.

3.3.2.4 Principal Component Analysis (PCA)

Based on the allele frequencies of seven polymorphic *Alu* loci, Principal Component Analysis (PCA) was performed to examine and to visualize relative relatedness of populations in 2 or 3 dimensional spaces. Three axes of the Principal Component analysis covers 77.7928 % of the total variation. The first Principal component (PC1) covers 37.1744 % of the overall variation while 26.2202 % and 14.3982 % of the total variation were explained by the second and third principal components. Populations on the first two axes covering 62.39 % of the total variation were shown in Figure 20. It was observed that Balkan, Basque and Anatolian populations are clumped together with some of the Caucasian populations. However, Central Asian populations had an intermingled distribution with those of other Asian populations on the two-dimensional depiction and Anatolia is not particularly close to the Asian populations.

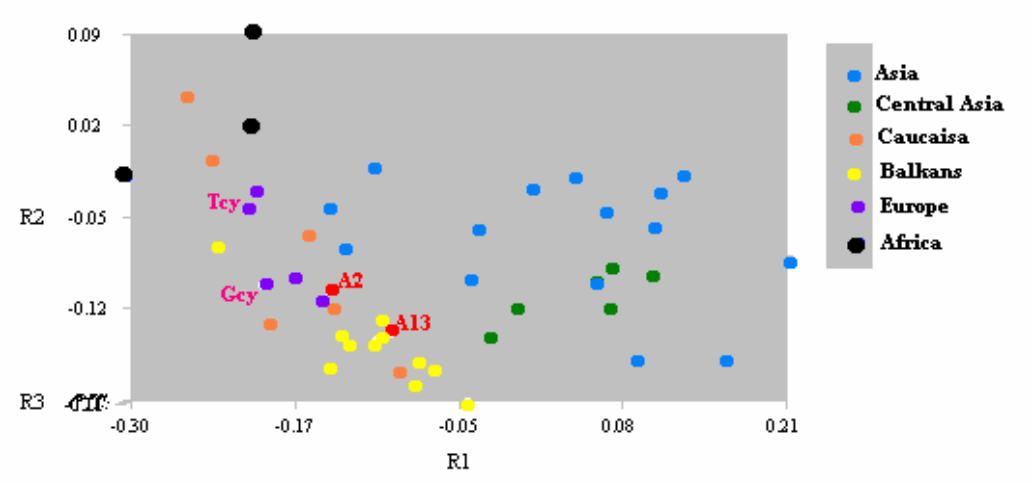

Figure 21. Populations on the first two principal components based on 7 *Alu* insertion loci. A13, A2 are as described on Table 12 and in the text on page 45.

3.3.2.5 Admixture Proportions

The relative genetic contribution of Central Asian genes to the current Anatolian gene pool was estimated on the basis of the *Alu* insertion polymorphism in eight loci (ACE, APO, A25, B65, FXIIIB, PV92, TPA25, and D1) and two parent populations Balkans (Greek, Romania, Albania, Hungarian) and Central Asia (Uighur, Uzbeks, Tajiks, Kyrgyzes, Kazakhs, Dungans). We decided on the two parent groups using the assumption that before the 11th century AD, the Anatolian human population was genetically similar to its neighbors in the Balkans.

Seven loci (ACE, APO, A25, B65, FXIIIB, PV92, TPA25) in approximately 297 samples from Balkans and 417 samples from Central Asia were collected by Comas et al. (2004), Mansoor et al. (2004), Khitrinskaya et al. (2003) and Xiao et al. (2002). D1 locus in 320 samples from Balkans and only 34 samples from Central Asia were collected by Comas et al. (2004), Mansoor et al. (2004), and Mayolo et al. (2002). And the frequency data was also updated via the database ALFRED.

Our estimates suggest roughly 28% Central Asian admixture for *Alu* insertion polymorphism. Table 14 summarizes the admixture estimates obtained by frequency approaches.

Table 14. Estimates of Central Asian admixture for *Alu* insertion polymorphism in Anatolian gene pool.

