GENETIC STRUCTURE ANALYSIS OF HONEYBEE POPULATIONS BASED ON MICROSATELLITES

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

BY

ÇAĞRI BODUR

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
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 Doctor of Philosophy.

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 Doctor of Philosophy.

Prof. Dr. Aykut Kence
Supervisor

Examining Committee Members

Prof. Dr. Işık Bökesoy (Ankara Unv., MED) _____

Prof. Dr. Aykut Kence (METU, BIOL) _____

Assoc. Prof. Dr. Meral Kence (METU, BIOL) _____

Prof. Dr. Semra Kocabıyık (METU, BIOL) _____

Prof. Dr. A. Nihat Bozcuk (Hacettepe Unv., BIOL) _____

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 : Çağrı BODUR

Signature :

ABSTRACT

GENETIC STRUCTURE ANALYSIS OF HONEYBEE POPULATIONS BASED ON MICROSATELLITES

Bodur, Çağrı
Ph. D., Department of Biological Sciences
Supervisor: Prof. Dr. Aykut Kence

September 2005, 116 pages

We analyzed the genetic structures of 11 honeybee (*Apis mellifera*) populations from Türkiye and one population from Cyprus using 9 microsatellite loci. Average gene diversity levels were found to change between 0,542 and 0,681. Heterozygosity levels, mean number of alleles per population, presence of diagnostic alleles and pairwise F_{ST} values confirmed the mitochondrial DNA finding that Anatolian honeybees belong to north Mediterranean (C) lineage. We detected a very high level of genetic divergence among populations of Türkiye and Cyprus based on pairwise F_{ST} levels (between 0,0 and 0,2). Out of 66 population pairs 52 were found to be genetically different significantly. This level of significant differentiation has not been reported yet in any other study conducted on European and African honeybee populations. High allelic ranges, and high divergence indicate that Anatolia is a genetic centre for C lineage honeybees.

We suggest that certain precautions should be taken to limit or forbid introduction and trade of Italian and Carniolan honeybees to Türkiye and Cyprus in order to preserve genetic resources formed in these territories in thousands of years. Effectivity at previously isolated regions in Artvin, Ardahan and Kırklareli was confirmed by the high genetic differentiation in honeybees of these regions. Genetically differentiated Karaburun and Cyprus honeybees

and geographical positions of the regions make these zones first candidates as new isolation areas.

Keywords: Honeybee, *Apis mellifera*, Türkiye, Cyprus, C lineage, population, microsatellite

ÖZ

BALARISI TOPLUMLARININ GENETİK YAPILARININ MİKROSATELİTLER KULLANILARAK İNCELENMESİ

Bodur, Çağrı
Doktora, Biyoloji Bölümü
Tez Yöneticisi: Prof. Dr. Aykut Kence

Eylül 2005, 116 sayfa

Türkiye'den 11 ve Kıbrıs'tan bir balarısı (*Apis mellifera*) toplumunun genetik yapılarını 9 mikrosatelit lokusu kullanarak inceledik. Ortalama gen farklılaşması düzeylerinin 0,542 ile 0,681 arasında değiştiği belirlendi. Heterozigotluk düzeyleri, toplum başına düşen ortalama alel sayıları, tanımlayıcı alellerin varlığı ve ikili F_{ST} değerleri mitokondriyel DNA bulgusu olan Anadolu balarılarının kuzey Akdeniz (C) soyhattına ait oldukları savını doğrulamıştır. İkili F_{ST} değerlerine (0,0 ile 0,2 arasında) göre Türkiye ve Kıbrıs toplumları arasında çok yüksek bir genetik farklılaşma bulduk. Altmış altı toplum çiftinden 52'si genetik olarak anlamlı düzeyde farklı bulundu. Bu düzeyde bir anlamlı farklılaşma henüz Avrupa ve Afrika balarısı toplumlarında yapılan hiçbir çalışmada belirtilmemiştir. Yüksek alel kapsamları ve yüksek oranda farklılaşma Anadolu'nun C soyhattı balarıları için bir genetik merkez olduğunu göstermektedir.

Biz bu yaşam alanlarında binlere yılda oluşmuş gen kaynaklarının korunması için Türkiye ve Kıbrıs'ta İtalyan ve Karniyol arılarının girişi ve ticaretinin sınırlandırılması ya da yasaklanması için gerekli önlemlerin alınmasını önermekteyiz. Artvin, Ardahan ve Kırklareli'nde daha önce yalıtılan bölgelerdeki başarı bu bölgenin arılarında bulunan güçlü genetik yapı ile doğrulanmıştır. Karaburun ve Kıbrıs'taki balarılarının güçlü genetik

yapıları ve coğrafi konumları bu bölgeleri yeni yalıtım alanları olarak ilk adaylar yapmaktadır.

Anahtar sözcükler: Balarısı, *Apis mellifera*, Türkiye, Kıbrıs, C soyhattı, toplum, mikrosatelit

ACKNOWLEDGMENTS

I wish to express his deepest gratitude to his supervisor Prof. Dr. Aykut Kence for his guidance, advice, criticism, encouragements and insight throughout the research.

I would also like to thank Assoc. Prof. Dr. Meral Kence for her suggestions and comments.

I am grateful to Evren Koban for her equipment support.

I would also like to thank to Damla Beton, Havva Dinç, Sara Banu Akkaş, Rahşan Tunca İvgin and Mehmet Değirmenci for their help.

TABLE OF CONTENTS

LIST OF SYMBOLS AND ABBREVIATIONS

ARD: Ardahan

ART: Artvin

ANA: Anatolia

ANK: Ankara

CYP: Cyprus

D_{LR}: Likelihood ratio distance

D_S: Standard genetic distance

ESK: Eskişehir

HAK: Hakkari

HAT: Hatay

H_E: Expected heterozygosity

H_O: Observed heterozygosity

HWE: Hardy-Weinberg equilibrium

İZM: İzmir

KAS: Kastamonu

KIR: Kırklareli

MASH: Molecularly accessible size homoplasy

MUĞ: Muğla

Nm: Number of migrants

St. Dev.: Standard deviation

URF: Urfa

CHAPTER 1

INTRODUCTION

1.1. Geographical Distribution and Evolution of Honeybees

Bees living in social communities are classified within Apidae family. Honeybees are named under the Apinae subfamily of Apidae family. This subfamily is characterized with special pollen collecting organs. Apinae subfamily includes four species of honeybees namely: *Apis dorsata*, *A. florea*, *A. cerana* and *A. mellifera* (Ruttner 1988).

A. mellifera shows a wide geographical distribution throughout the world which caused evolution of highly divergent subspecies. Several hybridization experiments showed that even most distant subspecies are of the same species, *A. mellifera*. *Apis* specific characters first emerged in early Tertiary period. This original *Apis* type is thought to be retained up to now without important morphological and ecological diversification. This "conservative" *Apis* type became extinct in Europe where climatic conditions deteriorated at the end of the Tertiary since then it was confined to the tropical conditions (Ruttner 1988). In the early Pleistocene (1-2 million years ago) a temperate climate *Apis* type evolved having new behavioral characteristics such as cavity nesting, temperature homeostasis, and elaborated dance communication. This Apinae subfamily succeeded in gaining independence from environmental effects by these features and a great radiation started to recolonize Europe and colonize Africa and great morphological and ecological diversification at subspecies level Thank to its high fitness and plasticity the new type also rapidly spreaded through several climatic zones of the New World recently (Ruttner 1988).

Among *Apis* species *A. mellifera* and *A. cerana* have very similar characteristics and it is evidenced that it had evolved more recently than two other *Apis* species since they do not have a pre-mating barrier. These two *Apis* species are believed to be at an immature stage of speciation which started by sexual isolation at last glaciation period. Therefore they should heve been existed for at most 50.000 years (Ruttner 1988). According to an hypothesis, two very similar *Apis* species, *A. mellifera* and *A. cerana* separated from each

other at south coast of the Caspian Sea not earlier than during the Pleistocene and the two spreaded towards opposite directions: *A. mellifera* to the west and *A. cerana* to the east. *A. cerana* shows a sympatric distribution with *A. m. dorsata* and *A. m. florea* in the southeast Asia whereas *A. m. mellifera* follows a distribution ranging from south of Caspian Sea to western Europe through Anatolia and it is also distributed in Africa without any other Apis species. Thus the Mediterranean is thought to be a gene center for all *A. mellifera* because it was firmly connected Africa at those times.

Apis florea, also called as "Dwarf Honeybee" and *Apis dorsata*, "Giant Honeybee" species are distributed throughout South Asia. And the "Eastern Honeybee" *A. cerana* is occupying almost all of Asia. The western honeybee, *Apis mellifera* L. has been adapted to many kinds of climates, cold, temperate, tropical, humid, and semi-deserts. Some subspecies of western honeybees including *anatoliaca* are known to have evolved to survive during long, hard winter conditions (Adam 1983).

1.1.1. Honeybees of Middle East

Middle East honeybee races comprise *Apis mellifera syriaca, adami, anatoliaca, meda, cypria, caucasica* and *armeniaca* (Ruttner 1988) (Figure 1). Among this group of subspecies *syriaca* and *cypria* are substantially smaller and very yellow compared to species at the north. Middle East is a zone of huge diversification and evolution for *Apis mellifera* species. It is thought to be an isolated part containing distinct subspecies adapted to diverse climate and habitat conditions. Anatolia is the genetic center of this group (Ruttner 1988). Before human interference honeybees of this region were isolated from other western honeybee subspecies. At the north there are dry steppes of Russia, at the west Ukraine did not have honeybee colonies 500 years ago, at the east border of Iran no honeybee existed and the remaining borders of the region are all sea coasts except a contact zone in Thrace, Türkiye.

The distribution of 5 subspecies out of 26 recorded so far seems to overlap within borders of Türkiye; These are *Apis m. carnica* in Thrace, *A. m. anatoliaca* in central Anatolia, *A.m. caucasica* in northeastern Anatolia, *A.m. meda* in eastern Anatolia, and *A. m.syriaca in* southeastern Anatolia (Kandemir *et al*. 2000).

Beekeeping tradition in Anatolia has origins long before 1300 B.C. as understood from an old Hittite code found in Boğazköy (Ruttner 1998). Ruttner (1988) argued that honeybees of Western Anatolia (*anatoliaca* subspecies) seems to be eastern genetic center of *Apis mellifera* based on phenetic similarities of these populations with southeast Europe, central Mediterranean and north African populations. Among excellent performances of Anatolian honeybees in extreme climatic conditions of Central Anatolia are wintering ability in harsh weather, energetic food collecting activity and adjustments to save energy and reserves at dearth times.

A.m. cypria is an island (Cyprus) subspecies well known for its aesthetic appearence especially because of its bright orange color. According to morphometric analyses these honeybees are almost equally distant from *anatoliaca, syriaca*, and *meda* (Ruttner 1998). Honeybees belonging to *A. m. syriaca* subspecies is the smallest of all Middle East subspecies and distributed around Israel, Jordan, Lebanon, Syria, and Hatay region of Türkiye. Morphometric analyses showed that this subspecies is the closest subspecies of the Middle East region to African honeybees (Ruttner 1998). They are known to be excellently adapted to the ecological conditions of their region. They produce more honey than well known Italian bees in their habitat and have more powerful defensive tactics against their predators (Ruttner 1988). But because of these defensive aggressivity colony management of *syriaca* may sometimes become problematic. *A. m. meda* subspecies is distributed within Iran, Iraq, and southeast Türkiye. This subspecies is occupying one of the largest territory among *Apis mellifera* subspecies.

A worldly renown subspecies *A. m. caucasica* is another honeybee subspecies that has a distribution in Türkiye. Northeast region of Türkiye is occupied by this race which is famous for its long probiscus. These bees have the longest tongues among all *mellifera* subspecies of the world. Other Middle East distribution areas of this so called "Grey Caucasion Mountain Bee" include east coast of the Black Sea, Georgia, and parts of Azerbaijan. When distribution areas are examined, this subspecies seems to be limited by climate. A subtropical humid climate at the sea level and cool temperate climate at mountains determine their living areas (Ruttner 1998).

Figure 1. Honeybee subspecies of Middle East

1.2. Genetic studies on honeybee distribution

According to Ruttner (1998) the western honeybee *Apis mellifera* is originated in Asia and invaded Africa and Europe in four evolutionary distinct branches. These branches are Near East (O), Tropical Africa (A), Western Mediterranean (M), and Central mediterranean and Southeastern European (C) branches. The original distribution areas of *A. mellifera* includes south and west of Asia, Europe and Africa. Currently 26 subspecies of *A.mellifera*

are formally recognized, based primarily on morphometric characters (Sheppard and Smith 2000). Although basic honeybee studies were almost exclusively based on morphometry, use of morphological characters has the disadvantage of polygenic determinism and these characters are not very suitable since they are sensitive to environmental selection pressures.

Allozyme analyses have brought very little information about honeybee evolution and population structure because of their low variability within this species (Pamilo *et al*. 1978; Sheppard 1986; Packer and Owen 1992) which should be a result of haplodiploidy (Pamilo *et al*. 1978).

Among DNA markers mitochondrial DNA (mtDNA) and microsatellite marker analyses have proved to be very useful in studying honeybee evolution and resolving the relationships between honeybeee populations, among and between lineages. The preliminary studies on mtDNA, a powerful discriminator at subspecies level, confirmed the existence of three evolutionary branches A, C and M (Smith 1991, Garnery *et al*. 1992, Arias and Sheppard 1996). In addition to 3 lineages the presence of the fourth lineage, O, was confirmed by a mitochondrial DNA study later (Franck *et al*. 2000a). Within lineage level mitochondrial DNA polymorphism among *A.m.mellifera* subspecies have been studied by researchers (Smith *et al*. 1989, 1991; Garnery *et al*. 1993, 1995; Franck *et al*. 1998). However one drawback of mtDNA is its uniparental inheritance. When formerly isolated populations come into contact via range expansion or human interference mtDNA introgression to new populations occur which may cause discordance between morphometric and mtDNA analyses. Moreover maternally inherited mitochondrial DNA, although being useful in population genetics, has been reported to have little genetic differences between honeybee subspecies (Arias and Sheppard 1996). Since in mtDNA analysis one bee represents the entire colony it is most powerful when used in conjunction with biparentally inherited nuclear markers (Sheppard and Smith 2000).

Microsatellites (look at page 10) which were reported to be abundant and highly variable in *A. mellifera* (Estoup *et al*. 1994) proved to be appropriate to discriminate subspecies and populations within these subspecies (Estoup *et al*. 1995a, Franck *et al*. 1998). Much larger samples (200-750 workers) were reported to be needed in order to determine genetic structure within a lineage if morphometry is used instead of microsatellites to reach the

same level of resolution. Twenty or 30 unrelated honeybee workers were shown to be sufficient for determining genetic differentiation among honeybee populations for even 7 microsatellite loci.

Existence and composition of three evolutionary honeybee lineages, A, M, and C, each represented with three different subspecies, was confirmed by seven microsatellites (Estoup *et al*. 1995a). Number of alleles for each locus was found to be between 7 and 30 in this study among 9 European and African honeybees. Average heterozygosities for populations were reported to be in the range of 0,291 and 0,872 in this study.

Microsatellite studies on genetic structures of honeybee populations from three evolutionary lineages A, M and C revealed that genetic variation is far higher in A and C lineages than M subspecies in terms of heterozygosity and allelic number (Garnery *et al.* 1998, Estoup *et al.* 1995a, Franck *et al.* 2001). In several studies genetic structures of honeybee populations in Slovenia, Spain, Canary Islands, Balearic Islands, continental Italy and Sicily Island and Africa continent have been analysed using microsatellite markers (Susnik *et al.* 2004, De La Rua *et al.* 2002,2001,2003,Franck *et al.* 2000b,2001). *A. m. carnica* honeybees of Slovenia and Croatia were found to have a uniform genetic structure without much differentiation (Susnik *et al*. 2004)

Microsatellites were shown to be able to assign a given honeybee colony to its original population even by using four microsatellite loci (Estoup *et al*. 1994). In this test, parental structure of the colony was found not to be significantly different than the original population, but colony structure was found to be significantly different than other populations in comparison. Increasing the microsatellite loci number to 12 did not change the situation. A single colony can give a very approximate estimation of average heterozygosity within the population. Microsatellites were also used in understanding the amount of gene flow. Introgression of commercial *A. m. ligustica* honeybees in northwest Europe were reported to represent a gene flow threat in a microsatellite analysis although native *A. m. mellifera* honeybees still exist (Jensen *et al.* 2005).

Number of the microsatellite loci isolated in most of the species is not high since these molecular markers are used mainly for population genetic studies. A few vertebrate species and cultivated plants are among species for which a large number of microsatellite loci were identified. Because of their economic and academic importance honeybees are among

very few invertebrate species that have a large number of isolated microsatellite loci. Flanking regions are well conserved among different honeybee subspecies and lineages as revealed by rarity of null alleles detected (Solignac *et al.* 2003). A total of 552 microsatellite loci containing mono, di, tri or tetra nucleotide repeat motifs were isolated and sequenced for *A. mellifera* (Solignac *et al*. 2003). Variability at 36 loci analysed for populations representing three mitochondrial lineages A, M and C showed that African lineage has a much higher variation compared to populations belonging to M and C lineage. A cross-species priming test showed that about 30% of 552 isolated *A. mellifera* microsatellites were also amplified in all four *Apis* species including *A. m. cerana, florea* and *dorsata* other than *mellifera*. This proportion of cross-priming should be even higher since only standart polymerase chain reaction (PCR) conditions were applied to all loci. Cross-priming efficiency shows that these loci could be exploited in comparative genome analyses among four different honeybee species (Solignac *et al*. 2003).

Nuclear restriction fragment length polymorhism (RFLP),random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP) are among other DNA markers used in honeybee population genetic studies (e.g., Hall 1990; Suazo *et al*. 1998; Suazo and Hall 1999). Although they have many polymorphic loci, nuclear RFLPs are not very suitable for large scale population studies because of impractical transfer hybridization detection and probes not widely available (Sheppard and Smith 2000). RAPD markers are dominantly inherited and they are difficult to replicate in different laboratories. AFLP method is very useful at intraspecifc level since it reveals high polymorphism and it is repeatable among laboratories,economical and fast (Vos *et al*. 1995, Sheppard and Smith 2000).

1.2.1. Genetic studies on honeybees of Middle East

Honeybee samples from Lebanon was analysed by mtDNA and microsatellites (Franck *et al*. 2000a)**.** High genetic divergence found between Lebanon honeybee samples and other samples representing A, M and C lineages supports the existence of a fourth evolutionary lineage (O) in the Middle East. However Lebanese population showed a little differentiation with Greek population from Chalkidiki based on microsatellites. Furthemore

mtDNA data for honeybees of Egypt shows that lineage O may extend to Northeast Africa outside the Middle East.

Studies based on morphometric characters and allozymes have shown great amount of variability in honeybees of Türkiye and thus supported the idea that Anatolia has been a genetic center for Middle East populations (Darendelioğlu and Kence 1992, Kandemir and Kence 1995, Asal *et al*. 1995, Kandemir *et al*. 2000). Thrace and southeastern Anatolia samples were found to be separate units, Black Sea and eastern Anatolia samples clustered closely such as central Anatolian ,Aegean and Mediterranean samples did (Kandemir *et al*. 2000). A mean heterozygosity of 0.072±0.007 among *A.mellifera* populations of Türkiye was obtained which was higher than the value, 0,038, which was obtained from 23 colonies of European honeybees (*A.mellifera*) by Sheppard (1986).

So far three evolutionary mtDNA lineages were identified based on restriction site and sequence polymorhism studies. Türkiye is at the crossroads of Europe, Asia and Middle East and therefore comprises diverse ecological conditions among which five honeybee subspecies exist (Kandemir *et al*. 2000). Kandemir has reported a low level of variation among the honey bee populations of Türkiye (Kandemir 1999).

Smith *et al*. (1997) showed that honeybees of Anatolia belongs to east Mediterranean mitochondrial lineage (C) in their work on disagnostic mtDNA sites. Four diagnostic restriction sites and a noncoding sequence in mtDNA were analysed among 16 honeybee populations of Türkiye. Three of the four mtDNA haplotypes detected in Türkiye were belonged to eastern Mediterranean lineage (C). But one mtDNA haplotype, detected in almost 50 % of Hatay samples, was novel for four restriction sites and a noncoding sequence. This haplotype was detected in almost 50% of Hatay samples. This haplotype does not belong to any of the three mitochondrial lineages (A, M and C) and may represent a new mitochondrial lineage. Kandemir *et al*. (submitted) found an African mtDNA haplotype in six colonies from Hatay. This region is known as the location where African faunal elements entered Anatolia (Kosswig 1955). Hatay samples clustered with *A. m. meda* and *A. m. lamarckii*, strengthening the argument for a different phylogeographic origin for this haplotype. Restriction site and sequence analyses of mitochondrial DNA in honeybee populations of Türkiye supported the previous findings that Türkiye honeybees primarily belong to eastern Mediterranean lineage (C). Central and western Türkiye

honeybees (*anatoliaca*) were in close relationships with northern Mediterranean bees (submitted).