Parents	
Balkan	Average mR = 0.7158 ± 0.0585
Central Asia	(1 – mR) = 0.2842 ± 0.0585

CHAPTER 4

DISCUSSION

In the present study, we report on the nine polymorphic *Alu* insertion loci in a survey of 100 individuals from Anatolian human population. Insertion frequency of each locus was calculated and it was concluded that all loci were in Hardy-Weinberg equilibrium ($p<0.05$). Presence of Hardy-Weinberg equilibrium indicates that there is no evolutionary force (mutation, recent migration, random genetic drift or selection) operating on the gene pool. It also shows that there is no detectable mistyping in these loci. In one of the previous studies (Gerçeker, 1998), all of the loci were exhibiting heterozygote deficiency. Later, it was realized that it is due to the technique used in the study. The technique in its initial developmental stage and in its modified stages is presented below.

After comparing observed insertion frequencies of each loci with those of the previous observations (Dinç, 2003; Comas *et al*., 2004), it was found that frequencies of PV92, TPA25, FXIIIB and B65 loci were significantly ($p<0.05$-0.001) different from those of Dinç (2003)'s. May be one of the reasons for these discordance was typing failure. *Alu* elements are about 300 nucleotides in length and amplification of heterozygous samples is more difficult than both homozygous amplification because

of the competition between the smaller and larger bands. So it is very likely that after the amplification only one band that mimics the lack of insertion is seen. Because the smaller band is amplified more efficiently than the larger bands, visualization of heterozygous samples on agarose gels will show the smaller band being more intense than the larger band. Hence heterozygous samples can often be interpreted as homozygous (-/-) because of a faint upper band. Although "hot start" method (Ueda et al., 1996) which resolves mistyping problem was applied to the genomic DNA before the PCR amplification in both of the studies, increasing the annealing and extended time may also help to get intrinsic genotype from the gel. In the present study the mistyping problem was resolved not only by applying hot start method but also by increasing the annealing and extended time from 1 to 3 minutes (Chbel *et al*., 2003; Cotrim *et al*., 2004) for a more accurate genotyping. Hence, more accurate genotyping might have been achieved and the frequencies were different from these obtained before by Dinç (2003).

To examine genetic relationships between populations from Eurasia, Africa and pooled data (data from present study and data for Turks from Comas *et al*. (2004)) were used. The results of the phylogenetic analysis, either using the frequencies of the present study or pooled data were nearly the same. Since the pooled data covers higher sample size, the results of it will be discussed in detail in the following.

Fst statistic, applied to obtain significance of the Reynold's genetic distance values with the help of the permutations, shows that there is higher genetic similarity between Anatolia and all of the Balkans (Albania, Macedonia, Macedonia Stip-Aromuns, Albania Aromuns, Romania Aromuns, Macedonia Krusevo-Aromuns, Greeks) and some of the Caucasian populations (Georgians and Cherkessians). After the estimation of the pairwise Reynold's genetic distances the Neighbor Joining Tree (NJ) tree was constructed. NJ tree indicted a clear differentiation of Asian (also Central Asian) populations from those of Anatolians, Balkan and European populations based on the genetic distance values. Genetic similarity of the Anatolian populations with those of all the Balkans and some Caucasian populations was

visualized more clearly by means of the Principal Component Analysis (PCA) using the raw data of allele frequencies. Populations on the first two principal components based on 7 *Alu* insertion loci show that Balkan, Basques and Anatolian populations are clumped together with some of the Caucasian populations (Georgians and Cherkessians). However, Central Asian populations had a intermingled distribution with those of other Asian populations but Anatolia isn't there. Both the NJ tree using Reynold's genetic distance network and PCA showing genetic distance between populations as an Euclidean distance in two or three dimensions consistently show that *Alu* insertion polymorphisms present features are significantly distant from Asian or Central Asian populations.

When we focus on humans, a single species, the time scale is small and the evolutionary frame work becomes a historical framework (Comas *et al*., 1996). Genetic interpretation of the past can, thereafter, be compared to other disciplines traditionally devoted to the human past, like archaeological, historical or linguistic evidence (Renfrew 1992; Jobling *et al*., 2004). With the Oghuz Turk's invasion of Anatolia in 1071 their language was imposed upon resident populations. It is not clear to what extent the Anatolian gene pool was affected by the Oghuz invasion (Benedetto *et al*., 2001). The proportion of gene flow associated with the arrival of Central Asian Turkic speaking people to Anatolia were estimated. Benedetto *et al.* (2001) suggest 30 % contribution based upon mtDNA control region sequences and six microsatellite loci of the Y- chromosome and one *Alu* insertion (YAP) analyzed in 118 Turkish samples. Conversely, another study suggest only 3.4 % based on analysis of the 89 biallelic polymorphisms in 523 Turkish Y chromosomes (Cinnioğlu *et al*., 2004).