Ruttner (1998) stated that his morphometric groupings may not represent true phylogenetical history. There is no exact match between morphometric and mtDNA based honeybee analyses. For instance, according to Ruttner (1998) *anatoliaca* and *caucasica* subspecies belong to O branch, however these two subspecies are found to belong eastern Mediterranean (C) branch based on mtDNA analyses.

Detection of a different restriction site pattern in Thrace honeybee samples which is also found in *A. m. carnica* maternal gene flow suggests that a maternal gene flow between the bees of Thrace, Balkans and southern Austria.

Honeybee populations of Thrace has been shown to be distinct from Anatolian populations by allozyme, morphometry and micosatellite analyses (Kandemir *et al.* 2000, Bodur 2001). Ruttner's (1988) suggestion that Anatolia is close to the center of speciation of *A. m. mellifera* is supported by a high diversity in mtDNA and allozymes found in Anatolia.

Microsatellite variation among five populations from Türkiye and one Cyprus population was studied using 5 loci and average heterozygosity levels changing between 0,502 and 0,687 were found (Bodur *et al.* 2004). Genetic variation among honeybee populations were reported to still existed although migratory beekeeping activities that cause gene flow.

1.2. Microsatellites

Microsatellites are short (2-6 nucleotides), tandemly repeated DNA sequences that are ubiqitiously interspersed in eukaryotic genomes (Tautz *et al*. 1999). They are present in prokaryotes in only low numbers. The larger repeat units (10-30 bp) form minisatellites which is differing in mutation mechanisms also (Ellegren 2004). Microsatellites are sometimes called short sequence repeats (SSR). Their variability, codominant inheritance and abundance cause them to be exploited as genetic markers in population and evolutionary genetics (Di Rienzo *et al*. 1998), linkage analyses and genetic mapping studies.

There is no consensus on the lower limit for iterations of a repetitive sequence. Also there is no a certain rule about how imperfect a microsatellite sequence can be. In many microsatellites there are interruptions between tandem repeats and even in a microsatellite more than one repeat motifs may occur. Most of the microsatellite repeats are known to be located on intergenic regions or introns and thus these markers are accepted as neutral markers. If they were on coding regions selection pressure would inhibit frameshift expansions. Some expanded trinucleotide repeats seen in human diseases are exceptions since they are on coding sequences. These repeats are not sharing similar mutational processes with the ones used in population genetics (Ellegren 2004).

Their high variation, abundance and genome wide distribution makes microsatellite markers extremely useful in population and evolutionary genetic inference areas such as forensic science, parentage testing, conservation genetics and molecular anthropology (Sainudiin *et al*. 2004). Microsatellite mutation rates at human autosomal chromosomes were reported to change between 10^{-2} and 10^{-4} (Weber and Wong 1993). Microsatellites are so variable that even with a few loci, it is possible to obtain unique multilocus genotypes and thus they are effective also at individual level for discrimination studies together with relationship, population structure and classification studies (Estoup *et al*. 2002). Microsatellite markers have been showed to be very efficient in differentiating populations or groups of populations within a species (Bowcock *et al* 1994).

Considerably higher assignment scores for highly variable microsatellite markers than those found for moderately variable allozymes, were obtained (Estoup *et al*. 1995a). Interrupted microsatellites are believed to be less variable than uninterrupted ones since interruptions seem to stabilize the tract in core region (Estoup *et al.* 1995b). These high resolution (fast evolving) neutral genetic markers are generally identified by sizes (in base-pairs) of polymerase chain reaction (PCR) amplified fragments with designed primers based on flanking region sequence.

Ubiquitious occurrence of microsatellites is not possibly explained by chance events. Hundreds of microsatellite motifs may be available on chromosomes (Ellegren 2004). Their extensive availability leads to questions about genomic organization and microsatellite

evolution. Whether they have a function or they are just junk sequences, is a challenge to be solved.

Genome sequencing studies are providing us with a more comprehensive view of genomic distribution of microsatellites in different species. Sequencing results in eukaryotes show that microsatellite density is generally positively correlated with genome size (Ellegren 2004). Mammals have been found to have the highest density, but within mammals rodents have higher microsatellite density than humans (Ellegren 2004). Moreover in plant kingdom this correlation is not positive but seem to be negative (Ellegren 2004). These contrasting results in different genomes suggest that there are differences between species in mutation processes or repair mechanisms or both. Microsatellite density seems to be similar at intergenic and intron sequences and dependent on base composition which is expected when random generation of mutations is considered (Ellegren 2004). These markers have been reported to have a higher density near chromosome arms in genome sequence studies of human and mouse (Ellegren 2004).

1.3. Mutation Mechanisms and Evolution Models for Microsatellites

Dynamics of microsatellite evolution are not resolved yet. Actually they have just poorly understood. These complex mutation processes are known to be influenced by DNA slippage, mismatch repair system efficiencies in different species, length constraints, selection, point mutations, repeat numbers, repeat types, flanking regions, recombination rates, sex and age (Schlötterer 2000).

Two mutational mechanisms that generate variability were proposed initially: replication slippage and unequal recombination between homologous chromosomes. Among the mutation mechanisms of microsatellites, the DNA slippage is the predominant one. DNA slippage is observed to occur when microsatellite repeat length exceeds 7 typically (Sainudiin *et al*. 2004).

In DNA slippage DNA polymerase enzyme pauses during DNA replication and dissociate from template DNA and this causes terminal portion of nascent DNA to be disattached from template. After pause nascent DNA realignes to another repeat unit on the template.

Most of these misassociations are repaired by mismatch repair system in the organism, it is the small amount of mismatches that could not repaired that lead new microsatellite alleles having more or less repeats in the array. Empirical studies generally indicate replication slippage (Samadi *et al*. 1998) as the main mechanism. According to a simulation study there is no evidence that unequal recombination between homologous chromosomes is taking role in evolution of most microsatellites (Samadi *et al*. 1998).

Recombination events like gene conversion and unequal crossing over have little evidence to contribute microsatellite evolution (Ellegren 2004). No correlation could have been found between recombination rates and microsatellite density and also no evidence is available that there is obvious difference between autosomal and Y chromosome linked regions for microsatellite distribution and mutation pattern (Ellegren 2004). Y chromosome is not involved in meiotic crossing over. These kind of recombination like events are thought to lead mutations in minisatellites actually (Ellegren 2004).

There could not found any association between microsatellite variation and recombination rates in a study of human dinucleotide microsatellites (Huang *et al*. 2002). This result was reported to be consistent with previous results obtained in *Drosophila* and *E.coli* studies (Huang *et al*. 2002). In an *E. Coli* and yeast study the mutations that eliminates recombination events in these organisms, any change in microsatellite stability has not been observed (Levinson and Gutman 1987). However in Schug *et al*.'s (1998) study on *Drosophila melanogaster* a strong positive correlation was observed between microsatellite variation and recombination rate.

1.3.1. Mutation Models

For a neutral marker the polymorphism is directly related with mutation rate. Although these markers have been extensively used in population genetics in recent years all the proposed theoretical evolution models for microsatellites failed to fully explain the allelic distribution patterns in natural populations (Ellegren 2004). A better understanding of mutation mechanisms and evolutionary properties of microsatellites is a prerequisite for interpretation of microsatellite data in population genetics. Findings so far show that the

mutation process is differing among loci and species. Rates and mutations patterns seem heterogeneous (Ellegren 2004).

1.3.1.1. Basic models

Stepwise mutation model (SMM) and infinite alleles model (IAM) are the two basic mutation models introduced for genetic markers. SMM states that microsatellite alleles evolve with addition or loss of one repeat motif and with an equal probability for addition and loss (Huang *et al*. 2002). Thus SMM predicts that the newly formed allele is possibly an allele that is already present in the population (Estoup *et al*. 1995a). However IAM predicts that a mutation event causes a change of any number of repeat units and always creates a novel allele which did not existed in the population (Estoup *et al*. 1995a).

SMM is attractive to researchers since it can easily be modelled and contains information about the closeness of alleles based on their repeat lengths. On the contrary, infinite alleles model (IAM) based methods are preferred by some researchers which do not make assumptions on the relationships between different alleles (Anderson *et al*. 2000).

1.3.1.2 Alternative models

The classical microsatellite evolution model, SMM, has two major weaknesses: first, it does not introduce an equilibrium distribution for allele lengths and second, it cannot explain the absence of very long microsatellite alleles (Huang *et al*. 2002). There are many studies that reports the occurrences of multi-step mutations in microsatellite alleles which seriously undermines this model (Huang *et al*. 2002).

In SMM, allele number is free to increase infinitely, but it is apparent that number of allelic states is finite (Paetkau *et al*. 1997). This could be explained by them being highly constrained (Ostrander *et al*. 1993). An equilibrium stage for microsatellite length distribution seems not possible by original SMM (Ellegren 2004). Actually microsatellites show an upper limit for size and this cannot be explained by original SMM (Ellegren 2004).

Different mutation models were introduced as alternatives to SMM which include two phase stepwise mutation model (TPM), one allowing an upper length constraint and mutation rate changes among loci (Feldman *et al*. 1997), biased models and ones introducing length constraints because of deletions or point mutations (Garza *et al*. 1995, Kruglyak *et al*. 1998).

The two phase model (TPM) allows for mutations of one repeat unit and more than one repeat units at one time (Sainudiin *et al*. 2004). According to both models, SMM and TPM, mutation rates are constant independent of repeat length and there is no mutational bias in favor of contraction or expansion. Hence microsatellites are predicted to increase or decrease in length unconstrained through time (Sainudiin *et al*. 2004).

Proportional slippage model (Kruglyak *et al*. 1998) is an alternative to SMM and leads to a stationary distribution phase which fits well to observations on humans, mice, fruit flies and yeasts (Kruglyak *et al*. 1998). This is a symmetric model assuming expansion or contraction is equally possible for microsatellites, slippage is proportional to repeat length and point mutations break large microsatellites. In an interspesific study length variaton predicted by this model was found to be higher than the observed values. This could be explained by a contraction bias which is supported by a *Drosophila* study (Calabrese and Durrett 2003). On the contrary there are other studies on human pedigrees and barn swallows that show a bias for expansion (Amos *et al*. 1996,Primmer *et al*. 1996).

Another view which may solve this contrast about the upward and downward bias is that there may be a target microsatellite length which is tried to be attained by either contraction if allele is larger than target length or expansion if allele is shorter than the target length (Garza *et al*. 1995).

In symmetric (i.e. rates of slippage up and down are the same) PSwK model, slippage occurs only when the microsatellite length exceeds a treshold value (Calabrese and Durrett 2003). Another model is the constant exponential model (ConExp) which assumes a constant expansion rate but an exponentially increasng rate for contraction (Calabrese and Durrett 2003). In assymmetric linear (AsyLin) and quadratic (AsyQuad) models the up and down slippage rates were different linear and quadratic length dependent functions respectively (Calabrese and Durrett 2003). Another asymmetric model is piecewise linear

bias model (PLBias) which assumes a constant mutation rate but an upward or downward bias is a linear function dependent on microsatellite length (Calabrese and Durrett 2003).

1.3.1.3. Testing the models

Because of high mutation rates direct observations of microsatellite mutations give us an opportunity to try to understand which of the proposed evolution models for microsatellites is closer to the actual process of evolution (Ellegren 2004). In additon to these direct genome sequence and pedigree analyses, computer simulations which were run by certain assumptions to be tested against heterozygosity measures and microsatellite distributions in genomic databases, are serving us for this purpose (Ellegren 2004).

1.3.1.3.1 Direct Observations

There are allelic distribution and pedigree analysis studies supporting SMM together with several other studies showing deviation from SMM (Huang *et al*. 2002). Slatkin and Goldstein argued that IAM is not appropriate to apply for microsatellites since they have high mutation rates and mutational process retains memory of ancestral allelic states (Slatkin 1995, Goldstein *et al*. 1995). Valdes, Slatkin and Freimer (1993) reported allelic frequencies found for 108 dinucleotide human microsatellite loci were consistent with SMM.

Pedigree analyses and genomic sequence analyses of microsatellite loci showed that mutational processes are heterogeneous among species, repeat types and loci. Single step SMM is not supported by evidence since many mutation events containing changes at more than one repeat unit are observed (Ellegren 2004). Studies on human pedigrees, swallows give evidences for SMM (Weber and Wong 1993, Primmer *et al*. 1996,1998). However some sequencing studies revealed that indels in flanking regions are playing important role in generating microsatellite variation (Angers and Bernatchez 1997). Five out of 12 sequenced loci showed multiple sources of length variation which cannot be explained solely by gain or loss of one or two repeats as in the case of SMM based models. Indels in flanking regions, and microsatellite containing minisatellites were sources of variation (Anderson *et al*. 2000).

Flanking regions of microsatellites are relatively conserved among different animal groups (Moore *et al*. 1991). This conservation is confirmed by one order of magnitude lower mutation rate at flanking regions of a salmonid locus than mutation rate in microsatellite region (Angers and Bernatchez 1997). Within microsatellite loci among species, several mutation types which do not conform to SMM were observed both within repeat arrays and non-repeat sequences in addition to repeat number changes (Angers and Bernatchez 1997). Similar complex mutational patterns that show deviation from SMM were also reported at within species level (Estoup *et al* 1995b).

Many other observations showing that some microsatellite loci do not obey SMM were reported. In a dinucleotide *Drosophila melanogaster* microsatellite, both single step and larger mutations were detected. In a barn swallow tetranucleotide locus 7 out of 44 mutations were shown to involve 2-5 repeat unit changes when the remaining mutations followed single unit changes (Primmer *et al*. 1998). According to Jones *et al*. (1999) 23 out of 26 mutations at a tetranucleotide microsatellite locus of pipefish *Syngnathus typhle* had mutations conform to SMM, but three other mutations contain multi-unit changes. Shriver *et al*. (1993) showed that 35 % of the mutations in a dinucleotide locus were not congruent with SMM while, the remaining in this locus and and tri-penta loci obeyed SMM. In a ten microsatellite loci study among Sardinian human population, allelic frequency distributions fit to TPM (Colson and Goldstein 1999). In an extensive work in 3 closely related species of *Drosophila* microsatellites, only 7 out of 19 loci were reported to show variation consistent with SMM (Colson and Goldstein 1999). In this study 63 % of dinucleotide microsatellite mutations in humans showed multistep changes. Observed and expected values of number of alleles and heterozygosities were used to test the adequecies of both IAM and SSM. It was reported that IAM could never be ruled out for the studies on 7 microsatellite loci 4 of which have more than one repeat type which is likely to prevent evolution under SMM (Estoup *et al*. 1995b).

Resolution power of microsatellites decreases with evolutionary time under SMM which is understood by higher proportion of stepwise mutations at within species level than between species level (Angers and Bernatchez 1997). However a study of a imperfect microsatellite locus in salmonid species showed that complex non-stepwise mutations are also involved

between closely related populations and even within alleles of the same population (Angers and Bernatchez 1997).

Imperfect microsatelites are relatively common in animal genomes and routinely used in microsatellite studies (Weber 1990). Base substitutions may be the driving forces for derivation of imperfect microsatellites from perfect ones since such mutations interrupt contiguous repeat arrays and thus reduce slippage probability (Angers and Bernatchez 1997). Since a minimal number of repeats is neceessary to create a microsatellite variability (Weber 1990) these events reduce variability sharply.

Anderson *et al*. (2000) suggested that IAM based models are more suitable than SMM based ones for many microsatellite loci in *Plasmodium falciparum*. The rate of rearrangements have been reported to be much higher than rate of point mutations in trinucleotide repeat microsatellites of *Plasmodium falciparum*. Mutation rates for di and trinucleotide loci have been reported to be more positively correlated with repeat length than repeat type in this study.

Assumptions of SMM such as infinite population size, sufficient number of alleles and random mating are rarely met in the nature. These results obviously show that microsatellite mutational processes are more complicated than SMM predicts. Thus caution should be taken in order to use SMM to understand genetic relatedness of natural populations. Microsatellite loci that are known to follow this model must only be used to calculate distance measures assuming SMM (Huang *et al*. 2002).

There are contrasting results about the directionality of mutations in microsatellites. Many observations showed that direction of microsatellite mutations are in favor of expansion rather that contraction of microsatellites. But there are also other studies that did not report a bias between gain and loss of repeat units. Moreover presence of some studies showing that long alleles have bias toward contraction may help to understand the stationary phase of microsatellite lengths in genomic distribution as well as increasing the complexity of mutational processes in microsatellites. In a study performed on dinuclotide microsatellites on human autosomes, an overall upward bias has not been observed for microsatellite legth (Huang *et al*. 2002). Instead a size dependent bias has been detected. Longer alleles had a tendency to lose repeats more than the shorter alleles and shorter alleles had a higher

tendency to gain repeats than the longer ones did. Consistent with these results some other studies also showed that contraction is more common longer alleles. In a study on human tetranucleotide microsatellites Xu *et al*. (2000) found that contraction rate increases exponentially with allele size but expansion rate remains constant.

Longer alleles have more chance to be broke by point mutations and this decreases the mutation rate making these alleles more prone to contract toward a focal length than expansion (Huang *et al*. 2002). A lineage specific variation is the case for pure AC repeats studied on humans and chimps (Sainudiin *et al*. 2004). There may be two sound explanations for this difference in different lineages. The first is the differences between efficiencies of mismatch repair systems in different species and the second, selection against longer alleles and differences in effective population sizes (Sainudiin *et al*. 2004).

1.3.1.3.2. Simulation studies

According to computer simulations, mutation and genetic drift cannot alone explain microsatellite evolution in the long term. Both lower and higher allelic size limits should be assumed to obtain an equilibrium state of allelic distribution. Either a selection on allelic size or an upward biased asymmetric mutation process could make this possible (Samadi *et al*. 1998).

Three asymmetric models out of 7 models have been found to show best fits for every dinucleotide repeat motif type to the genomic data from both humans and *Drosophila* in a simulation study of uninterrupted microsatellites (Calabrese and Durrett 2003). Hence bias up or down were changing according to functions based on microsatellite lengths. Moreover for long microsatellites this bias was in the favor of contraction always (Calabrese and Durrett 2003). An equilibrium distribution was reached by every model since it was assumed that point mutations break microsatellites whose rate is proportional to repeat length. These length distributions have been used to calculate likelihoods of the genomic data for each model. All simple symmetric models failed to explain microsatellite length distribution.

Mutational bias and proportionality between mutation rate and repeat length were found to be necessary components of a realistic mutation model for pure dinucleotide microsatellite

data homologous between humans and chimpanzees in another simulation study (Sainudiin *et al*. 2004). This study indicated that the models best fit to the real data were the ones with a linear bias toward a focal length. Together with Garza *et al*. (1995) and Zhivotovsky *et al*.'s (1997) models, these results support Calabrese and Durrett's (2003) findings about the insufficiency of proportional slippage in the absence of mutational bias to predict equilibrium distributions of human microsatellites. The observed linear bias may be explained by counteracting mutational forces in microsatellites which means that an upward bias caused by slippage event could be balanced by a downward mutational bias in longer alleles because of mismatch repair system (Harr *et al*. 2002). Natural selection may also be in action in favor of contractions dirctly when longer microsatellites confer a disadventage on indirectly by affecting mismatch repair system. In unbiased models repeat lengths reach to unrelistically large values when upper bound parameter is high (Sainudiin *et al*. 2004). Two-phase models did not prove to be significantly better than one-phase models. Two phase models were reported to mimic one-phase models to fit the real data (Sainudiin *et al*. 2004). Some variation in microsatellite alleles have been reported to be caused by indels in flanking regions which are amplified with core sequences (Angers and Bernatchez 1997) . This variation may be attributed to multi step changes in some emprical studies (Sainudiin *et al*. 2004).

1.3.1.4. Choosing the model

Microsatellites are known to deviate from SMM frequently (Takezaki and Nei 1996). For obtaining correct tree topology, details of mathematical model of microsatellite evolution were found to be unimportant for phylogeny reconstruction (Takezaki and Nei 1996). However evolutionary processes in microsatellite allele genesis are very complicated and seem to involve an upper limit for alleles. Moreover microsatellite polymorphism may change drastically between different populations. These factors should be accounted when computer simulations are extrapolated (Takezaki and Nei 1996).

An important question in microsatellite evolution is: What prevents infinite growth? So far studies indicated that the answer contains biased mutations in microsatellites and a balance betwen DNA slippage and point mutations and selection (Huang *et al* 2002). An ideal microsatellite evolution model should consider mutational bias and a balance betweeen slippage and point mutations (Huang *et al*. 2002). In order to be able to make correct

inferences in such areas, biologically realistic models of microsatellite evolution should be developed.

According to a study, an interrupted honeybee microsatellite, A113, does not follow SMM but mutational processes follow IAM (Estoup *et al*. 1995b). Same conclusion also holds for another interrupted microsatellite locus, B121, in bumblebees. On this locus, rather than single unit jumps, multi unit jumps and differences in the location and number of interruptions occur to create new alleles. More complex events like gene conversion and unequal recombination should be considered to understand the allelic distribution at A113 locus (Estoup *et al*. 1995b).

So far any ideal mutation model did not prove to be valid in all cases for microsatellites. This probably reflects the much more complex nature of mutational processes than to be evaluated by existing models and which show variation among microsatellite loci. Although their evolution is poorly understood, microsatellites are very useful to study closely related populations since classical markers are not sufficiently polymorphic in many cases for this purpose (Takezaki and Nei 1996).

1.4. Size Homoplasy in Microsatellites

Homoplasy is a term which is used for genetic markers in evolutionary genetics. It is said to occur when different copies of a locus are identical in state but not identical by descent. The similarity between these copies from different ancestors may be due to convergence, reversion or parallism. Mutations create these "identical in state" alleles, thus the way that mutations occur for that genetic marker is important for this phenomenon (Estoup *et al*. 2002).

Microsatellite alleles correspond to PCR amplified and electrophoretically sized DNA fragments which contain flanking regions together with microsatellite repeats. That is why homoplasy in microsatellite electromorphs is called "size homoplasy". Electromorphs are identical in state (same size), but may not be identical by descent. They may be descendants of different alleles that mutated in different ways (Estoup *et al*. 2002).

Size homoplasy is expected for microsatellites under SMM based mutation dynamics since every new mutation at allele "i" creates "i+1" or "i-1" alleles with equal probality in this model. But this "evolutionary noise" is not expected under IAM since every allele mutates to a novel one not already present in the population. Other than mutation models, size homoplasy depends on evolutionary factors such as divergence time, effective populations size and mutation rate (Estoup *et al*. 2002). Size homoplasy may take place among closely relates species and even within a species. Thus allele polymorphism, heterozygosity and genetic distances may be understimated (Van Oppen *et al*. 2000). The occurrance of size homoplasy is expected to increase with time of divergence among populations and mutation rate (Estoup *et al* 1995b). Size homoplasy is a drawback of microsatellites to infer population parameters such as genetic distances, effective population sizes and migration rates (Estoup *et al*. 1995b). Allele size constraints and homoplasy that homogenize mutations, possibly limit usefulness of microsatellites (Richard and Thorpe 2001).

A fraction of size homoplasy in microsatellite electromorphs can be detected by single stranded conformational polymorphism (SSCP) or DNA sequencing since alleles which are not identical by descent may contain different sequences in repeat region (e.g. interruptions) or within flanking regions. This fraction of size homoplasy is called molecularly accessible size homoplasy (MASH) (Estoup *et al*. 2002).

To make inferences about size homoplasy from MASH is problematic since this relation is affected by different evolutionary factors such as mutation rate, mutation model, effective population size, and type of microsatellite loci (Estoup *et al*. 2002). Variation in the amount of MASH was reported between different microsatellite loci. (Garza and Freimer 1996, Viard *et al*. 1998). Interrupted and compound microsatellite loci represent suitable candidates for MASH studies. For example, microsatellites with core regions (AT)nTT(AT)mAT(AT)x or (AT)n(CT)m are useful for detecting size homoplasy since same size electromorphs of these loci may represent different alleles with point mutations at interruptions or different combinations of repeat units respectively. But a significant fraction of size homoplasy remains undetected since their sequence is the same. An important problem in estimating size homoplasy from MASH is the less homoplasious nature of perfect microsatellites that have pure repeat motifs than compound or interrupted microsatellites. This is because size homoplasy is not detectable for pure repeats unless a

mutation occurs in flanking region which is rare when compared to mutations in repeat region (Estoup *et al*. 2002).

MASH studies showed that size homoplasy is lower among populations of same species than among species and even rarer at within population level (Estoup *et al*. 2002). Noise effect of homoplasy is important for phylogeny receonstruction but its effect on population genetic studies at intraspecific level is crucial to understand. Theoritical simulations and emprical MASH studies showed that size homoplasy causes a decrease in allelic polymorphism and heterozygosity (Estoup *et al*. 2002).

Although genetic markers are performing better under non-homoplasious IAM than homoplasious SMM, for various genotype assignment methods, mutation models seem to be less important than the variability of selected genetic marker (Estoup *et al*. 2002). Between closely related populations, the genetic divergence is mostly related with genetic drift. Thus it is not expected that mutation model and size homoplasy are very effective at this level which is supported by findings that classical distance measures D_S and D_C which do not consider size homoplasy, perform better to construct phylogenies than SMM based $(\delta\mu)^2$ distance (Takezaki and Nei 1996). However for distantly related populations in which divergence is at mutation-drift equilibrium SMM based models, which take allele size differences into account, perform better phylogeny reconstruction (Goldstein and Pollock 1997). This shows that effect of size homoplasy is higher for studying distantly related populations.

Sequencing uninterrupted microsatellite alleles may not provide information about size homoplasy, but number and location of interruptions introduce a new level of interruption for interrupted microsatellites. Repeat number of interrupted microsatellites has a large variance and thus these loci have lower size homoplasy than pure repeat microsatellites. Hence genetic information saturation effect of size homoplasy is slower in interrupted microsatellites and these loci are more suitable for studying distantly related populations (Estoup *et al*. 1995b).

The phenomenon of size homoplasy has been evidenced in honeybees when electromorphs of the same size from different lineages were sequenced (Estoup *et al*. 1995b). Most of the electromorphs seemed to have different sequences for an interrupted locus A113. However

sequences of electromorphs of the same size were identical when they are sampled from same population and even when they are sampled from populations belonging to the same honeybee lineage. For all electromorphs, the flanking region sequences of A113 microsatellites were found to be identical in all *Apis mellifera* subspecies and lineages studied. Thus size homoplasy has not been detected in honeybees from same subspecies and even in the individuals from same lineage for A113 locus. But for distantly related populations (from different lineages) size identity did not prove identity by descent and hence size homoplasy may cause underestimation of genetic distances between such distantly related populations. Interrupted microsatellites are believed to be less variable than uninterrupted ones since interruptions seem to stabilize the tract in core region. (Estoup *et al*. 1995b)

Size homoplasy were reported to not represent a significant problem for many purposes in population genetic studies and high variability of microsatellite markers compensate to a high extent for the reduction in polymorphism due to homoplasy (Estoup *et al*. 2002). Hence MASH data obtained for routine population genetic studies is not essential in most cases. In closely related populations increasing the number of microsatellite loci is more important than focusing on mutation model and size homoplasy (Estoup *et al*. 2002).

1.5. Genetic Distance Measures

In populations genetic studies, microsatellites are exploited to understand relatedness among populations or species and to reconstruct phylogenies. In order to achieve this, one should calculate a genetic distance measure. There are some genetic distance measures specifically designed for microsatellite data. However a real disadvantage for these measures is that they assume that the microsatellite evolution in nature obeys the SMM

Genetic distance statistics based on SMM use variance in repeat numbers, however the statistics based on IAM use variance in allelic frequencies (Richard and Thorpe 2001). Since mutational processes in microsatellites are not following only one model in different conditions it is not suitable to talk about an ideal genetic distance statistic for these markers (Richard and Thorpe 2001). Large variances cause poorer performance of SMM based

statistics than IAM based statistics unless sample size and locus number is very high (Gaggiotti *et al*. 1999).

Nei's (1972) standart genetic distance D_S, Nei's (1973) minimum genetic distance, Latter's (1972) F_{ST} distance, Rogers' (1972) distance D_R, Cavalli-Sforza and Edwards' (1967) chord distance D_C, Nei *et al*.'s (1983) D_A distance, Shangvi's (1953) X^2 distance, Goldstein *et al*.'s (1995) $(\delta\mu)^2$ distance and Shriver *et al*.'s (1995) D_{SW} distances were tested for their performances under both IAM and SMM to be used with microsatellites (Takezaki and Nei 1996). D_A and D_S distances were found to be the best ones in obtaining correct phylogenetic tree topology both under IAM and SMM under various conditions. However D_S and $(\delta\mu)^2$ were reported to be more useful for branch length estimations under IAM and SMM respectively (Takezaki and Nei 1996). Different distance measure are suggested to be used for different purposes (Nei *et al*. 1983).

When the divergence is high sample size (at least 20) were reported not to matter for correct topology performances under both SMM and IAM. Sample size again is not important at low divergence, as among closely related populations when average heterozygosity is not high. But when heterozygisity is high among closely related populations (0.5 for IAM and 0.8 for SMM) then large sample size (up to 50) increases performance in giving correct tree topology (Takezaki and Nei 1996).

It is essential to test performances of genetic distances on microsatellite data using organisms of known evolutionary history. Traditional genetic distance measures performed better than distance measures specifically designed for microsatellites in another study between arctic brown bears from adjacent areas where climate, latitude and habitat was similar without any barrier to movement (Paetkau *et al*. 1997). Among six tested genetic distance statistics (D_S, D_A, D_m, D_{SW}, $(\delta\mu)^2$ and D_{LR}), D_S and D_{LR} performed extremely well when genetic distance graphs were drawn against geographic distances. All distances had significant linear regressions on geogrophical distance except $(\delta\mu)^2$ which did worst possibly because of large variance. At the continuous variation scale the main mechanism of evolution is drift and thus choosing correct mutation model was not of crucial importance. When data from distantly located populations were used, every genetic distance measure lost linearity after relatively short period of independent evolution. Power of microsatellites at interspecific level population genetic studies seem to be low since even

$(\delta\mu)^2$ plateaus after very short periods of time in evolutionary terms. Every genetic distance statistics was affected by heterozygosity levels within studied bear populations which further complicated the extrapolation of results (Paetkau *et al*. 1997).

A novel distance likelihood ratio distance (D_{LR}) (Paetkau *et al*. 1997) which is based on a genotype assignment test was reported to be suitable as an independent measure to confirm the relationships that D_S suggested (Paetkau *et al*. 1998). Nei's standard distance, D_S, is calculated from genotypic frequencies and D_{LR} is calculated from genotype probabilities. These two distance measure had a correlation have a high correlation although they treat data in radically different ways in a population genetic study (Paetkau *et al*. 1998). Estimates of D_S and D_{LR} parallelled the results from both pairwise F_{ST} and assignment tests (Kyle and Strobeck 2001). These two distance measures were reported to be able to provide meaningful insights into biological relationships even for 8 microsatellite loci (Paetkau *et al*. 1998).

Phylogenetic reconstruction assumes that the effect of migration is not important when compared to mutation. Thus for microsatellites to be useful in phylogeny reconstruction mutation rate should be much higher than the migration rate but not high enough to cause size homoplasy to cause problems. Hence, microsatellites must be most useful in phylogeny construction of closely related, small, allopatric populations (Richard and Thorpe 2001). Currently population phylogenies are mostly based on mitochondrial DNA data and microsatellites as nuclear markers could be used to independently test these phylogenies. In a study where performance of microsatellites in phylogeny reconstruction were tested against phylogenetic trees based on mtDNA genetic distances in 12 populations of western Canary Island lizards *Gallotia galloti* using 5 microsatellite loci (Richard and Thorpe 2001). With moderate sample size (30) and a limited number of microsatellite loci (5) IAM based metrics performed better than SMM based metrics to elucidate the historical relationships among populations. It is possible to construct a phylogenetic tree compatible with mtDNA constructed ones by using relatively low number of microsatellite loci as shown with works of Estoup *et al*.'s (1995a) on honeybees, Berube *et al*.'s (1998) on fin whales and Forbes *et al*.'s (1995) on sheep with 7, 6, and 6 microsatellite loci respectively. SMM based microsatellite distances are more sensitive to recent demographic changes (e.g. bottlenecks) in populations than IAM based classical genetic distances and thus perform poorly with moderate sample sizes and few loci (Richard and Thorpe 2001).

In general it is believed to be more important to use more microsatellite loci than to increase sample size except if average heterozygosity level is high when closely related populations are under study to obtain correct phylogenetic tree (Takezaki and Nei 1996). Hundreds of microsatellite loci is needed to calculate divergence times correctly (Zhivotovsky 1999). However to construct a correct topology is possible with a much lower microsatellite loci (Zhivotovsky 1999). It seems that different distance measures could be used for microsatellites according to different complicated mutational events the follow (Zhivotovsky 1999).

CHAPTER 2

MATERIALS AND METHODS

2.1. Biological material

In population genetics, sampling is of crucial importance. Random samples that we collect should reflect the actual variation in natural populations. Because the honeybee workers of individual colonies are generally descended from a single queen, it is not preferred for a location to be sampled extensively from few colonies. Instead, collecting few worker bees from a high number of colonies is more suitable.

We have used 349 honeybee workers collected from 45 different locations belonging to 12 provinces (Figure 1). We sampled only one or two individuals per colony from the laboratory stock except Artvin which we have sampled. Names of the provinces, locations and number of bees collected from each are given in Appendix A. Samples have been kept in absolute ethanol until DNA isolation.

Figure 1. Sampling areas.

2.2. DNA Isolation

Bee heads were removed after taking the bees out of alcohol. Each head was then grinded in a 1,5 ml tube with a sterile pestle immediately after immersing the tube containing head into liquid nitrogen and 750 µl of Wilson buffer (Appendix C) was added into the tube. Twenty five µl of 10 mg/ml Proteinase K was added into each tube. After mixing briefly, the tubes have been incubated for two hours in a water bath at 50°C. After a centrifugation step at 10000 rpm for 10 minutes, the upper phase solution was poured into a new tube. Seven hundred and fifty µl of phenol:chloroform:isoamylalcohol (25:24:1 vol.) was added and tubes were centrifuged at 10000 rpm for 20 minutes after gentle inversions of five minutes. Then 600 µl of aqueouse phase was removed into a new tube and same extraction procedure was performed twice first by adding 600 µl of phenol:chloroform:isoamylalcohol (25:24:1 vol.) and then 450 µl of chloroform: isoamylalcohol (24:1 vol.) to the removed 450 µl of aqueouse phase. Recovered 300 µl of aqueouse phase was transferred into a new tube and added with 30 µl of 3 M sodium acetate, 600 µl of absolute alcohol and stored at -20 ºC overnight after mixing for a few minutes.

The tubes were centrifuged at 13,000 rpm for 30 minutes and the supernatant was discarded. 900 µl of 70% ethanol was added and the tubes were centrifuged at 13000 rpm for 20 minutes. After pouring alcohol off, the pellet was dried in a desiccator for 30 minutes. The pellets in the tubes were added 50 µl of sterile water and kept at room temperature for one hour. DNA solutions were examined under UV illumination at 230, 260 and 280 nm for detection of absorptions of RNA, DNA and protein parts respectively, if available in solution and run on 1 % agarose gel electrophoretically to confirm the presence of DNA.

2.3. Microsatellite amplification by PCR

Nine *Apis mellifera* specific microsatellite loci namely;A24, A113, A7, A43, A28, Ap226, Ap43, Ap68 and Ac306 (Solignac *et al*. 2003) were exploited in this study whose core regions, primer sequences and polymerase chain reaction (PCR) conditions are given (Table 1 and 2).

PCR amplifications of genomic sample DNAs were performed as reported in Estoup *et al*. (1995). Twenty five microliter of amplification reactions were performed with 50 ng of template DNA, 400 nM of each primer, 75 µM of each 2'-deoxythymidine 5'-triphosphate (dTTP), 2'-deoxyguanidine 5'-triphosphate (dGTP) and 2'-deoxycytidine 5'-triphosphate (dCTP), 7.5 µM of 2'-deoxyadenosine 5'triphosphate (dATP), 0.25 µCi of α^{33}P-dATP, 20 µg/ml bovine serum albumin (BSA), 1x reaction buffer containing $(NH_4)_2SO_4$, 0.4 unit of Taq polymerase and 1-1.2 mM $MgCl_2$. PCR started with a denaturation step of 3 minutes at 94 ºC and continued with 30 cycles, containing; a 30 second denaturation segment at 94 ºC, a 30 second annealing segment at the optimum temperature, and a 30 second elongation segment at 72 ºC. The final elongation step was extended to 10 minutes in order to allow all the products to be fully extended. The annealing temperatures and $MgCl_2$ concentrations that were used for each microsatellite loci are given in Table 1.

Table 1. Core sequences, Magnesium concentrations (M) and annealing temperatures (ºC) of microsatellites used in polymerase chain reactions.

Locus	Core Region	Mg	$T_{annealing}$
A24	$(CT)_{11}$	1,2	56
A113	$(TC)_5TT(TC)_8TT(TC)_5$	1,2	60
A7	$(CT)_{24}$	1,0	60
A43	$(CT)_{12}$	1,5	55
A28	$(AG)_6(GAG)_6$	1,7	55
Ap226	$(CT)_8$	1,5	50
Ap43	$(TA)_6GATA(GA)_{10}$	1,2	60
Ap68	$(CT)_{12}(TA)_8$	1,5	50
Ac306	$(CT)_{11}$	1,2	55

Table 2. Primer sequences (5'-3')

Locus	Forward primer	Reverse Primer
A24	CACAAGTTCCAACAATGC	CACATTGAGGATGAGCG
A113	CTCGAATCGTGGCGTCC	CCTGTATTTTGCAACCTCGC
A7	GTTAGTGCCCTCCTCTTGC	CCCTTCCTCTTTCATCTTCC
A43	CACCGAAACAAGATGCAAG	CCGCTCATTAAGATATCCG
A28	GAAGAGCGTTGGTTGCAGG	GCCGTTCATGGTTACCACG
Ap226	AACGGTGTTCGCGAAACG	AGCCAACTCGTGCGGTCA
Ap43	GGCGTGCACAGCTTATTCC	CGAAGGTGGTTTCAGGCC
Ap68	TGTCTGCCCTCCTCTCTGTT	CACATCGAGCGAGAAGGC
Ac306	GAATATGCCGCTGCCACC	TTTCGTTGCATCCGAGCG

2.4. Sequencing polyacrylamide gel electrophoresis

Sequencing polyacrylamide gel electrophoresis apparatus (Owl S4S) has been used in order to achieve discrimination between alleles differing with one or a few nucleotides.

2.4.1. Cleaning the glass plates

Glass plates with edges of twenty and forty five centimeters were used in electrophoresis. One side of each plate were cleared carefully first with distilled water and then by absolute ethanol in order to prevent any debris on the surface to interfere with the progress of DNA fragments during electrophoresis. Then a silanizing solution was applied to one clean surface of a glass plate to make it easier to remove one of the plates after electrophoresis, the intact gel remaining on the other plate.

2.4.2. Preparation of the gel

A 6% denaturing polyacrylamide gel was used in electrophoresis. A 6% acrylamide-urea mix (Appendix C) containing 8 molar of urea was prepared and put in a light-tight bottle and kept at 4°C. Six hundred and fifty µl of 10% (v/v) ammoniumpersulfate (APS) and 30 µl of N,N,N',N'-tetramethylethylenediamine (TEMED) was added to 50 ml of acrylamide/urea mix just before pouring the gel.

2.4.3. Pouring the gel

A gel caster, a comb and 0.4 mm. plastic spacers were used. Gel mix was poured on one of the glass plates which is fixed horizantally in the gel caster, immediately after the addition of TEMED by using a syringe. The upper glass plate is slided slowly on the other plate as the gel is poured. Being spacers adhered by water drops to the lower plate, gel solution fills the area between the plates. After pouring the gel, comb is inserted and metal clamps were used at the edges of the plates to squeeze them.

2.4.4. Loading and running the gel

PCR reactions containing 25 µl of DNA solution were added 10 µl of loading dye solution (Appendix C) each and 2.5 µl of these mixes were loaded to the gel placed in the vertical gel apparatus using an ordinary micropipettor. A sequencing reaction done by USB Sequenase Version 2.0 DNA Sequencing Kit using α^{33}P-dATP, was exploited as size marker to determine the exact sizes of DNA fragments. Upper and lower reservoirs of sequencing gel electrophoresis apparatus were filled with 1x Tris-Boric Acid-EDTA (TBE) buffer, and it was run at 40 Watts and for 2,5 hours.

2.5. Autoradiography

After the run, siliconized plate was removed and the gel which remained on the other plate, was taken onto a chromatography paper (Whatman 3MM). Gel was covered with an ordinary stretch film and dried on a vacuum dryer at 80 °C for 30 minutes. Special autoradiography films (Kodak Biomax MR) handled in a dark room, were exposed to the dried gels in light-tight metal cassettes for 2-5 days depending on the time passed after the radioactive material purchased. The exposed films were developed in the medical center of Middle East Technical University.