Since the current day Anatolian gene pool is molded by migration from east to west (such as Paleolithic and Neolithic from Middle East to Europe), west to east (such as Greeks to Persia) from north to south (such as Cimmers and Synthians (Ascherson, 1995; San, 2000; King, 2004) through the Caucasus. The mR estimation for Asia in the present study as well as the estimation of the Benedetto et al. (2001). The net

contribution of all these events definitely can not be attributed to the Oghuz's. It perhaps provides the upper bound for the Oghuz contribution. However, as one of the previous estimations (Benedetto *et al*., 2001) revealed 30 % Asian contribution in Anatolia compared to Balkans, in the present study based on a totally different marker (*Alu*) estimation of similar magnitude (28 %) of Asian contribution is noteworthy.

To contribute to the knowledge on history of Anatolian gene pool in addition many more loci in many more geographic locations (for instance from Middle East, Caucasian and Central Asian) must be examined. Many important issues in recent human evolution can be considered to be questions of admixture but it is important that any method of estimation of admixture proportions relies on the correct identification of the hybrid and parental populations. Because gene flow processes, including admixture, affect the entire genome, the greater the number of system considered, the more robust the inference about admixture (Bertorelle and Excoffier, 1998). Recently a model-based clustering method (STRUCTURE) has been deviced that determines the most likely number of clusters, the frequency of any given allele in each cluster and the proportion of each individual's genome that owes ancestry to each cluster. However, a study using either eight or 21 *Alu* insertions as genetic markers was unable to identify continent specific groups (Romualdi *et al*., 2002). Individuals are partitioned more clearly into the three classes as the number of *Alu* insertion polymorphisms increases from 20 to 100 (Bamshad *et al*. (2003).

Finally, considering greater number of loci will contribute to the knowledge on history of Anatolian gene pool.

CHAPTER

CONCLUSION

Human specific polymorphic *Alu* loci provide information about the genetic relatedness of different human populations. Although seven or eight loci do not provide a robust and accurate estimate for the relatedness, because for example bootstrap values are low in NJ tree or Basques are grouped together with Balkans they reveal the rough outline of the relation. Yet, since they are easy and cheap to study and since they provide an independent estimate of the evolutionary history compared to mtDNA and or Y chromosomes, they are worth of study.

In the present study nine *Alu* loci were examined. The reported frequencies will contribute to the data pool of these loci. Although, comparative studies indicated that, Asian contribution to Anatolia is 28% more compared to Balkans, Anatolian gene pool is closer to the populations of Balkans, Caucasians and Middle East, rather than the Central Asian populations in accordance with the mtDNA studies (Comas et al., 1996; Callafell et al., 1998).

Finally, it is known that 28 % Asian contribution can not be solely attributed to Oghuz invasion.

REFERENCES

Allele frequency database (ALFRED), http://alfred.med.yale.edu/alfred/index.asp, August 2005.

Ascherson, N., (2001). Black Sea. Hill and Wang; New York, USA.

Arcot, S.S., Shaikh, T.H., Kim, J., Benneth, L., Alegria-Hartman, M.,Nelson, D.O., Deininger P.L., and Batzer, M.A., (1995). Sequence diversity and chromosal distribution of "young" *Alu* repeats. *Genetics*, 163:273-278.

Bamshad, M.J., Wooding, S., Watkins W.S., Ostler C.T., Batzer M.A., and Jorde L.B., (2003). Human Population genetic structure and inference of group membership. *American Society of Human Genetics*, 72:578-589.

Barbujani G., (2001). Dna diversity and population admixture in Anatolia. *American Journal of Physical Antropolgy*, 115:144-156.

Batzer, M.A., Stoneking, M., Alegria-Hartman, M., Bazan, H., Kaas, D.H., Herrera, R.J., and Deininger P.L., (1994). African origin of human-specific polymorphic *Alu* insertions. *Evolution*, 91:12288-12292.

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., Boundreau D.A., Scheer W.D., Keats B.J.B., Deininger, P.L. and Stoneking M., (1996). Genetic variation of recent *Alu* insertions in human populations. *Journal of Molecular Evalution*, 42:22-29.

Belkhir, K., Borse, P., Chikki, 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.html, August 2005.

Benedetto, G.D., Ergüven A., Stenico M., Castri L., Bertorelle G., Togan İ., and Bertorelle, G., and Excoffier L., (1998). Inferring admixture proportions from molecular data. *Molecular Biology and Evolution*, 15(10):1298-1311.

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

Cann, R.L., Stoneking, M., and Wilson, A.C., (1987). Mitochondrial Dna and human evolution. *Nature*, 325:31-35.

Chbel, F.,Pancorbo M.M., Martinez –Bouzas, C., Azeddoug, H.,Alvarez-Alvarez, M., Rodriquez-Tojo and Nadifi. S., (2003). Polymorphism of six *Alu* insertions in Morocco: Comparative study between Arabs, Berbers and Casablanca residents. *Human Genetics*, 39(10):1398-1405.

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

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

Comas, D., Calafell, F., Benchemsi, N., Helal, A., Lefranc, 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 starits. *Annals of Human Genetics*, 107:312-319.

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

Cost, G,J., Feng, Q., Jacquier, A., Boeke J.D., (2002). Human L1 element target-primed reverse transcription in vitro. *The EMBO Journal*, 21(21):5899-5910.

Cotrim, N.H., Auricchio, M.T.B.M., Vicente, J.P., Otto, P.A., and Mingroni-Netto, R.C., (2004). Polymorphic *Alu* insertions in six Brazillian African-derived populations. *American Jounal of Human Genetics*, 16:264-227.

Daniel (1999). Biostatistics: A foundation for analysis in the health sciences. John Wiley &sons, Inc., New York.

Deininger, P.L., and Batzer, M.A., (2002). Mammalian retroelements. *Genome Research*, 12:1455-1465.

Dinç, H., (2003). *Alu* insertion polymorphisms in Anatolian Turks. Msc Thesis, METU.

El-sawy, M. and Deininger, P., (2005). Tandem insertions of *Alu* elements. *Cytogenetic and Genome Research*, 108:58-62.

Eswaran,V., Harpending H. and Rogers A.R., (2005). Genomic refutes an exclusively african origin of humans. *Journal of Human Evolution*.49:1-18.

Felsenstein, J., (1993). Phylogeny Inference Package (PHYLIP) version 3.6c. Department of Genetics, University of Washington, Seatle, http://evolution.genetics.washington.edu/phylip.html, August 2005.

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

Hendrick, P.W., (1985). Genetics of Populations. Jones and Bartlett Publishers, Inc., UK, (2nd Ed.).

Jobling et al., 2004. Human Evolutionary Genetics: Origins, Peoples and Diesase. Garland Science., New York.

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

Khitrinskaya, Y., Stepanov, V.A., Puzyrev, V.P., Spiridonova, M.G. and Voevoda, M.I., (2003). Genetic differentiation of the population of central Asia inferred from autosomal markers. *Human Genetics.* 39:1389-1397.

King, C., (2004). The Black Sea: A History. Oxford University Press, Oxford.

Mansoor, A., Mazhar, K., Khaliq.S., Hameed.A., Rehman, S., Siddiqi, S., Papaioannou. M., Cavalli-Sforza, L.L., Mehdi, S.Q., Ayub, Q., (2004). Investigation of the Greek ancestry of populations from nothern Pakistan. *Human Genetics.* 114:484-490.

Mayola, G., Antunez-de-Mayola, A., Antunez-de-Mayola, P., Papiha, S.S.,Hammer. M.,Yunis J.J.,Yunis E.J., Damodaran, C., Pancorbo, M.M., Caeiro J.L.,Puzyrev V.P. and Herrera, R.J., (2002). Phylogenetics of wordwide human populations as determined by polymorphic *Alu* insertions. *Electrophoresis*, 23: 3346-3356.

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

Perna, N.T., Batzer, M.A., Deininger P.L., And Stoneking, M., (1992).”*Alu*” insertion polymorphism:a new type of marker for human population studies. *Human Biology*, 64:5. 641-648.

Renfrew, C.(1992). Archaelogy, genetics and linguistic diversity. Oxford Press. 27:445-478.

Reynolds, J., Weir, B. S., & Cockerham, C. C., (1983). Estimation of co-ancestrycoefficient: basis for a short-term genetic distance. *Genetics* 105**,** 767–779.

Rolf, F.Y., (1993). NTSYS-pc: Numerical Taxonomy and Multivariate Analysis System, version 2.10 g., App. Biostatistcs ,Inc.