2.6. Statistical Analyses

Statistical analyses, containing allele frequencies, heterozygosities, gene diversity, pairwise F_{ST} measures; population differentiation, Hardy Weinberg equilibrium, linkage disequilibrium tests and genetic distance calculations and phenogram constructions were performed using population genetic softwares.

2.6.1. Genetic variation

Number of alleles, numbers and frequencies of private alleles, allele frequencies were all calculated from raw allele frequency data obtained from Basic Information option of Genepop program which is available at the web address http://wbiomed.curtin.edu.au/genepop/ freely (Raymond and Rousset 1995). Allele frequencies are calculated for each population as the proportion of the observed number of the allele to the total number of alleles in that population.

Observed and expected heteozygosities for each population and each locus were calculated using Hardy Weinberg Option of Arlequin ver. 2.000 program (Schneider *et al*. 2000). Then the averages and standard deviations were calculated for each population. Observed heterozygosities are the proportions of heterozygote individuals within populations. Expected heterozygosity which is also called "gene diversity" for diploid data, may be defined as the probability of two randomly chosen haplotypes (genes) to be different in the sample (Nei 1987).

2.6.2. Genetic structure

Significance of departures from Hardy Weinberg Equilibrium were tested using Hardy Weinberg option of Arlequin ver. 2.000 program (Schneider *et al*. 2000). Tests are performed by the program by testing the null hypothesis that assumes random association of gametes as described by Guo and Thompson (1992). An initial contingency table is created by using observed allele counts and then alternative tables are prepared by decreasing and increasing certain counts by one unit each time.The P value calculated, corresponds to the proportion of the visited tables that have probabilities equal to or smaller than the original contingency table.

Presence of pairwise linkage between two microsatellite loci was tested using Linkage Disequilibrium option of Genepop program which is available at the web address http://wbiomed.curtin.edu.au/genepop/ freely (Raymond and Rousset 1995). All contingency tables containing counts of unions of different alleles of loci pairs are prepared for all pairs of populations and a probability test for each table is performed by the

program. Here the tested null hypothesis is: Genotypes at one locus are independent from genotypes at the other locus.

Genic and genotypic differentiation tests among 12 populations have been done using population differentiation option of Genepop program (Raymond and Rousset 1995). In genic differentiation test, the null hypothesis of identical allelic distribution among populations was tested and P values were calculated according to Raymond and Rousset (1995). The null hypothesis of identical genotypic distribution among populations was tested in genotypic differentiation test. An unbiased estimate of the P-value was performed according to Goudet *et al*. (1996).

Genetic distinctness of populations were analysed by calculating F coefficients, number of migrant (Nm) values, by performing assignment tests and by constructing phylogenetic trees based on genetic distances among populations.

The fixation index F_{ST} is the most inclusive measure of population substructure (Hartl and Clark 1997). It is used to analyse the genetic divergence among subpopulations of a total population. Theoretically F_{ST} measures changes between 0 (no divergence) and 1(fixation of different alleles in different populations). However the F_{ST} levels are generally much lower than 1. According to Wright (Hartl and Clark 1997) the F_{ST} levels between 0 and 0,05 indicate little genetic differentiation, between 0,05 and 0,15 indicate moderate level genetic differentiation, levels between 0,15 and 0,25 indicate great genetic differentiation and levels higher than 0,25 indicate very great genetic differentiaition.

Pairwise F_{ST} values were reported to be used as short-term genetic distances with a slight transformation (Reynolds *et al*. 1983; Slatkin 1995). Pairwise F_{ST} values and their P values giving the proportion of the permutations (Distribution of F_{ST} values under the null hypothesis of no difference among populations is obtained by permutation of haplotypes between the populations) giving an F_{ST} greater or same with the observed one, were obtained by using Population comparisons option of the Arlequin ver. 2.000 program (Schneider *et al*. 2000).

F_{IS} and F_{IT} are inbreeding coefficients that give deviations from Hardy Weinberg equilibrium within subpopulations and within the total population respectively. Positive values indicate a deficit and negative values indicate an excess of heterozygote individuals.

F_{ST}, F_{IS} and F_{IT} measures of the total population composed of 12 honeybee populations for each of 9 microsatellite loci were calculated according to Weir and Cockerham (1984) using Genepop software (Raymond and Rousset 1995) option 6.

Nm estimates for the total population and pairwise Nm estimates for population pairs were calculated using Nm estimates option of Genepop software (Raymond and Rousset 1995). This method exploits the average private allele frequencies to estimate the effective number of migrants per generation (Nm) (Slatkin 1985). Private alleles are the alleles that are observed in only one population. When Nm is smaller than 2 it is thought that there is still a considerable opportunity for genetic divergence among subpopulations (Hartl and Clark 1997).

The assignment tests were performed by using "Doh asignment test calculator" available online at http://www2.biology.ualberta.ca/jbrzusto/Doh.php. The method is based on the articles of Paetkau *et al*. (1995, 1997) and Waser and Strobeck (1998). The assignment test calculates the probability for an individual to be belong to the population it actually sampled and the probabilities of the same individual to be originally belong to the other populations in comparison. Then an individual is assgigned to the population that has the highest probability value for that individual. This is done by using allelic frequencies within populations. However the tested individual's genotype is removed from calculations when the allelic frequencies are calculated for each population. The population that has the highest probability for the emergence of tested genotype is assigned as the origin of tested individual.

The individuals that were assigned to the population that were actually sampled are called "correctly assigned individuals" in this thesis. The pairwise log likelihood graphics were drawn by taking the logarithms of likelihoods for individuals and placing these values for population pairs on a spreadsheet. On these graphics x=y line represents the region that the tested individual is equally likely to be from one or the other population in comparison.

Different data randomizations were done to test the overall structure of the total population consisting of 12 honeybee populations using the same assignment calculator. First data randomization were conducted by drawing existing individuals from combined gene pool of eleven populations with replacement to reform the populations. This randomization

assumes that 12 populations are actually one well mixed population. Second data randomization was done by drawing new individuals from combined gene pool of 12 populations with replacement to reform the populations. This randomization assumes that 12 populations are actually one well mixed population at Hardy-Weinberg Equilibrium. The third data randomization was applied by drawing new individuals from gene pools of each population with replacement to reform the populations. This randomization assumes that each population is in Hardy-Weinber Equilibrium but the populations are distinct.

Two alternative genetic distances that treat the data in radically different ways were used. Genotype likelihood ratio distance, D_{LR}, as an independent measure to confirm the famous D_S. Nei's (1972) standard genetic distance Ds, standard errors of standard genetic distances (Nei 1978) were calculated among populations in order to create an input distance matrix by using DISPAN software (Ota 1993).

Nei's standard distance is calculated as;

$$Ds = -\ln\left[\frac{J_{XY}}{J_X J_Y}\right]$$

where

$$J_X = \frac{\sum_j^r \sum_i^{m_j} x_{ij}^2}{r}$$

and

$$J_Y = \frac{\sum_j^r \sum_i^{m_j} y_{ij}^2}{r}$$

are the average homozygosities over loci in populations **X** and **Y**, respectively, and

$$J_{XY} = \frac{\sum_j^r \sum_i^{m_j} x_{ij} y_{ij}}{r}$$

$\mathbf{X_{ij}}$ and $\mathbf{Y_{ij}}$ are the frequencies of the ***i***th allele at the ***j***th locus in populations **X** and **Y**, respectively, $\mathbf{m_j}$ is the number of alleles at the ***j***th locus, and **r** is the number of loci examined.

D_{LR} distance matrix was formed using Doh asignment test calculator. D_{LR} was described in Paetkau *et al*.' Study (1997). This distance measure is based on assignment test (Paetkau *et al*. 1995) that was also used in this thesis. It is defined as

$$D_{LR} = \left(\frac{1}{n_X} \sum_{i}^{n_X} \log \frac{L_{iXX}}{L_{iXY}} + \frac{1}{n_Y} \sum_{i}^{n_Y} \log \frac{L_{iYY}}{L_{iYX}} \right) \div 2.$$

X and Y refers to populations with n_X and n_Y individuals in the formula. L_{iXX} and L_{iXY} are the likelihoods of individual i from X population in population X and in population Y respectively. When D_{LR} is 3 this means that the likelihood of this genotype in its own population is three times higher than its likelihood in the other population.

Population trees based on Ds were constructed using the neighbour-joining (NJ) method of Saitou and Nei (1987) using DISPAN software (Ota 1993). Population trees based on D_{LR} were drawn by using D_{LR} matrix into PHYLIP (v3.6) software (Felsenstein 1988) as an input.

CHAPTER 3

RESULTS

3.1. DNA Isolation and Genotyping

Availibility of DNA and amount of protein contamination were checked by spectroscopy at wavelengths 260 and 280 nm. We used DNA solutions that had 1,75-2,00 absorbance ratios (A_{260}/A_{280}). DNA concentrations changed between 0,1 and 0,9 µg/µl after isolation. Genomic DNA existence were further controlled by 1% agarose gel electrophoresis and positive DNA solutions were used for polymerase chain reaction (PCR) amplification (Figure 1).

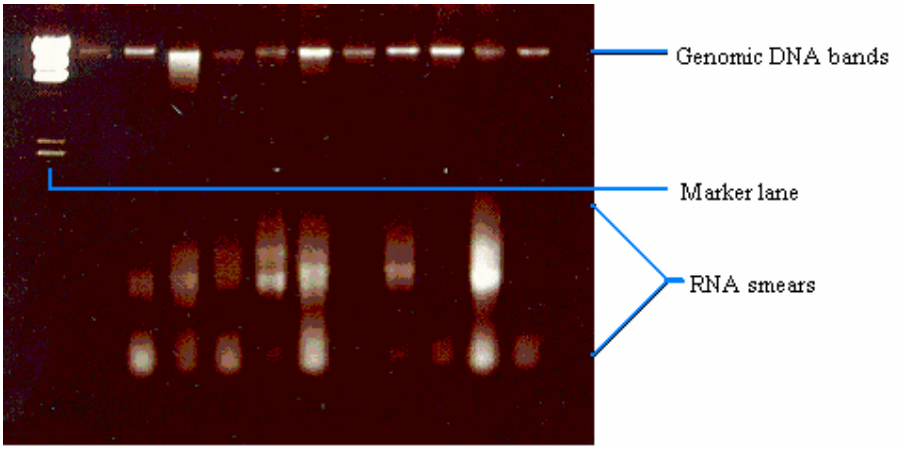

Figure 1. Agarose gel electrophoresis of genomic DNAs isolated from various honeybee samples.

Genotyping of worker honeybees were performed on autoradiograms representing microsatellite alleles as bands (Figures 2-10).

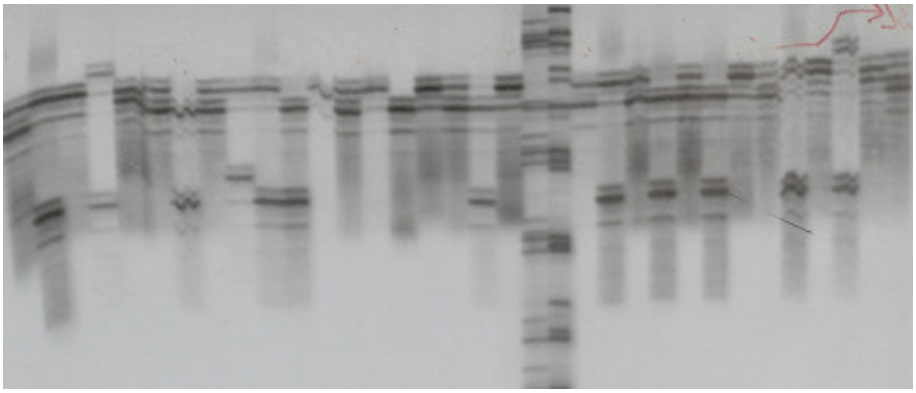

Figure 2. Bands that refer to alleles of A24 locus on auturadiography film.

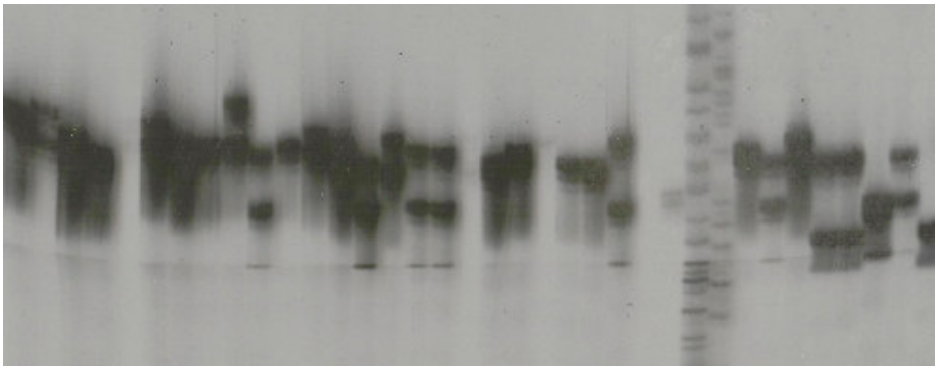

Figure 3. Bands that refer to alleles of A113 locus on auturadiography film.

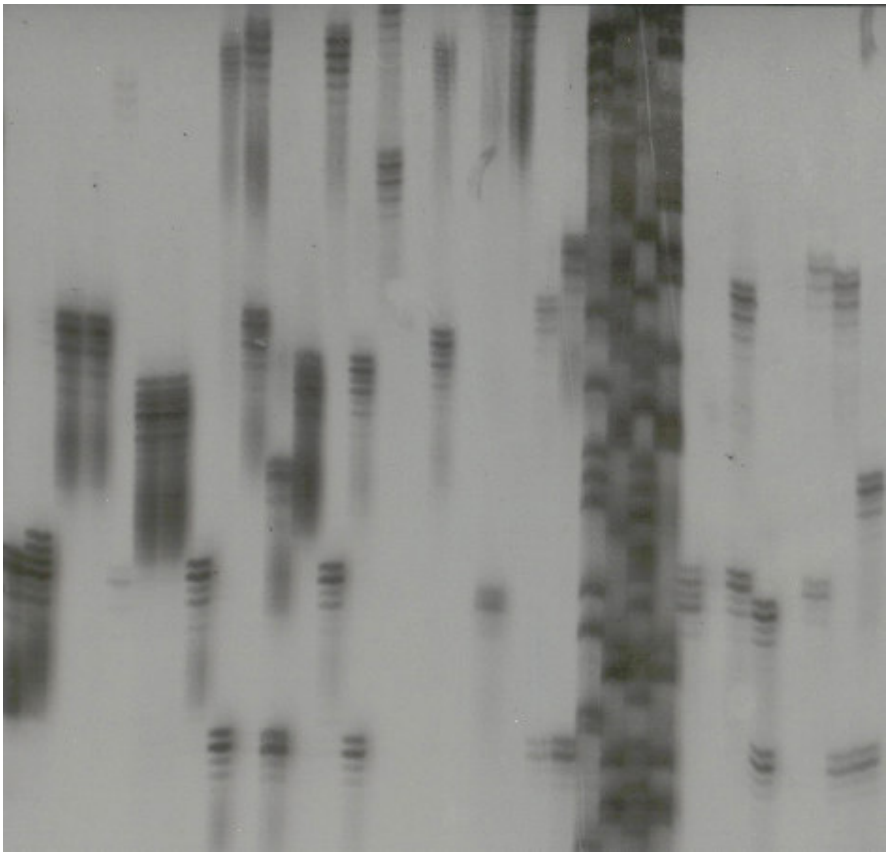

Figure 4. Bands that refer to alleles of A7 locus on auturadiography film.

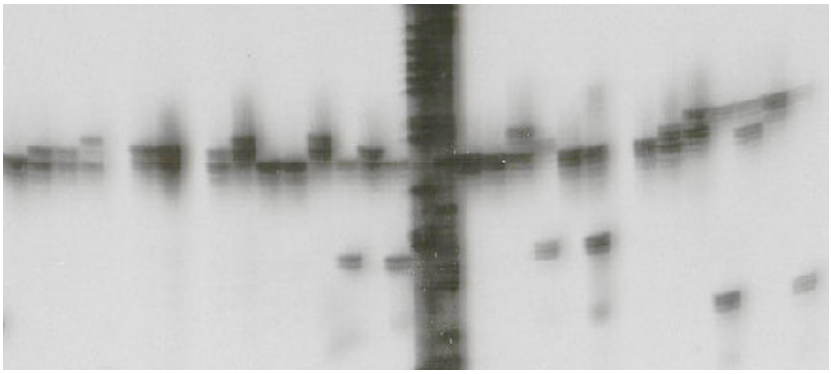

Figure 5. Bands that refer to alleles of A43 locus on auturadiography film.

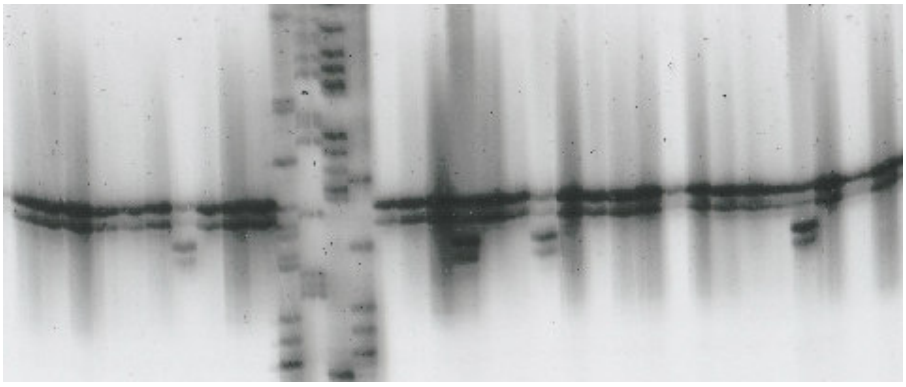

Figure 6. Bands that refer to alleles of A28 locus on auturadiography film.

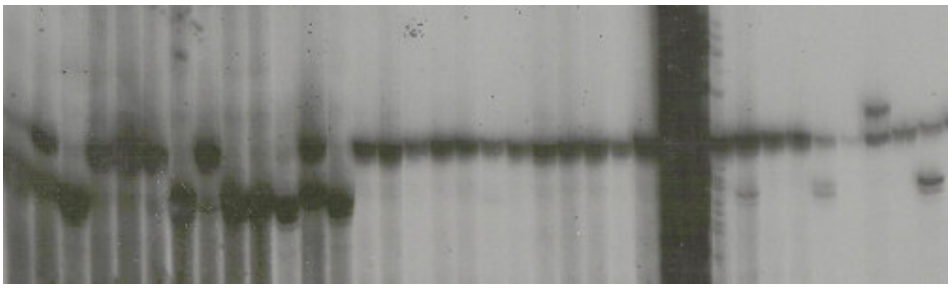

Figure 7. Bands that refer to alleles of Ap226 locus on auturadiography film.

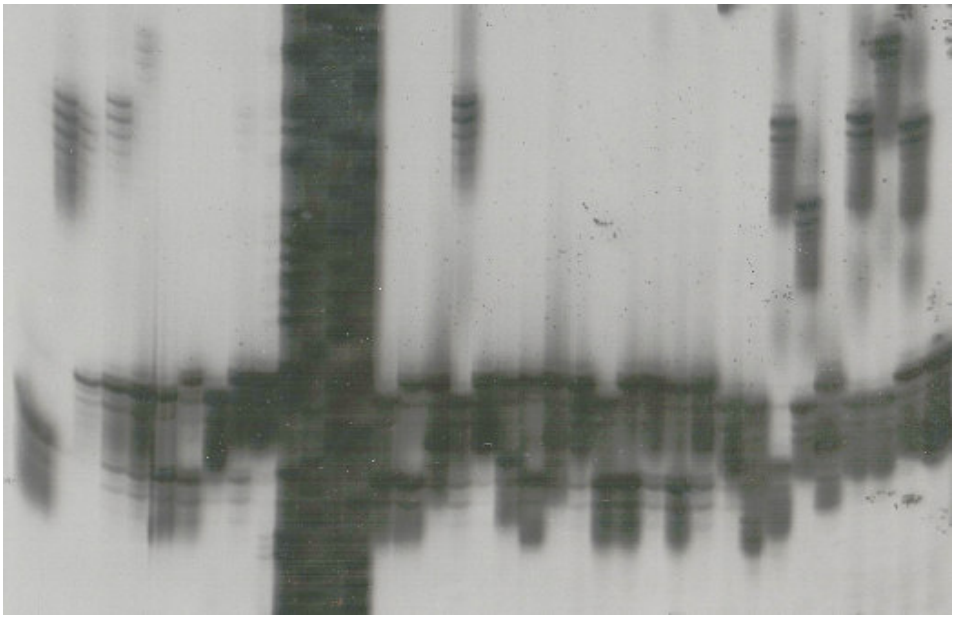

Figure 8. Bands that refer to alleles of Ap43 locus on auturadiography film.

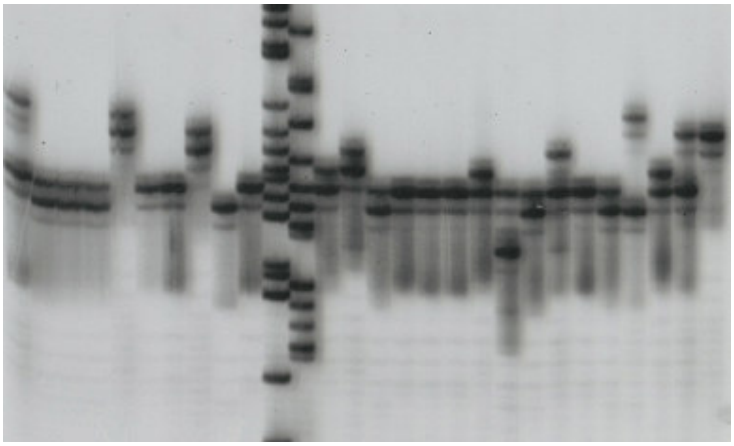

Figure 9. Bands that refer to alleles of Ap68 locus on auturadiography film.

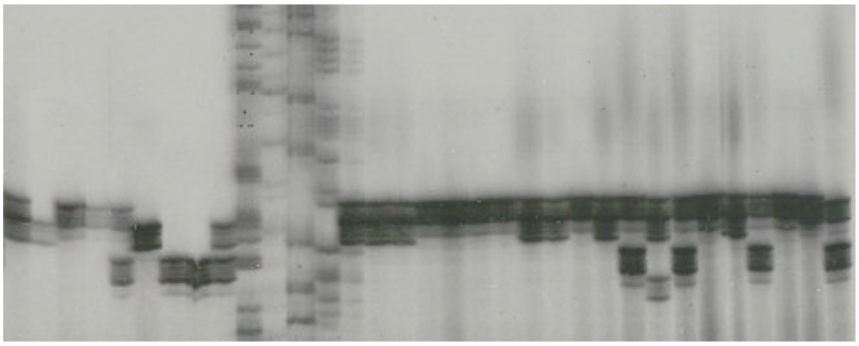

Figure 10. Bands that refer to alleles of Ac306 locus on auturadiography film.

3.2. Genetic variation

Genetic variation among eleven honeybee populations were analysed by determination of number of alleles, allele frequencies and heterozygosity measures among populations.

3.2.1. Allele polymorphism

A total of 167 alleles were found for nine microsatellite loci in 349 worker bees from 12 honeybee populations. All microsatellite loci were found to be polymorphic whose number of alleles changed between 6 (A24) and 68 (A7). Mean number of alleles per locus were between 5,67 (İzmir) and 8,33 (Hakkari) as could be seen in Table 1.

Table 1. Number of alleles and average number of alleles per locus values (Av: Average number of alleles per locus).

	A24	A113	A7	A43	A28	Ap226	Ap43	Ap68	Ac306	Av
ESK	4	10	24	3	3	2	10	6	4	7,33
ART	5	9	16	4	5	1	15	2	5	6,89
HAK	5	10	23	6	3	3	13	7	5	8,33
HAT	4	8	18	5	3	4	13	6	5	7,33
KIR	4	9	14	4	2	4	9	8	4	6,44
CYP	4	10	19	4	2	5	8	4	4	6,67
ARD	3	8	18	4	5	1	10	3	4	6,22
İZM	3	6	17	3	2	1	10	5	4	5,67
KAS	4	9	25	3	2	2	14	5	4	7,56
MUĞ	5	9	20	3	2	1	12	6	4	6,89
URF	5	10	20	6	1	3	13	7	3	7,56
ANK	3	8	18	5	2	2	13	5	3	6,56
Total	6	16	68	14	8	8	30	9	8	6,95

Alleles that have been observed in only one population are called private alleles. Private alleles and their frequencies could be seen in Tables 2a and 2b. Mean frequency of private alleles for the total population was found to be 0,036 (Table3). The highest average private allele frequency was seen in population from Cyprus with a 0,088 value and the lowest average private allele frequencies were detected in Urfa (0,021) and Ankara (0,000) populations (Table 3).

Alleles that are in relatively high proportions in one population and either absent or in very low frequencies in all other populations are named diagnostic alleles in this thesis. So called diagnostic alleles are given in Table 4.

The abbreviations for population names in all the tables of this chapter are: ESK: Eskişehir, ART: Artvin, HAK: Hakkari, Hat: Hatay, KIR: Kırklareli, CYP: Cyprus, ARD: Ardahan, İZM: İzmir, KAS: Kastamonu, MUĞ: Muğla, URF: Urfa, ANK: Ankara, ANA: Anatolia.

Table 2a. Private alleles and their frequencies in relevant populations.

POPULATION	LOCUS	ALLELE	FREQUENCY
Muğla	A113	248	0,020
Ardahan	A7	110	0,024
Cyprus	A7	111	0,080
Kastamonu	A7	114	0,017
Cyprus	A7	116	0,040
Cyprus	A7	118	0,020
Ardahan	A7	120	0,024
Cyprus	A7	126	0,060
Ardahan	A7	128	0,071
Cyprus	A7	130	0,160
Ardahan	A7	148	0,024
Ardahan	A7	152	0,048
İzmir	A7	156	0,026
Kastamonu	A7	158	0,086
Eskişehir	A7	160	0,037
Hatay	A7	161	0,028
Ardahan	A7	162	0,024
Muğla	A7	163	0,023
Hatay	A7	169	0,028
İzmir	A7	170	0,053
İzmir	A7	172	0,026
Artvin	A7	173	0,026
İzmir	A7	175	0,026
Artvin	A7	177	0,053
Kastamonu	A7	200	0,017
Cyprus	A43	117	0,130
Hatay	A43	119	0,175
Hakkari	A43	127	0,037
Kırklareli	A43	128	0,017
Hakkari	A43	139	0,019
Hatay	A43	141	0,025
Urfa	A43	143	0,017
Urfa	A43	148	0,034
Artvin	A28	125	0,023
Eskişehir	A28	127	0,033
Ardahan	A28	131	0,024
Artvin	A28	140	0,023
Ardahan	A28	144	0,024
Kırklareli	Ap226	241	0,050
Cyprus	Ap226	253	0,180

Table 2b. Private alleles and their frequencies in relevant populations.

POPULATION	LOCUS	ALLELE	FREQUENCY
Hatay	Ap226	255	0,032
Urfa	Ap43	131	0,017
Cyprus	Ap43	133	0,031
Hakkari	Ap43	149	0,017
Hatay	Ap43	157	0,017
Ardahan	Ap43	168	0,036
Kastamonu	Ap43	189	0,067
Muğla	Ap43	191	0,022
Urfa	Ap43	193	0,017
Ardahan	Ap43	201	0,036
Kırklareli	Ap68	151	0,017
İzmir	Ac306	173	0,024
Artvin	Ac306	178	0,125

Table 3. Number of private alleles (N_P) and their average frequencies.

Population	Number of private alleles	Average Frequency
Ardahan	10	0,034
Cyprus	8	0,088
Hatay	6	0,051
İzmir	5	0,031
Artvin	5	0,050
Kastamonu	4	0,047
Urfa	4	0,021
Kırklareli	3	0,028
Muğla	3	0,022
Hakkari	3	0,024
Eskişehir	2	0,035
Ankara	0	0,000
Total	53	0,036

Table 4. Diagnostic alleles for 12 populations and Anatolia.

ESK	127 (A28)
ART	177 (A7),181,185 (Ap43),178 (Ac306)
HAK	127 (A43)
HAT	123 (A7),119 (A43),155 (Ap226),147 (Ap43)
KIR	214 (A113),115 (A7),237, 239 (Ap226),167 (Ac306)
CYP	111,116,126,130,134 (A7),117 (A43),253 (Ap226),133 (Ap43)
ARD	220 (A113),168,201 (Ap43),138 (A43),128,138,152 (A7)
İZM	104,170 (A7),139 (Ap43)
KAS	158 (A7),251 (Ap226),189 (Ap43)
MUĞ	141 (A7)
URF	133 (A7),148 (A43)
ANK	-
ANA	96 (A24),210, 232 (A113),144 (A43),179 (Ac306)

3.2.2. Allele frequencies and heterozygosity values

Allelic frequencies of 9 microsatellite loci in each of 12 honeybee populations could be seen in Tables 5-13 together with number of alleles sampled from populations. Observed and expected heterozygosities for each locus were calculated for 12 populations and tabulated in Tables 14a and 14b.

Table 5. Frequencies of A24 alleles (N: number of alleles).

	96	98	102	104	106	108	N
ESK	0,107	0,000	0,054	0,375	0,464	0,000	56
ART	0,097	0,000	0,081	0,435	0,355	0,032	62
HAK	0,232	0,018	0,018	0,321	0,411	0,000	56
HAT	0,188	0,062	0,000	0,229	0,521	0,000	48
KIR	0,000	0,000	0,089	0,536	0,357	0,018	56
CYP	0,000	0,111	0,000	0,278	0,593	0,019	54
ARD	0,139	0,000	0,000	0,444	0,417	0,000	38
İZM	0,000	0,000	0,042	0,438	0,521	0,000	48
KAS	0,080	0,000	0,000	0,420	0,420	0,080	50
MUĞ	0,048	0,000	0,065	0,306	0,532	0,048	62
URF	0,367	0,033	0,000	0,167	0,417	0,017	60
ANK	0,188	0,000	0,000	0,208	0,604	0,000	48

Table 6a. Frequencies of A113 alleles (N: number of alleles).

	210	212	214	216	218	220	222	224	N
ESK	0,017	0,000	0,000	0,000	0,000	0,069	0,069	0,190	58
ART	0,000	0,000	0,000	0,016	0,065	0,000	0,258	0,016	62
HAK	0,000	0,000	0,017	0,034	0,000	0,034	0,103	0,310	58
HAT	0,023	0,205	0,023	0,068	0,000	0,000	0,114	0,523	44
KIR	0,000	0,000	0,438	0,125	0,000	0,000	0,021	0,042	48
CYP	0,083	0,000	0,021	0,146	0,042	0,000	0,062	0,312	48
ARD	0,000	0,000	0,000	0,000	0,000	0,192	0,077	0,077	26
İZM	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,114	44
KAS	0,125	0,000	0,000	0,000	0,042	0,000	0,042	0,000	24
MUĞ	0,040	0,000	0,000	0,000	0,000	0,000	0,100	0,100	50
URF	0,150	0,117	0,033	0,033	0,000	0,000	0,000	0,117	60
ANK	0,125	0,000	0,000	0,000	0,062	0,000	0,062	0,250	

Table 6b. Frequencies of A113 alleles (N: number of alleles).

	226	228	230	232	234	236	238	248	N
ESK	0,224	0,086	0,172	0,069	0,069	0,034	0,000	0,000	58
ART	0,097	0,210	0,242	0,081	0,000	0,016	0,000	0,000	62
HAK	0,155	0,155	0,138	0,034	0,000	0,017	0,000	0,000	58
HAT	0,023	0,000	0,000	0,000	0,000	0,000	0,023	0,000	44
KIR	0,167	0,062	0,104	0,000	0,021	0,000	0,021	0,000	48
CYP	0,104	0,188	0,021	0,000	0,000	0,021	0,000	0,000	48
ARD	0,154	0,269	0,115	0,077	0,000	0,038	0,000	0,000	26
İZM	0,250	0,341	0,136	0,091	0,000	0,068	0,000	0,000	44
KAS	0,042	0,125	0,250	0,208	0,083	0,083	0,000	0,000	24
MUĞ	0,160	0,180	0,180	0,160	0,000	0,060	0,000	0,020	50
URF	0,300	0,133	0,050	0,017	0,000	0,050	0,000	0,000	60
ANK	0,250	0,031	0,062	0,156	0,000	0,000	0,000	0,000	32

Table 7a. Frequencies of A7 alleles (N: number of alleles).

	99	102	104	105	108	110	111	112	114	115	N
ESK	0,000	0,019	0,000	0,037	0,000	0,000	0,000	0,000	0,000	0,074	54
ART	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	38
HAK	0,000	0,000	0,019	0,000	0,019	0,000	0,000	0,000	0,000	0,037	54
HAT	0,028	0,000	0,000	0,028	0,028	0,000	0,000	0,000	0,000	0,000	36
KIR	0,000	0,025	0,000	0,050	0,000	0,000	0,000	0,000	0,000	0,550	40
CYP	0,000	0,000	0,000	0,000	0,000	0,000	0,080	0,020	0,000	0,000	50
ARD	0,000	0,000	0,024	0,000	0,000	0,024	0,000	0,000	0,000	0,000	42
İZM	0,000	0,000	0,237	0,000	0,000	0,000	0,000	0,000	0,000	0,105	38
KAS	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,034	0,017	0,000	58
MUĞ	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,068	44
URF	0,018	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	56
ANK	0,000	0,028	0,000	0,028	0,000	0,000	0,000	0,000	0,000	0,000	36

Table 7b. Frequencies of A7 alleles (N: number of alleles).

	116	117	118	119	120	121	123	124	125	126	N
ESK	0,000	0,019	0,000	0,093	0,000	0,000	0,000	0,000	0,019	0,000	54
ART	0,000	0,184	0,000	0,000	0,000	0,000	0,000	0,000	0,026	0,000	38
HAK	0,000	0,148	0,000	0,019	0,000	0,056	0,019	0,000	0,037	0,000	54
HAT	0,000	0,000	0,000	0,056	0,000	0,056	0,139	0,000	0,028	0,000	36
KIR	0,000	0,000	0,000	0,075	0,000	0,050	0,025	0,000	0,000	0,000	40
CYP	0,040	0,000	0,020	0,000	0,000	0,020	0,000	0,040	0,020	0,060	50
ARD	0,000	0,214	0,000	0,071	0,024	0,000	0,000	0,000	0,000	0,000	42
İZM	0,000	0,105	0,000	0,000	0,000	0,000	0,000	0,053	0,000	0,000	38
KAS	0,000	0,000	0,000	0,000	0,000	0,086	0,017	0,000	0,034	0,000	58
MUĞ	0,000	0,000	0,000	0,091	0,000	0,000	0,023	0,000	0,023	0,000	44
URF	0,000	0,018	0,000	0,054	0,000	0,107	0,018	0,000	0,071	0,000	56
ANK	0,000	0,028	0,000	0,000	0,000	0,083	0,028	0,000	0,139	0,000	36

Table 7c. Frequencies of A7 alleles (N: number of alleles).

	127	128	129	130	131	132	133	134	135	136	N
ESK	0,093	0,000	0,037	0,000	0,056	0,000	0,000	0,000	0,130	0,000	54
ART	0,184	0,000	0,079	0,000	0,000	0,000	0,000	0,000	0,026	0,000	38
HAK	0,037	0,000	0,111	0,000	0,074	0,000	0,074	0,000	0,037	0,000	54
HAT	0,139	0,000	0,000	0,000	0,056	0,000	0,028	0,000	0,111	0,000	36
KIR	0,050	0,000	0,025	0,000	0,000	0,000	0,025	0,000	0,025	0,000	40
CYP	0,020	0,000	0,020	0,160	0,000	0,040	0,000	0,220	0,020	0,100	50
ARD	0,000	0,071	0,000	0,000	0,048	0,000	0,000	0,071	0,000	0,048	42
İZM	0,000	0,000	0,000	0,000	0,000	0,026	0,000	0,053	0,000	0,000	38
KAS	0,017	0,000	0,000	0,000	0,017	0,052	0,017	0,000	0,069	0,000	58
MUĞ	0,068	0,000	0,068	0,000	0,045	0,000	0,023	0,000	0,045	0,000	44
URF	0,107	0,000	0,036	0,000	0,107	0,000	0,143	0,000	0,054	0,000	56
ANK	0,056	0,000	0,028	0,000	0,000	0,000	0,111	0,000	0,167	0,000	36

Table 7d. Frequencies of A7 alleles (N: number of alleles).

	137	138	139	140	141	142	143	144	145	146	N
ESK	0,019	0,019	0,037	0,074	0,019	0,000	0,000	0,000	0,019	0,000	54
ART	0,079	0,000	0,026	0,000	0,053	0,000	0,000	0,000	0,079	0,026	38
HAK	0,074	0,000	0,037	0,019	0,037	0,000	0,019	0,000	0,037	0,000	54
HAT	0,056	0,000	0,111	0,000	0,028	0,000	0,000	0,000	0,000	0,000	36
KIR	0,000	0,000	0,000	0,000	0,000	0,025	0,000	0,025	0,000	0,000	40
CYP	0,000	0,000	0,000	0,000	0,000	0,020	0,000	0,000	0,000	0,000	50
ARD	0,000	0,119	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,024	42
İZM	0,000	0,026	0,000	0,053	0,000	0,000	0,026	0,026	0,000	0,053	38
KAS	0,000	0,000	0,052	0,017	0,000	0,052	0,000	0,000	0,034	0,052	58
MUĞ	0,023	0,000	0,114	0,023	0,159	0,000	0,045	0,000	0,023	0,000	44
URF	0,018	0,000	0,018	0,000	0,000	0,000	0,036	0,000	0,054	0,000	56
ANK	0,111	0,000	0,028	0,000	0,000	0,000	0,028	0,000	0,028	0,000	36

Table 7e. Frequencies of A7 alleles (N: number of alleles)

	147	148	149	150	151	152	153	154	155	156	N
ESK	0,037	0,000	0,037	0,019	0,000	0,000	0,019	0,000	0,000	0,000	54
ART	0,079	0,000	0,026	0,026	0,000	0,000	0,000	0,000	0,000	0,000	38
HAK	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,037	0,000	54
HAT	0,000	0,000	0,028	0,000	0,028	0,000	0,000	0,000	0,000	0,000	36
KIR	0,025	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	40
CYP	0,020	0,000	0,060	0,000	0,000	0,000	0,000	0,000	0,000	0,000	50
ARD	0,000	0,024	0,024	0,000	0,000	0,048	0,000	0,048	0,000	0,000	42
İZM	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,026	38
KAS	0,017	0,000	0,086	0,034	0,000	0,000	0,000	0,052	0,000	0,000	58
MUĞ	0,000	0,000	0,023	0,000	0,023	0,000	0,068	0,000	0,000	0,000	44
URF	0,018	0,000	0,000	0,000	0,000	0,000	0,036	0,000	0,036	0,000	56
ANK	0,000	0,000	0,000	0,000	0,028	0,000	0,000	0,000	0,000	0,000	36

Table 7f. Frequencies of A7 alleles (N: number of alleles).

	157	158	159	160	161	162	163	165	167	169	N
ESK	0,056	0,000	0,019	0,037	0,000	0,000	0,000	0,000	0,000	0,000	54
ART	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	38
HAK	0,019	0,000	0,019	0,000	0,000	0,000	0,000	0,000	0,000	0,000	54
HAT	0,000	0,000	0,000	0,000	0,028	0,000	0,000	0,000	0,000	0,028	36
KIR	0,025	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	40
CYP	0,000	0,000	0,020	0,000	0,000	0,000	0,000	0,000	0,000	0,000	50
ARD	0,000	0,000	0,000	0,000	0,000	0,024	0,000	0,048	0,048	0,000	42
İZM	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	38
KAS	0,034	0,086	0,017	0,000	0,000	0,000	0,000	0,017	0,069	0,000	58
MUĞ	0,023	0,000	0,000	0,000	0,000	0,000	0,023	0,000	0,000	0,000	44
URF	0,036	0,000	0,018	0,000	0,000	0,000	0,000	0,000	0,000	0,000	56
ANK	0,000	0,000	0,000	0,000	0,028	0,000	0,000	0,028	0,028	0,000	36

Table 7g. Frequencies of A7 alleles (N: number of alleles).

	170	171	172	173	175	177	179	200	N
ESK	0,000	0,019	0,000	0,000	0,000	0,000	0,000	0,000	54
ART	0,000	0,000	0,000	0,026	0,000	0,053	0,026	0,000	38
HAK	0,000	0,000	0,000	0,000	0,000	0,000	0,019	0,000	54
HAT	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	36
KIR	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	40
CYP	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	50
ARD	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	42
İZM	0,053	0,026	0,026	0,000	0,026	0,000	0,079	0,000	38
KAS	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,017	58
MUĞ	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	44
URF	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	56
ANK	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	36

Table 8a. Frequencies of A43 alleles (N: number of alleles).