Romualdi, C., Balding, D., Nasidze I.S., Risch, G.,Robicaux 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.* 12:602-612.

Saitou, N. and Nei, M., (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mollecular Biology and Evolution*, 4(4):406-425.

San, O., (2000). Bazı bulgular ışığında Anadolu'da Kimmer ve İskit varlığı üzerine gözlemler. *Belleten*, 239:1-23.

Sherry, S.T. and Batzer, M.A., (1997). Modelling human evolution-to tree or not to tree? *Genome Research*, 7:947-949.

Simoni, L.,Calafell, F., Pettener, D., Bertranpetit, J. And Barbujani, G.(1999). Geographical patterns of mtdna diversity in Europe. *American Society of Human Genetics,* 66:262-278.

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

Strachan, T., Read, A.P., (2004). Human Molecular Genetics 3. Garland Science, London and New York.

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

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

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

Watkins, W.S., Rogers A.R.,Ostler C.T., Wooding S., Bamshad M.J., Brassington A.E., Caroll M.L., Nguyen S.V., Walker J.A., Prasad, B.V.R., Reddy P.G., Das P.K., Batzer M.A. and Jorde L.B.(2003). Genetic variation among world populations:inferences from 100 *Alu* insertion polymorphisms. *Genome Research*,13:1607-1618.

Watkins, W.S., Ricker, C.E., Bamshad M.J., Caroll, M.L., Nguyen, S.V., Barzer 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. *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(6):1358-1370.

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

APPENDIX A

QUESTIONNAIRE AND CONSENT

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 (father's) birth place:

Address and telephone number:

CONSENT :

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

Signature:

APPENDIX B

ALU INSERTION FREQUENCIES

Table A.1 *Alu* insertion frequencies at seven *Alu* loci and their sample sizes in 49 populations.

Populations	Insertion frequencies							Sample sizes						
	ACE	APO	PV92	TPA25	FXIIIB	B65	A25	ACE	APO	PV92	TPA25	FXIIIB	B65	A25
[a]Albanians	0.467	1.000	0.203	0.557	0.600	0.670	0.075	60	60	60	60	60	60	60
[a]Romanians	0.469	0.915	0.200	0.577	0.408	0.569	0.031	65	65	65	65	65	65	65
[a]Maceds	0.482	0.962	0.200	0.519	0.491	0.600	0.127	55	55	55	55	55	55	55
[a]Greeks	0.260	0.977	0.190	0.552	0.500	0.651	0.098	51	51	51	51	51	51	51
[a]AlbaniaAr	0.367	1.000	0.306	0.500	0.684	0.638	0.051	49	49	49	49	49	49	49
[a]RomaniaAr	0.483	1.000	0.322	0.517	0.678	0.737	0.051	59	59	59	59	59	59	59
[a]MacStipAr	0.492	0.977	0.192	0.531	0.608	0.631	0.115	65	65	65	65	65	65	65
[a]MacKrusAr	0.361	0.986	0.118	0.514	0.521	0.563	0.090	72	72	72	72	72	72	72
[b]Kyrgyzes	0.630	0.904	0.539	0.470	0.808	0.546	0.087	104	104	102	101	99	97	103
[b]Tajiks	0.329	0.902	0.537	0.585	0.622	0.590	0.128	41	41	40	41	41	39	39
[b]Uzbeks	0.598	0.913	0.467	0.511	0.554	0.585	0.085	46	46	45	46	46	47	45
[b]Dungans	0.602	0.886	0.773	0.585	0.682	0.420	0.045	44	44	44	41	44	44	44
[b]Kazakhs	0.601	0.924	0.550	0.533	0.797	0.346	0.072	79	79	80	75	69	78	76
*Anatolia (A13)	0.419	0.963	0.232	0.401	0.544	0.681	0.061	143	138	127	143	132	127	140
[c]Bretons	0.280	0.900	0.270	0.320	0.160	0.510	0.170	66	68	71	67	62	67	72
[d]China	0.670	0.820	0.860	0.350	0.710	0.350	0.100	49	49	49	49	49	49	49
[c]EurAmeri	0.500	0.930	0.230	0.540	0.060	0.580	0.140	71	63	71	70	62	71	71
[c]French	0.410	0.950	0.160	0.600	0.290	0.490	0.090	71	68	69	70	68	68	70
[c]German	0.460	0.870	0.100	0.510	0.050	0.350	0.140	69	64	65	70	61	69	70
[c]Gcypriot	0.330	0.950	0.150	0.510	0.180	0.530	0.000	49	47	47	50	44	48	47
[d]InChris	0.600	0.670	0.480	0.570	0.610	0.310	0.140	27	27	27	27	27	27	27
[d]InHindu	0.520	0.850	0.520	0.340	0.660	0.350	0.050	28	28	28	28	28	28	28
[d]InMuslum	0.520	0.860	0.300	0.410	0.660	0.400	0.120	26	26	26	26	26	26	26
[d]Pakistan	0.440	0.720	0.300	0.510	0.230	0.370	0.070	42	42	42	42	42	42	42
[d]Pushtoon	0.520	0.860	0.330	0.550	0.570	0.490	0.180	50	50	50	50	50	50	50
[c]Swiss	0.390	0.940	0.160	0.500	0.160	0.570	0.150	70	51	66	69	65	71	71
[d]Taiwan	0.500	0.930	0.900	0.640	0.970	0.540	0.220	46	46	46	46	46	46	46
[d]Tamill	0.690	0.810	0.560	0.560	0.610	0.550	0.170	47	47	47	47	47	47	47
[c]TCypriot	0.360	0.960	0.150	0.400	0.010	0.390	0.080	59	56	59	59	69	71	71
[d]UAE	0.330	0.970	0.300	0.440	0.390	0.410	0.120	42	42	42	42	42	42	42
[d]Filipino	0.530	0.980	0.800	0.630	0.720	0.570	0.140	47	47	47	47	47	47	47
[d]Malaysian	0.640	0.760	0.720	0.500	0.730	0.420	0.020	47	47	47	47	47	47	47
[c]Moluccas	0.560	0.750	0.690	0.550	0.780	0.250	0.000	51	50	50	50	38	47	46
[d]Java	0.860	0.780	0.840	0.390	0.920	0.580	0.060	97	97	96	92	90	86	92