	117	119	126	127	128	138	139	N
ESK	0,000	0,000	0,000	0,000	0,000	0,000	0,000	58
ART	0,000	0,000	0,016	0,000	0,000	0,000	0,000	62
HAK	0,000	0,000	0,000	0,037	0,000	0,000	0,019	54
HAT	0,000	0,175	0,000	0,000	0,000	0,000	0,000	40
KIR	0,000	0,000	0,052	0,000	0,017	0,000	0,000	58
CYP	0,130	0,000	0,000	0,000	0,000	0,000	0,000	54
ARD	0,000	0,000	0,000	0,000	0,000	0,100	0,000	40
İZM	0,000	0,000	0,000	0,000	0,000	0,021	0,000	48
KAS	0,000	0,000	0,000	0,000	0,000	0,000	0,000	24
MUĞ	0,000	0,000	0,000	0,000	0,000	0,000	0,000	60
URF	0,000	0,000	0,000	0,000	0,000	0,000	0,000	58
ANK	0,000	0,000	0,000	0,000	0,000	0,042	0,000	48

Table 8b. Frequencies of A43 alleles (N: number of alleles).

	140	141	142	143	144	146	148	N
ESK	0,879	0,000	0,103	0,000	0,017	0,000	0,000	58
ART	0,677	0,000	0,210	0,000	0,097	0,000	0,000	62
HAK	0,463	0,000	0,185	0,000	0,278	0,019	0,000	54
HAT	0,275	0,025	0,075	0,000	0,450	0,000	0,000	40
KIR	0,862	0,000	0,069	0,000	0,000	0,000	0,000	58
CYP	0,463	0,000	0,296	0,000	0,111	0,000	0,000	54
ARD	0,600	0,000	0,200	0,000	0,100	0,000	0,000	40
İZM	0,917	0,000	0,062	0,000	0,000	0,000	0,000	48
KAS	0,500	0,000	0,250	0,000	0,250	0,000	0,000	24
MUĞ	0,767	0,000	0,167	0,000	0,067	0,000	0,000	60
URF	0,241	0,000	0,086	0,017	0,534	0,086	0,034	58
ANK	0,562	0,000	0,271	0,000	0,104	0,021	0,000	48

Table 9. Frequencies of A28 alleles (N: number of alleles).

	125	127	129	131	133	138	140	144	N
ESK	0,000	0,033	0,000	0,000	0,100	0,867	0,000	0,000	30
ART	0,023	0,000	0,182	0,000	0,068	0,705	0,023	0,000	44
HAK	0,000	0,000	0,117	0,000	0,050	0,833	0,000	0,000	60
HAT	0,000	0,000	0,031	0,000	0,031	0,938	0,000	0,000	32
KIR	0,000	0,000	0,000	0,000	0,146	0,854	0,000	0,000	48
CYP	0,000	0,000	0,000	0,000	0,111	0,889	0,000	0,000	54
ARD	0,000	0,000	0,048	0,024	0,048	0,857	0,000	0,024	42
İZM	0,000	0,000	0,000	0,000	0,312	0,688	0,000	0,000	48
KAS	0,000	0,000	0,000	0,000	0,172	0,828	0,000	0,000	58
MUĞ	0,000	0,000	0,000	0,000	0,167	0,833	0,000	0,000	48
URF	0,000	0,000	0,000	0,000	0,000	1,000	0,000	0,000	60
ANK	0,000	0,000	0,000	0,000	0,105	0,895	0,000	0,000	38

Table10. Frequencies of Ap226 alleles (N: number of alleles).

	235	237	239	241	249	251	253	255	N
ESK	0,000	0,000	0,017	0,000	0,983	0,000	0,000	0,000	60
ART	0,000	0,000	0,000	0,000	1,000	0,000	0,000	0,000	50
HAK	0,000	0,016	0,047	0,000	0,938	0,000	0,000	0,000	64
HAT	0,000	0,016	0,065	0,000	0,887	0,000	0,000	0,032	62
KIR	0,000	0,217	0,400	0,050	0,333	0,000	0,000	0,000	60
CYP	0,020	0,020	0,000	0,000	0,740	0,040	0,180	0,000	50
ARD	0,000	0,000	0,000	0,000	1,000	0,000	0,000	0,000	36
İZM	0,000	0,000	0,000	0,000	1,000	0,000	0,000	0,000	48
KAS	0,000	0,000	0,000	0,000	0,768	0,232	0,000	0,000	56
MUĞ	0,000	0,000	0,000	0,000	1,000	0,000	0,000	0,000	64
URF	0,017	0,000	0,000	0,000	0,933	0,050	0,000	0,000	60
ANK	0,000	0,000	0,000	0,022	0,978	0,000	0,000	0,000	46

Table 11a. Frequencies of Ap43 alleles (N: number of alleles).

	131	133	135	137	139	143	145	147	149	157	N
ESK	0,000	0,000	0,233	0,067	0,000	0,417	0,017	0,017	0,000	0,000	60
ART	0,000	0,000	0,143	0,054	0,000	0,089	0,036	0,018	0,000	0,000	56
HAK	0,000	0,000	0,083	0,100	0,000	0,400	0,200	0,017	0,017	0,000	60
HAT	0,000	0,000	0,034	0,000	0,000	0,397	0,103	0,207	0,000	0,017	58
KIR	0,000	0,000	0,211	0,026	0,053	0,184	0,421	0,000	0,000	0,000	38
CYP	0,000	0,031	0,156	0,000	0,000	0,281	0,250	0,000	0,000	0,000	32
ARD	0,000	0,000	0,143	0,000	0,000	0,393	0,000	0,036	0,000	0,000	28
İZM	0,000	0,000	0,071	0,190	0,262	0,310	0,048	0,000	0,000	0,000	42
KAS	0,000	0,000	0,150	0,033	0,017	0,267	0,183	0,000	0,000	0,000	60
MUĞ	0,000	0,000	0,174	0,152	0,043	0,283	0,043	0,000	0,000	0,000	46
URF	0,017	0,000	0,050	0,050	0,000	0,400	0,150	0,017	0,000	0,000	60
ANK	0,000	0,000	0,194	0,000	0,028	0,194	0,056	0,000	0,000	0,000	36

Table 11b. Frequencies of Ap43 alleles (N: number of alleles).

	163	165	167	168	169	171	173	175	177	179	N
ESK	0,000	0,000	0,000	0,000	0,000	0,017	0,117	0,000	0,017	0,050	60
ART	0,018	0,000	0,000	0,000	0,036	0,018	0,000	0,161	0,054	0,036	56
HAK	0,000	0,000	0,017	0,000	0,017	0,050	0,017	0,033	0,033	0,017	60
HAT	0,000	0,069	0,034	0,000	0,000	0,017	0,000	0,017	0,034	0,000	58
KIR	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,026	0,026	0,000	38
CYP	0,000	0,000	0,000	0,000	0,000	0,062	0,094	0,094	0,031	0,000	32
ARD	0,000	0,000	0,000	0,036	0,071	0,071	0,000	0,143	0,036	0,036	28
İZM	0,000	0,000	0,000	0,000	0,024	0,024	0,024	0,024	0,000	0,000	42
KAS	0,017	0,050	0,000	0,000	0,000	0,033	0,083	0,033	0,000	0,000	60
MUĞ	0,022	0,000	0,000	0,000	0,022	0,022	0,043	0,130	0,000	0,043	46
URF	0,000	0,083	0,000	0,000	0,000	0,083	0,000	0,050	0,033	0,033	60
ANK	0,000	0,000	0,000	0,000	0,111	0,056	0,056	0,028	0,028	0,083	36

Table 11c. Frequencies of Ap43 alleles (N: number of alleles).

	181	183	185	187	189	191	193	195	199	201	N
ESK	0,000	0,000	0,000	0,050	0,000	0,000	0,000	0,000	0,000	0,000	60
ART	0,196	0,000	0,107	0,000	0,000	0,000	0,000	0,018	0,018	0,000	56
HAK	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	60
HAT	0,000	0,034	0,017	0,000	0,000	0,000	0,000	0,000	0,017	0,000	58
KIR	0,000	0,026	0,000	0,026	0,000	0,000	0,000	0,000	0,000	0,000	38
CYP	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	32
ARD	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,036	28
İZM	0,000	0,000	0,000	0,024	0,000	0,000	0,000	0,000	0,000	0,000	42
KAS	0,017	0,000	0,017	0,033	0,067	0,000	0,000	0,000	0,000	0,000	60
MUĞ	0,000	0,000	0,000	0,000	0,000	0,022	0,000	0,000	0,000	0,000	46
URF	0,000	0,000	0,000	0,000	0,000	0,000	0,017	0,017	0,000	0,000	60
ANK	0,111	0,028	0,000	0,028	0,000	0,000	0,000	0,000	0,000	0,000	36

Table 12. Frequencies of Ap68 alleles (N: number of alleles).

	149	151	153	155	157	159	161	163	167	N
ESK	0,000	0,000	0,083	0,783	0,033	0,050	0,033	0,000	0,017	60
ART	0,000	0,000	0,404	0,596	0,000	0,000	0,000	0,000	0,000	52
HAK	0,000	0,000	0,204	0,481	0,167	0,037	0,037	0,056	0,019	54
HAT	0,000	0,000	0,210	0,435	0,242	0,032	0,065	0,016	0,000	62
KIR	0,034	0,017	0,207	0,483	0,086	0,069	0,052	0,052	0,000	58
CYP	0,000	0,000	0,333	0,463	0,056	0,000	0,148	0,000	0,000	54
ARD	0,000	0,000	0,450	0,400	0,150	0,000	0,000	0,000	0,000	40
İZM	0,000	0,000	0,310	0,500	0,048	0,024	0,119	0,000	0,000	42
KAS	0,000	0,000	0,222	0,370	0,074	0,074	0,259	0,000	0,000	54
MUĞ	0,000	0,000	0,167	0,593	0,130	0,056	0,037	0,019	0,000	54
URF	0,017	0,000	0,183	0,567	0,100	0,017	0,083	0,033	0,000	60
ANK	0,000	0,000	0,175	0,450	0,100	0,050	0,225	0,000	0,000	40

Table 13. Frequencies of Ac306 alleles (N: number of alleles).

	167	171	173	175	177	178	179	181	N
ESK	0,050	0,250	0,000	0,000	0,483	0,000	0,217	0,000	60
ART	0,000	0,196	0,000	0,000	0,196	0,125	0,464	0,018	56
HAK	0,031	0,109	0,000	0,031	0,469	0,000	0,359	0,000	64
HAT	0,016	0,016	0,000	0,016	0,355	0,000	0,597	0,000	62
KIR	0,467	0,183	0,000	0,000	0,283	0,000	0,067	0,000	60
CYP	0,000	0,068	0,000	0,000	0,523	0,000	0,386	0,023	44
ARD	0,000	0,053	0,000	0,000	0,474	0,000	0,447	0,026	38
İZM	0,000	0,429	0,024	0,000	0,381	0,000	0,167	0,000	42
KAS	0,033	0,283	0,000	0,000	0,533	0,000	0,150	0,000	60
MUĞ	0,000	0,300	0,000	0,000	0,317	0,000	0,350	0,033	60
URF	0,000	0,093	0,000	0,000	0,352	0,000	0,556	0,000	54
ANK	0,000	0,350	0,000	0,000	0,400	0,000	0,250	0,000	20

Table 14a. Observed (H_O) and expected (H_E) heterozygosities for microsatellite loci.

	A24 H_O	A24 H_E	A113 H_O	A113 H_E	A7 H_O	A7 H_E	A43 H_O	A43 H_E	A28 H_O	A28 H_E
ESK	0,714	0,641	0,862	0,871	0,963	0,955	0,241	0,250	0,200	0,246
ART	0,871	0,678	0,968	0,823	0,842	0,920	0,484	0,517	0,591	0,507
HAK	0,857	0,697	0,793	0,837	0,926	0,950	0,667	0,702	0,333	0,322
HAT	0,833	0,681	0,818	0,681	0,833	0,944	0,700	0,737	0,125	0,181
KIR	0,750	0,588	0,583	0,768	0,650	0,723	0,276	0,253	0,208	0,290
CYP	0,667	0,570	0,833	0,842	0,880	0,912	0,778	0,681	0,222	0,234
ARD	1,000	0,627	1,000	0,868	0,905	0,929	0,800	0,595	0,286	0,307
İZM	0,750	0,543	0,955	0,810	0,947	0,920	0,167	0,197	0,542	0,467
KAS	0,880	0,647	0,750	0,884	0,931	0,962	0,500	0,652	0,276	0,319
MUĞ	0,645	0,624	0,960	0,876	0,955	0,945	0,300	0,417	0,333	0,284
URF	0,667	0,679	0,867	0,859	0,964	0,940	0,690	0,651	0,000	0,000
ANK	0,583	0,568	1,000	0,849	1,000	0,935	0,458	0,611	0,211	0,240

Table 14b. Observed (H_O) and expected (H_E) heterozygosities for microsatellite loci.

	Ap226 H_O	Ap226 H_E	Ap43 H_O	Ap43 H_E	Ap68 H_O	Ap68 H_E	Ac306 H_O	Ac306 H_E
ESK	0,033	0,066	0,667	0,762	0,300	0,409	0,667	0,681
ART	0,000	0,000	0,786	0,901	0,654	0,491	0,679	0,708
HAK	0,125	0,121	0,633	0,794	0,704	0,711	0,750	0,647
HAT	0,226	0,240	0,690	0,812	0,774	0,728	0,645	0,526
KIR	0,433	0,711	0,789	0,758	0,690	0,717	0,800	0,692
CYP	0,520	0,456	0,875	0,837	0,667	0,674	0,591	0,586
ARD	0,000	0,000	0,571	0,844	0,350	0,672	0,579	0,634
İZM	0,000	0,000	0,762	0,842	0,571	0,653	0,762	0,659
KAS	0,250	0,398	0,800	0,876	0,852	0,766	0,733	0,622
MUĞ	0,000	0,000	0,739	0,864	0,481	0,638	0,600	0,729
URF	0,067	0,160	0,833	0,819	0,600	0,637	0,593	0,602
ANK	0,043	0,086	0,778	0,905	0,650	0,751	0,700	0,689

Average expected heterozygosities (gene diversities) changed between 0,542 (Eskişehir) and 0,681 (Kastamonu). Mean gene diversity for all 12 populations was found to be 0,612 ± 0,036.

A grand mean of average observed heterozygosities was found as 0,609 for the total population consisting of 12 populations with a low standard deviation of 0,046. Among average observed heterozygosities the value for Eskişehir population (0,516) determined the lower part of the range and the value Cyprus population (0,670) was at the higher end (Table15).

Table 15. Average observed heterozygosities and their standart deviations.

	Average H_O	St.Dev.	Average H_E	St. Dev.
ESK	0,516	0,328	0,542	0,311
ART	0,653	0,287	0,616	0,284
HAK	0,643	0,257	0,642	0,260
HAT	0,627	0,266	0,614	0,255
KIR	0,575	0,221	0,611	0,199
CYP	0,670	0,210	0,644	0,213
ARD	0,610	0,349	0,608	0,294
İZM	0,606	0,330	0,566	0,305
KAS	0,664	0,258	0,681	0,219
MUĞ	0,557	0,315	0,597	0,312
URF	0,587	0,338	0,594	0,315
ANK	0,603	0,325	0,626	0,293
MEAN	0,609	0,046	0,612	0,036

3.3. Genetic structure

Hardy-Weinberg tests, linkage disequilibrium tests and population differentiation measures were calculated and represented in order to analyse genetic structure of 12 honeybee populations from Türkiye and Cyprus.

3.3.1. Hardy-Weinberg Tests

Deviations from Hardy Weinberg were detected at nine microsatellite loci for all populations and at 8 out of 108 population-locus combinations statistically significant deviations were detected at 0,05 level (Table 16). Three of these 8 deviations were detected at northeastern Türkiye population Ardahan. All deviations were in favor of homozygotes except at A24 locus in Ardahan population which showed an excess of heterozygotes.

Table 16. Significant deviations from HWE.

Population	Locus	P value
ARDAHAN	A24	0,001
ARDAHAN	Ap43	0,000
ARDAHAN	Ap68	0,002
KIRKLARELİ	A113	0,000
KIRKLARELİ	Ap226	0,001
ARTVİN	Ac306	0,001
HATAY	Ap43	0,000
MUĞLA	Ap68	0,008

3.3.2. Linkage disequilibrium

Linkage disequilibrium tests were performed for all pairs of loci at all populations in order to understand any linked inheritance among 9 microsatellite loci used. Out of 432 locus pair-population combinations 23 were found to show significant linkage disequilibriums (Table 18). P values for all loci pairs at total population indicated a disequilibrium between A24 and Ap43 and between A113 and Ap68 microsatellite loci (Table 17). But these total values were confirmed by significant disequilibriums at only 3 and 2 populations respectively.

Table 17. P values of linkage disequilibrium tests across all populations.

LOCUS PAIR	X^2	DF	P VALUE
A24 – A113	23,520	22	0,373
A24 – A7	17,183	20	0,641
A113 – A7	26,525	20	0,149
A24 – A43	27,447	24	0,284
A113 – A43	30,296	22	0,111
A7 - A43	25,064	20	0,199
A24 – A28	23,436	22	0,377
A113 – A28	18,941	20	0,526
A7 - A28	10,054	16	0,864
A43 – A28	18,632	22	0,668
A24 – Ap226	7,563	16	0,961
A113 - Ap226	14,609	14	0,405
A7 – Ap226	8,682	12	0,730
A43 – Ap226	8,899	16	0,917
A28 – Ap226	9,725	14	0,782
A24 – Ap43	44,525	22	0,003
A113 - Ap43	15,606	22	0,835
A7 – Ap43	18,287	18	0,437
A43 – Ap43	21,554	22	0,487
A28 – Ap43	8,009	20	0,992
Ap226 - Ap43	3,722	14	0,997
A24 – Ap68	27,037	24	0,303
A113 - Ap68	53,328	22	0,000
A7 – Ap68	15,472	20	0,749
A43 – Ap68	22,853	24	0,528
A28 – Ap68	13,434	22	0,920
Ap226 - Ap68	12,979	16	0,674
Ap43 - Ap68	24,123	22	0,341
A24 – Ac306	24,181	24	0,451
A113 - Ac306	29,055	22	0,143
A7 – Ac306	25,293	20	0,190
A43 – Ac306	18,977	24	0,753
A28 – Ac306	17,378	22	0,742
Ap226 - Ac306	14,252	14	0,431
Ap43 - Ac306	22,740	22	0,417
Ap68 - Ac306	24,205	24	0,450

Table 18. Significant linkage disequilibriums and their p values and standard errors.

LOCUS PAIR	POPULATION	P VALUE and ST. ERROR
A24-Ap43	ESKİŞEHİR	0,001 ± 0,001
A24-Ap43	CYPRUS	0,013 ± 0,001
A24-Ap43	KASTAMONU	0,005 ± 0,001
A113-Ap68	HAKKARİ	0,001 ± 0,000
A113-Ap68	URFA	0,041 ± 0,004
A113-A7	KIRKLARELİ	0,014 ± 0,001
A113-A7	CYPRUS	0,014 ± 0,002
Ap68-Ac306	HAKKARİ	0,023 ± 0,002
Ap68-Ac306	URFA	0,031 ± 0,001
A24-A43	ESKİŞEHİR	0,037 ±0,001
A113-Ac306	ESKİŞEHİR	0,027 ± 0,002
A7-Ap68	HAKKARİ	0,023 ± 0,003
A7-Ac306	HAKKARİ	0,006 ± 0,001
A113-A43	KIRKLARELİ	0,013 ± 0,001
A24-A7	CYPRUS	0,012 ± 0,001
A7-A43	CYPRUS	0,003 ± 0,001
A113-Ap226	CYPRUS	0,010 ± 0,001
A28-Ap226	CYPRUS	0,033 ± 0,000
A43-Ac306	CYPRUS	0,035 ± 0,001
A43-Ap43	ARDAHAN	0,032 ± 0,001
A24-A28	KASTAMONU	0,018 ± 0,001
A24-Ap68	MUĞLA	0,033 ± 0,002
Ap43-Ap68	URFA	0,005 ± 0,001

3.3.3. Population Differentiation

Genetic distinctness of populations were analysed by differentiation tests, calculating F coefficients, number of migrant (Nm) values, by performing assignment tests and by constructing phylogenetic trees based on genetic distances among populations.

3.3.3.1. Differentiation tests

Genic and genotypic differentiation tests both resulted in highly significant differentiation measures for both allelic and genotypic distribution among all populations. P values was 0,00 for all loci for both tests.

3.3.3.2. F coefficients

F_{ST}, F_{IS} and F_{IT} coefficients for the total population consisting of 12 subpopulations were given in Table 19. A significant F_{ST} measure of 0,077 is an indication of genetic differentiation is existing among 12 honeybee populations sampled. F_{IS} and F_{IT} measures that show deviations from Hardy-Weinberg equilibrium within subpopulations (within each of 12 populations) and within the total population, indicate a slight deficiency of heterozygotes within subpopulations and a higher defiency of heterozygotes within the total population respectively as revealed by positive F_{IS} and F_{IT} values.

Table 19. F coefficients of the total population consisting 12 subpopulations.

Locus	F_{ST}	F_{IT}	F_{IS}
A24	0,041	-0,163	-0,213
A113	0,074	0,051	-0,025
A7	0,053	0,090	0,039
A43	0,145	0,172	0,031
A28	0,046	0,034	-0,013
Ap226	0,238	0,428	0,249
Ap43	0,043	0,160	0,122
Ap68	0,034	0,097	0,065
Ac306	0,092	0,056	-0,039
All	0,074	0,083	0,010

A very high genetic structure is observed among honeybee populations of Türkiye and Cyprus as indicated by pairwise F_{ST} measures that revealed 52 population pairs are effectively differentiated out of 66 compared population pairs at the 0,05 significance level (Table 20 and 21). Among 12 honeybee populations from Kırklareli which is located at European region of Türkiye and İzmir-Karaburun which is at almost the west end of Anatolia showed a complete differentiation from all others according to F_{ST} measures.

Table 20. Pairwise F_{ST} values (ESK: Eskişehir, ART: Artvin, HAK: Hakkari, HAT: Hatay, KIR: Kırklareli, CYP: Cyprus, ARD: Ardahan, İZM: İzmir, KAS: Kastamonu, MUĞ: Muğla, URF: Urfa, ANK: Ankara).

	ESK	ART	HAK	HAT	KIR	CYP	ARD	İZM	KAS	MUĞ	URF	ANK
ESK	0,000											
ART	0,049	0,000										
HAK	0,042	0,043	0,000									
HAT	0,100	0,078	0,008	0,000								
KIR	0,120	0,124	0,139	0,200	0,000							
CYP	0,058	0,046	0,014	0,012	0,125	0,000						
ARD	0,029	0,000	0,003	0,019	0,135	0,027	0,000					
İZM	0,024	0,060	0,082	0,130	0,134	0,080	0,049	0,000				
KAS	0,020	0,000	0,000	0,042	0,076	0,035	0,010	0,033	0,000			
MUĞ	0,000	0,019	0,034	0,075	0,148	0,044	0,017	0,016	0,001	0,000		
URF	0,105	0,097	0,012	0,000	0,183	0,062	0,043	0,141	0,012	0,071	0,000	
ANK	0,000	0,006	0,000	0,018	0,120	0,000	0,006	0,029	0,000	0,000	0,013	0,000

Table 21. Significant pairwise F_{ST} values (ESK: Eskişehir, ART: Artvin, HAK: Hakkari, HAT: Hatay, KIR: Kırklareli, CYP: Cyprus, ARD: Ardahan, İZM: İzmir, KAS: Kastamonu, MUĞ: Muğla, URF: Urfa, ANK: Ankara).

	ESK	ART	HAK	HAT	KIR	CYP	ARD	İZM	KAS	MUĞ	URF	ANK
ESK	*	+	+	+	+	+	+	+	+		+	
ART	+	*	+	+	+	+		+		+	+	
HAK	+	+	*	+	+	+		+		+	+	
HAT	+	+	+	*	+	+	+	+	+	+		+
KIR	+	+	+	+	*	+	+	+	+	+	+	+
CYP	+	+	+	+	+	*	+	+		+	+	
ARD	+			+	+	+	*	+	+	+	+	
İZM	+	+	+	+	+	+	+	*	+	+	+	+
KAS	+			+	+		+	+	*		+	
MUĞ		+	+	+	+	+	+	+		*	+	
URF	+	+	+		+	+	+	+	+	+	*	+
ANK				+	+			+			+	*

The highest pairwise F_{ST} values are demonstarted as a bar graph that also contains pairwise number of migrant (Nm) values for comparison. Among these 15 population pairs a visible inconsistency for F_{ST} and Nm measures is detected between Kırklareli and Hakkari populations. This pair had a relatively high Nm value (2,587) against a high F_{ST} measure (0,139).

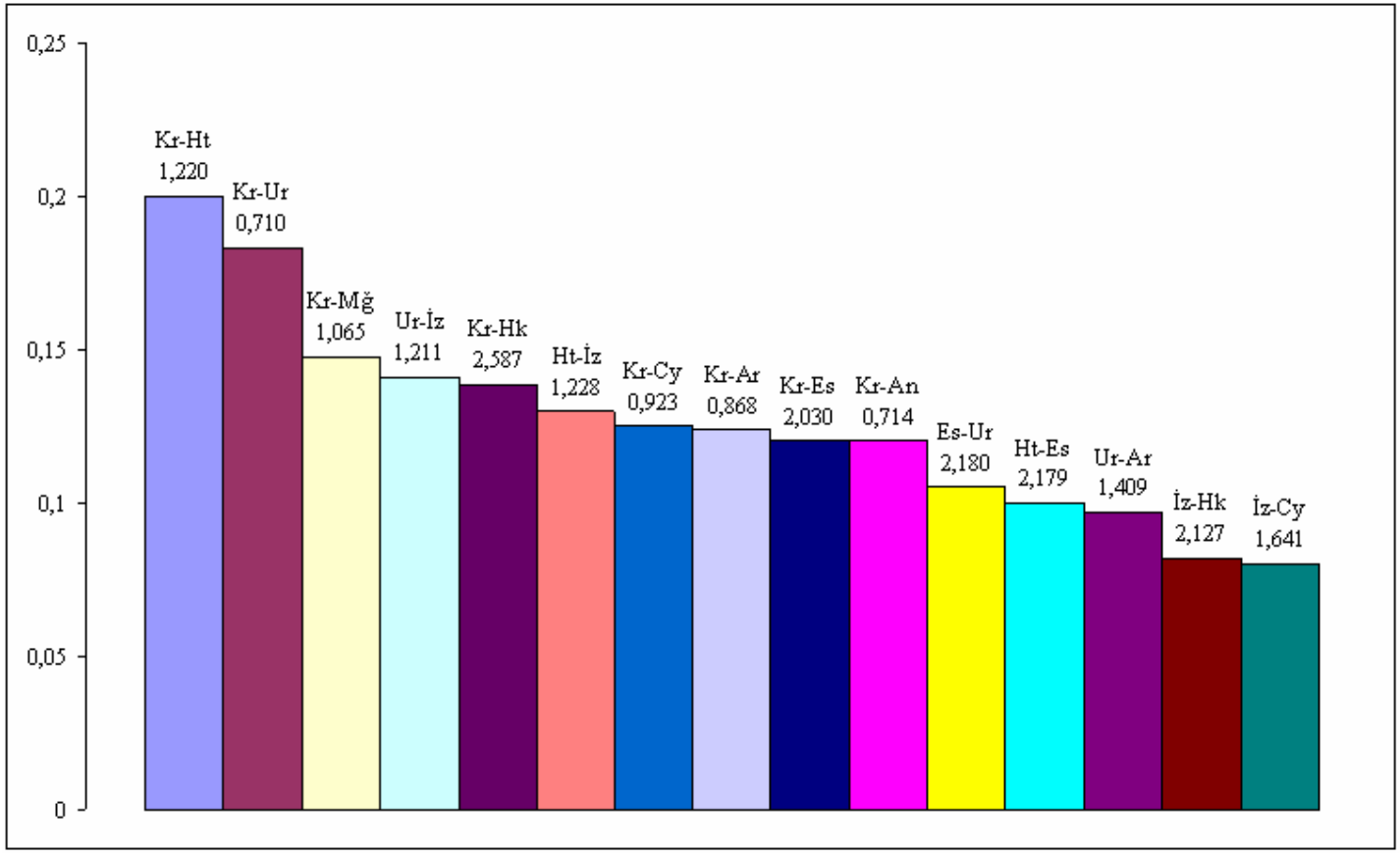

Figure 11. The highest pairwise F_{ST} values and their Nm statistics, Nm statistics are written above the data bars together with population pairs (Kr: Kırklareli, Ht: Hatay, Ur: Urfa, Mğ: Muğla, İz: İzmir, Hk: Hakkari, Cy: Cyprus, Ar: Artvin, Es: Eskişehir, An: Ankara).

3.3.3.3. Number of migrants

An overall number of migrants (Nm) value for all populations was found to be 2,716 Although total Nm value is higher than 2, pairwise Nm values (Table 22) shows that 35 out of 66 pairwise Nm values are higher than 2 which indicates an opportunity for divergence among 12 honeybee populations sampled throughout the Türkiye and Cyprus. And even 5 population pairwise Nm values of Kırklareli population with Urfa, Artvin, Kastamonu, Cyprus and Ankara, are below 1, at the region of so called high amount of differentiation. For Kırklareli population only two pairwise Nm values are higher than 2 level, namely its pairing with Eskişehir and interestingly Hakkari populations.

The highest pairwise Nm values are demonstrated on a bar graph (Figure 12) together with relevant F_{ST} measures of the same pair of populations in order to check the existance of congruence between high Nm values that means a decrease in genetic divergence and low F_{ST} measures that also means low divergence. Agreement between these two measures is interestingly disturbed by a high F_{ST} for Kırklareli-Hakkari population pair (0,139) against a high Nm for the same pair which seems contradictory.

Average pairwise Nm values for each population indicates that Kırklareli, Cyprus and Artvin populations had the lowest number of migrants (Table 23). Together with these three populations Izmir, Hatay and Kastamonu populations have also Nm values lower than two. The greatest Nm values were detected at Hakkari, Eskişehir, Ankara and Muğla populations.

Table 22. Pairwise Nm values and total Nm value for all populations (ESK: Eskişehir, ART: Artvin, HAK: Hakkari, HAT: Hatay, KIR: Kırklareli, CYP: Cyprus, ARD: Ardahan, İZM: İzmir, KAS: Kastamonu, MUĞ: Muğla, URF: Urfa, ANK: Ankara).

	ESK	ART	HAK	HAT	KIR	CYP	ARD	İZM	KAS	MUĞ	URF	ANK	TOTAL
ESK		1,888	4,137	2,179	2,030	1,786	2,879	2,262	2,406	4,039	2,180	2,845	
ART	1,888		2,842	1,463	0,868	1,263	1,845	1,451	1,600	1,776	1,409	2,395	
HAK	4,137	2,842		2,381	2,587	1,876	3,212	2,127	2,045	3,666	3,134	3,922	
HAT	2,179	1,463	2,381		1,220	1,557	1,857	1,228	1,377	1,684	3,384	2,375	
KIR	2,030	0,868	2,587	1,220		0,923	1,024	1,244	0,869	1,065	0,710	0,714	
CYP	1,786	1,263	1,876	1,557	0,923		1,704	1,641	1,459	1,555	1,389	1,800	
ARD	2,879	1,845	3,212	1,857	1,024	1,704		2,181	1,879	2,198	2,208	2,331	
İZM	2,262	1,451	2,127	1,228	1,244	1,641	2,181		1,853	2,282	1,211	1,664	
KAS	2,406	1,600	2,045	1,377	0,869	1,459	1,879	1,853		1,954	2,624	2,218	
MUĞ	4,039	1,776	3,666	1,684	1,065	1,555	2,198	2,282	1,954		2,329	2,575	
URF	2,180	1,409	3,134	3,384	0,710	1,389	2,208	1,211	2,624	2,329		3,031	
ANK	2,845	2,395	3,922	2,375	0,714	1,800	2,331	1,664	2,218	2,575	3,031		
TOTAL													2,716

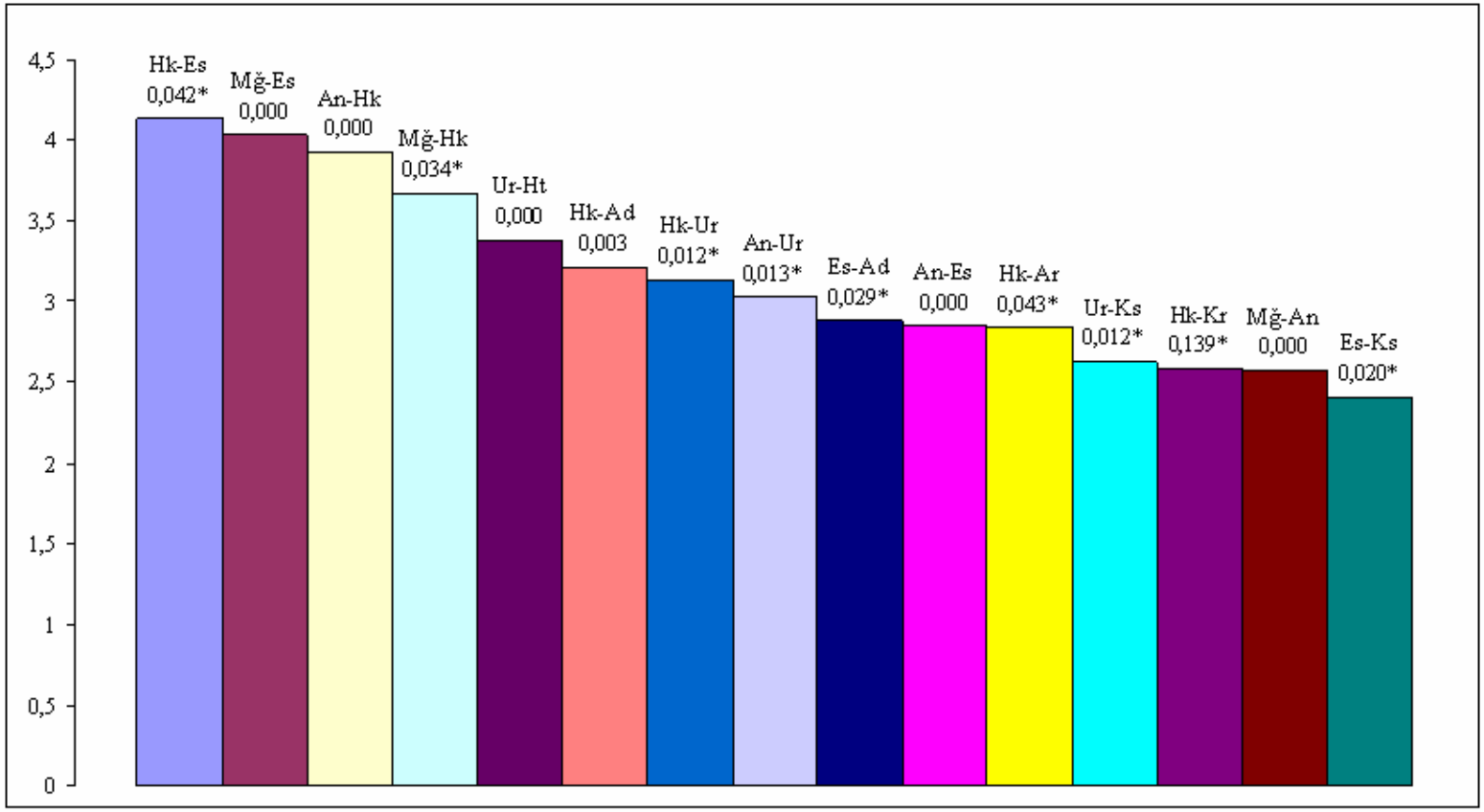

Figure 12. The highest pairwise Nm values and relevant F_{ST} statistics, F_{ST} statistics are written above the data bars together with population pairs (Kr: Kırklareli, Ht: Hatay, Ur: Urfa, Mğ: Muğla, Hk: Hakkari, Ar: Artvin, Es: Eskişehir, Ad: Ardahan, Ks: Kastamonu, An: Ankara). Significant F_{ST} values were indicated by a * sign.

Table 23. Averages and standard deviations of pairwise Nm values for each population.

ESK	ART	HAK	HAT	KIR	CYP	ARD	IZM	KAS	MUĞ	URF	ANK
2,579	1,641	2,801	1,833	1,254	1,515	2,099	1,748	1,807	2,255	2,058	2,352
0,850	0,519	0,739	0,668	0,594	0,277	0,611	0,448	0,510	0,927	0,866	0,819

3.3.3.4. Assignment tests

Assignment tests were performed in order to see likelihoods of belonging to different honeybee populations for each individual. Numbers of individuals assigned to the population in which its likelihood is the highest were given in Table 24. Percentages of correct assignments which means percentages of individuals that assigned to the population they were sampled, showed that genetic structures of Hakkari,Muğla and Ankara populations are very low (Table 25). These populations had only 25,34 and 35 % correct assignments for their individuals respectively. Individual honeybees from all of the remaining 9 populations showed correct assignment percentages higher than 55 %. Mean value of correct assignment were found to be 62 % with a big variance because of the 3 populations that had very low assignment percentages. Especially Kırklareli, Cyprus, Ardahan, Urfa, Artvin and Izmir populations had distinct genetic structures with correct assignment percentages higher than 70 %.

Table 24. Number of individuals assigned from population i (rows) to j (columns).

	ESK	ART	HAK	HAT	KIR	CYP	ARD	IZM	KAS	MUĞ	URF	ANK
ESK	17	0	0	1	0	0	0	2	2	4	0	4
ART	1	22	1	0	0	0	1	1	0	4	0	1
HAK	3	3	9	3	1	0	0	5	0	2	8	2
HAT	1	0	2	19	0	2	1	0	0	0	5	1
KIR	1	0	0	0	27	0	0	3	0	0	0	0
CYP	0	1	0	1	0	22	0	0	2	1	0	0
ARD	1	0	1	0	0	1	16	0	1	1	0	0
IZM	2	0	0	0	1	0	2	17	1	1	0	0
KAS	3	0	1	0	0	1	0	0	17	2	2	4
MUĞ	5	1	5	0	0	0	1	2	2	11	0	5
URF	0	0	4	3	0	0	0	0	0	1	22	0
ANK	5	2	1	0	0	0	2	2	1	2	2	9

Table 25. Percentages of correctly assigned individuals (n: Number of individuals assigned to population that they are sampled, N: Total number of individuals within each population, TOT: Total).

	ESK	ART	HAK	HAT	KIR	CYP	ARD	IZM	KAS	MUĞ	URF	ANK	TOT
n	17	22	9	19	27	22	16	17	17	11	22	9	215
N	30	31	36	31	31	27	21	24	30	32	30	26	346
%	57	71	25	61	87	81	76	71	57	34	73	35	62

Three different data randomizations were done in order to test three null hypotheses about the genetic structures of eleven honeybee populations. First data randomization were done by drawing existing individuals from combined gene pool of eleven populations with replacement to reform the populations. This randomization assumed that 12 populations are actually one well mixed population. Mean numbers and variances of assignments among populations and number of resamples with at least as many assignments from one population to another after randomization process if the assumed null hypothesis is true are given in Tables 26,27 and 28 respectively.

Second data randomization was done by drawing new individuals from combined gene pool of 12 populations with replacement to reform the populations. This randomization assumed that 12 populations are actually one well mixed population at Hardy-Weinberg Equilibrium. Mean numbers and variances of assignments among populations and number of resamples with at least as many assignments from one population to another after randomization process if the assumed null hypothesis is true are given in Tables 29,30 and 31 respectively.

The third data randomization was applied by drawing new individuals from gene pools of each population with replacement to reform the populations. This randomization assumed that each population is in Hardy-Weinber Equilibrium but the populations are distinct. Mean numbers and variances of assignments among populations and number of resamples with at least as many assignments from one population to another after randomization process if the assumed null hypothesis is true are given in Tables 32,33 and 34 respectively.

Table 26. Mean number of assignments from population i (rows) to j (columns) after randomization by drawing existing individuals from combined gene pool for all populations (1st randomization).

	ESK	ART	HAK	HAT	KIR	CYP	ARD	İZM	KAS	MUĞ	URF	ANK
ESK	2,3	2,5	2,2	2,5	2,3	2,6	2,7	2,6	2,4	2,3	2,4	2,6
ART	2,6	2,6	2,3	2,4	2,4	2,6	2,8	2,8	2,5	2,5	2,4	2,6
HAK	2,9	2,9	2,7	2,9	2,7	3,0	3,2	3,2	2,9	2,8	3,0	3,1
HAT	2,5	2,5	2,3	2,6	2,5	2,6	2,8	2,8	2,4	2,5	2,4	2,7
KIR	2,5	2,5	2,2	2,4	2,4	2,7	2,8	2,8	2,5	2,5	2,4	2,7
CYP	2,3	2,1	1,9	2,1	2,2	2,3	2,4	2,4	2,1	2,1	2,2	2,4
ARD	1,7	1,6	1,5	1,7	1,6	1,7	1,9	1,8	1,7	1,6	1,8	1,8
İZM	1,9	2,0	1,8	1,9	2,0	2,0	2,1	2,1	1,9	1,8	1,9	2,1
KAS	2,4	2,5	2,2	2,3	2,3	2,5	2,7	2,7	2,5	2,3	2,5	2,6
MUĞ	2,6	2,5	2,4	2,5	2,6	2,7	2,9	2,9	2,5	2,6	2,6	2,7
URF	2,4	2,3	2,3	2,4	2,3	2,5	2,7	2,6	2,5	2,4	2,5	2,5
ANK	2,1	2,1	1,9	2,0	2,1	2,2	2,4	2,3	2,1	2,1	2,1	2,2

Table 27. Variances of number of assignments from population i (rows) to j (columns) after randomization by drawing existing individuals from combined gene pool for all populations (1st randomization).