Table A.1. (continued)

[c]Tenggarans	0.610	0.780	0.510	0.390	0.810	0.400	0.050	97	97	96	92	90	86	92				
[e]Armenians	0.477	0.871	0.013	0.430	0.343	0.453	0.058	40	40	40	40	40	40	40				
[e]Azerbaijn	0.216	0.943	0.382	0.513	0.100	0.697	0.000	34	34	34	34	34	34	34				
[c]Cherkess	0.390	0.932	0.167	0.386	0.439	0.651	0.045	40	40	40	40	40	40	40				
[e]Darginian	0.167	0.864	0.167	0.361	0.143	0.321	0.028	16	16	16	16	16	16	16				
[e]Georgians	0.354	0.934	0.250	0.493	0.610	0.727	0.088	67	67	67	67	67	67	67				
[e]Ingushian	0.340	0.941	0.129	0.224	0.000	0.210	0.067	24	24	24	24	24	24	24				
[c]Syrians	0.400	0.930	0.180	0.510	0.280	0.310	0.000	70	68	68	70	69	70	66				
[c]Bantu	0.330	0.730	0.280	0.300	0.000	0.460	0.080	48	44	48	47	47	48	48				
[c]IKung	0.130	0.990	0.170	0.140	0.070	0.480	0.230	39	37	38	42	35	40	33				
[c]Nguni	0.360	0.600	0.180	0.200	0.090	0.340	0.270	43	39	42	44	43	41	42				
[c]Hungarian	0.380	0.970	0.120	0.500	0.000	0.450	0.040	69	65	65	66	59	67	70				
[f]Basques	0.443	0.953	0.188	0.568	0.484	0.604	0.219	96	96	96	96	96	96	96				
[f]Catalans	0.300	0.983	0.175	0.608	0.500	0.525	0.125	60	60	60	60	60	60	60				
[f]Andalusia	0.470	0.985	0.194	0.590	0.448	0.552	0.142	67	67	67	67	67	67	67				

[a] Comas et al., 2004; [b] Khitrinskaya et al., 2003; [c] Romualdi et al., 2002; [d] Stoneking et al., 1997; [e] Nasidze et al., 2001; [f] Comas et al., 2000; * Data from present study and data for Turks from Comas et al., 2004.

APPENDIX C

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)

1 M Tris-HCl

9) 5X Tris Borate Edta (TBE) Buffer (PH: 8)

0.45 M Tris (Base)

0.45 M Boric Acid