	ESK	ART	HAK	HAT	KIR	CYP	ARD	İZM	KAS	MUĞ	URF	ANK
ESK	3,7	2,5	2,3	2,5	2,2	2,5	2,8	2,6	2,4	2,5	2,6	2,6
ART	2,7	4,2	2,3	2,1	2,4	2,8	2,3	2,6	3,2	2,5	2,5	2,7
HAK	3,2	3,0	5,0	3,1	2,8	3,0	3,3	3,3	3,2	3,1	3,0	3,1
HAT	2,5	2,6	2,4	4,5	2,5	2,6	3,0	2,9	2,5	2,4	2,5	2,3
KIR	2,4	2,7	2,1	2,6	4,1	2,7	2,7	3,0	2,7	2,4	2,3	2,9
CYP	2,3	2,0	1,9	2,1	2,1	3,8	2,3	2,3	2,0	2,2	2,0	2,3
ARD	1,7	1,5	1,4	1,5	1,7	1,7	2,8	1,6	1,7	1,5	1,8	1,7
İZM	1,9	1,9	1,8	1,9	1,9	1,9	2,0	3,2	1,8	1,7	1,8	2,1
KAS	2,2	2,5	2,3	2,4	2,4	2,4	2,7	2,4	4,0	2,2	2,4	2,6
MUĞ	2,8	2,5	2,4	2,8	2,7	2,8	2,7	2,7	2,7	4,6	2,9	2,6
URF	2,6	2,2	2,3	2,2	2,3	2,6	2,9	2,6	2,5	2,3	4,5	2,6
ANK	2,1	1,9	2,0	2,1	2,1	2,0	2,0	2,4	2,3	2,1	2,2	1,9

Table 28. Number of resamples with at least as many assignments from population i (rows) to j (columns) after randomization by drawing existing individuals from combined gene pool for all populations (1st randomization).

	ESK	ART	HAK	HAT	KIR	CYP	ARD	IZM	KAS	MUĞ	URF	ANK
ESK	0	1000	1000	912	1000	1000	1000	743	718	231	1000	278
ART	909	0	907	1000	1000	1000	949	940	1000	256	1000	930
HAK	580	579	11	575	947	1000	1000	264	1000	771	12	830
HAT	924	1000	670	0	1000	741	933	1000	1000	1000	107	942
KIR	915	1000	1000	1000	0	1000	1000	552	1000	1000	1000	1000
CYP	1000	884	1000	890	1000	0	1000	1000	630	888	1000	1000
ARD	817	1000	786	1000	1000	834	0	1000	826	822	1000	1000
İZM	594	1000	1000	1000	874	1000	644	0	861	849	1000	1000
KAS	426	1000	883	1000	1000	927	1000	1000	0	688	723	269
MUĞ	142	923	100	1000	1000	1000	951	803	712	1	1000	131
URF	1000	1000	181	443	1000	1000	1000	1000	1000	916	0	1000
ANK	61	638	849	1000	1000	1000	718	689	890	642	629	3

Table 29. Mean number of assignments from population i (rows) to j (columns) after randomization by drawing new individuals from combined gene pool for all populations (2nd randomization).

	ESK	ART	HAK	HAT	KIR	CYP	ARD	İZM	KAS	MUĞ	URF	ANK
ESK	2,4	2,4	2,3	2,4	2,4	2,5	2,7	2,6	2,5	2,4	2,5	2,5
ART	2,6	2,5	2,3	2,5	2,5	2,6	2,8	2,7	2,4	2,5	2,5	2,6
HAK	3,0	3,0	2,8	2,8	2,9	3,0	3,3	3,2	2,9	2,8	2,9	3,0
HAT	2,6	2,4	2,3	2,6	2,5	2,6	2,8	2,6	2,5	2,5	2,5	2,6
KIR	2,6	2,5	2,3	2,5	2,5	2,6	2,8	2,8	2,5	2,4	2,5	2,6
CYP	2,3	2,2	1,9	2,2	2,2	2,3	2,4	2,3	2,3	2,1	2,2	2,3
ARD	1,7	1,6	1,6	1,7	1,7	1,7	1,8	1,8	1,7	1,6	1,7	1,8
İZM	1,9	1,9	1,8	1,9	1,9	2,0	2,1	2,1	1,9	1,9	2,0	2,0
KAS	2,5	2,4	2,2	2,4	2,4	2,5	2,7	2,7	2,5	2,3	2,5	2,5
MUĞ	2,6	2,6	2,3	2,6	2,5	2,7	2,9	2,8	2,6	2,5	2,7	2,8
URF	2,4	2,4	2,2	2,4	2,4	2,5	2,7	2,6	2,5	2,4	2,4	2,6
ANK	2,1	2,0	1,9	2,2	2,1	2,2	2,4	2,3	2,0	2,0	2,1	2,2

Table 30. Variances of number of assignments from population i (rows) to j (columns) after randomization by drawing new individuals from combined gene pool for all populations (2nd randomization).

	ESK	ART	HAK	HAT	KIR	CYP	ARD	İZM	KAS	MUĞ	URF	ANK
ESK	3,1	2,3	2,5	2,3	2,3	2,7	2,5	2,4	2,5	2,1	2,5	2,3
ART	2,3	3,1	2,2	2,2	2,4	2,4	2,5	2,5	2,4	2,5	2,6	2,7
HAK	2,8	2,7	3,6	2,7	2,7	3,0	3,2	3,3	2,9	2,7	3,0	2,7
HAT	2,4	2,5	2,2	3,1	2,2	2,6	2,7	2,5	2,5	2,5	2,5	2,6
KIR	2,6	2,3	2,1	2,4	2,1	2,5	2,7	2,7	2,5	2,5	2,3	2,5
CYP	2,2	1,9	1,9	2,1	2,1	2,8	2,2	2,2	2,0	1,9	2,1	2,2
ARD	1,8	1,6	1,5	1,5	1,6	1,8	1,9	1,8	1,6	1,5	1,7	1,9
İZM	1,8	1,7	1,7	1,8	1,8	2,0	2,1	2,6	2,0	1,9	1,9	1,9
KAS	2,5	2,3	2,2	2,3	2,3	2,4	2,6	2,4	3,2	2,0	2,4	2,3
MUĞ	2,6	2,6	2,2	2,5	2,5	2,7	2,8	2,8	2,5	3,2	2,8	2,8
URF	2,4	2,2	1,9	2,3	2,4	2,6	2,5	2,5	2,4	2,3	3,1	2,6
ANK	2,0	1,9	2,0	2,1	2,0	2,2	2,3	2,0	2,0	2,0	2,0	2,7

Table 31. Number of resamples with at least as many assignments from population i (rows) to j (columns) after randomization by drawing new individuals from combined gene pool for all populations (2nd randomization).

	ESK	ART	HAK	HAT	KIR	CYP	ARD	IZM	KAS	MUĞ	URF	ANK
ESK	0	1000	1000	918	1000	1000	1000	772	733	209	1000	253
ART	932	0	901	1000	1000	1000	953	941	1000	255	1000	927
HAK	582	596	5	543	962	1000	1000	231	1000	789	9	828
HAT	918	1000	685	0	1000	745	948	1000	1000	1000	108	924
KIR	923	1000	1000	1000	0	1000	1000	521	1000	1000	1000	1000
CYP	1000	912	1000	907	1000	0	1000	1000	693	893	1000	1000
ARD	824	1000	813	1000	1000	840	0	1000	838	833	1000	1000
İZM	597	1000	1000	1000	858	1000	649	0	860	845	1000	1000
KAS	460	1000	902	1000	1000	921	1000	1000	0	714	711	258
MUĞ	125	930	83	1000	1000	1000	948	767	750	0	1000	157
URF	1000	1000	179	454	1000	1000	1000	1000	1000	912	0	1000
ANK	61	625	860	1000	1000	1000	714	690	874	606	661	0

Table 32. Mean number of assignments from population i (rows) to j (columns) after randomization by drawing new individuals from each population's gene pool (3rd randomization).

	ESK	ART	HAK	HAT	KIR	CYP	ARD	IZM	KAS	MUĞ	URF	ANK
ESK	22,1	0,3	0,9	0,1	0,2	0,1	0,5	0,8	0,4	3,4	0,2	1,0
ART	0,5	26,1	0,7	0,0	0,0	0,1	0,9	0,4	0,2	1,6	0,1	0,5
HAK	1,8	1,2	20,4	1,8	0,3	0,5	1,8	0,7	0,4	2,2	3,1	1,7
HAT	0,3	0,1	1,5	26,1	0,0	0,4	0,2	0,0	0,1	0,2	1,8	0,4
KIR	0,4	0,0	0,1	0,0	30,0	0,0	0,0	0,2	0,1	0,1	0,0	0,0
CYP	0,2	0,1	0,2	0,1	0,0	24,8	0,3	0,2	0,3	0,2	0,2	0,3
ARD	0,4	0,6	0,6	0,1	0,0	0,3	17,7	0,5	0,1	0,3	0,1	0,2
İZM	0,6	0,1	0,1	0,0	0,0	0,1	0,4	21,7	0,1	0,8	0,0	0,1
KAS	1,1	0,2	0,4	0,2	0,1	0,5	0,2	0,5	23,5	0,9	0,8	1,5
MUĞ	3,9	1,4	1,3	0,2	0,0	0,2	0,8	2,0	0,8	19,3	0,5	1,6
URF	0,3	0,1	1,8	1,2	0,0	0,1	0,2	0,0	0,3	0,4	24,7	0,8
ANK	1,5	0,8	1,1	0,5	0,0	0,4	0,5	0,4	1,0	1,2	1,1	17,4

Table 33. Variances of number of assignments from population i (rows) to j (columns) after randomization by drawing new individuals from each population's gene pool (3rd randomization).

	ESK	ART	HAK	HAT	KIR	CYP	ARD	IZM	KAS	MUĞ	URF	ANK
ESK	5,6	0,3	0,8	0,1	0,2	0,1	0,5	0,8	0,4	3,0	0,2	1,0
ART	0,5	3,6	0,6	0,0	0,0	0,1	0,8	0,3	0,2	1,5	0,1	0,5
HAK	1,5	1,1	8,8	1,6	0,3	0,6	1,6	0,6	0,4	2,0	2,7	1,6
HAT	0,3	0,1	1,6	3,9	0,0	0,3	0,2	0,0	0,1	0,2	1,6	0,3
KIR	0,4	0,0	0,1	0,0	0,9	0,0	0,0	0,2	0,1	0,1	0,0	0,0
CYP	0,2	0,1	0,2	0,1	0,0	1,8	0,3	0,2	0,3	0,2	0,2	0,4
ARD	0,4	0,5	0,5	0,1	0,0	0,3	2,3	0,5	0,1	0,3	0,1	0,2
İZM	0,5	0,1	0,1	0,0	0,0	0,1	0,4	1,9	0,1	0,7	0,0	0,1
KAS	1,1	0,2	0,4	0,2	0,1	0,5	0,2	0,5	4,6	0,9	0,7	1,3
MUĞ	3,1	1,2	1,2	0,2	0,0	0,2	0,8	1,8	0,8	7,3	0,5	1,5
URF	0,3	0,1	1,6	1,1	0,0	0,1	0,2	0,0	0,3	0,4	3,9	0,8
ANK	1,4	0,7	1,0	0,4	0,0	0,3	0,5	0,4	0,9	1,2	1,0	6,1

Table 34. Number of resamples with at least as many assignments from population i (rows) to j (columns) after randomization by drawing new individuals from each population's gene pool (3rd randomization).

	ESK	ART	HAK	HAT	KIR	CYP	ARD	IZM	KAS	MUĞ	URF	ANK
ESK	988	100	1000	102	1000	1000	1000	191	58	443	1000	25
ART	369	985	487	1000	1000	1000	583	303	1000	78	1000	420
HAK	274	122	1000	281	290	1000	1000	0	1000	671	9	497
HAT	245	1000	450	999	1000	41	203	1000	1000	1000	24	304
KIR	308	1000	1000	1000	998	1000	1000	0	1000	1000	1000	1000
CYP	1000	118	1000	120	1000	990	1000	1000	41	167	1000	1000
ARD	341	1000	474	1000	1000	278	933	1000	107	291	1000	1000
İZM	110	1000	1000	1000	16	1000	50	1000	131	551	1000	1000
KAS	95	1000	345	1000	1000	416	1000	1000	1000	249	188	48
MUĞ	336	776	6	1000	1000	1000	529	592	201	1000	1000	18
URF	1000	1000	103	107	1000	1000	1000	1000	1000	362	944	1000
ANK	13	173	686	1000	1000	1000	91	68	643	363	303	1000

In order to test these three null hypotheses we calculated the probabilities of obtaining at least as many correct assignments as we observed from original data after randomization processes (Table 35). Null hypotheses that assumes that the 12 populations are actually one well mixed population either not or at HWE (1st and 2nd randomizations) were rejected. The

null hypothesis that assumes that each population is a seperate population at HWE is not rejected (3rd) at all with a very high probability (0,986).

Table 35. Probabilities of obtaining at least as many correct assignments as we observed after three randomization processes in case their differing null hypotheses are valid for populations.

Randomization Type	Probability
1st	0,013
2nd	0,000
3rd	0,986

Scatter graphs plotted using logarithms of assignment likelihoods of individuals for different population pairs provided us with a visual aid in understanding the genetic structures of populations. Assignment graphs for Kırklareli honeybees which had the highest differention in assignment tests and F_{ST} measures and Artvin and Ardahan honeybees, two northeastern Türkiye honeybee populations representing two ecotypes of *Apis mellifera caucasica*, could be seen below (Figures 13-42). The lines are x=y lines and the individuals located on these lines are equally assigned to both populations. As an individual goes far from the line the probability for it to be belonged to one of the populations increases.

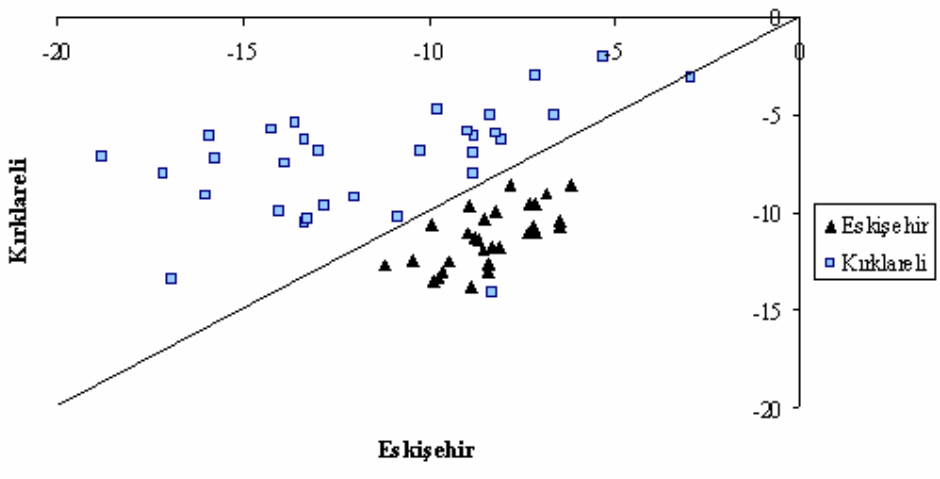

Figure 13. Log likelihood graph.

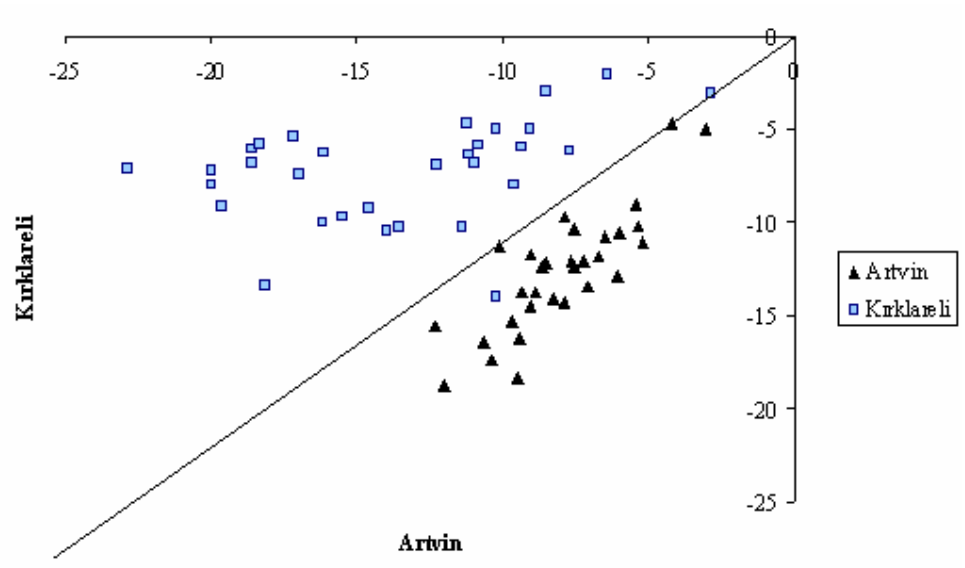

Figure 14. Log likelihood graph.

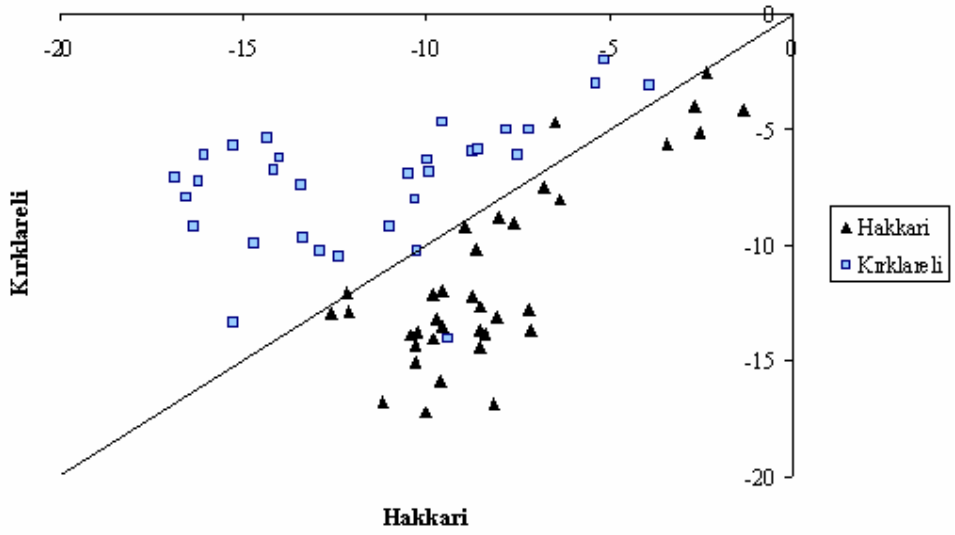

Figure 15. Log likelihood graph.

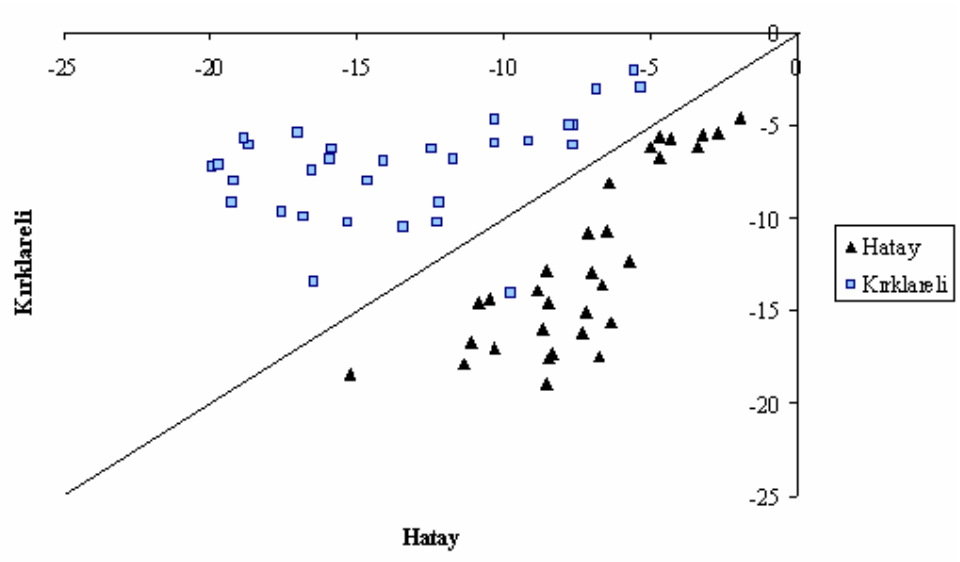

Figure 16. Log likelihood graph.

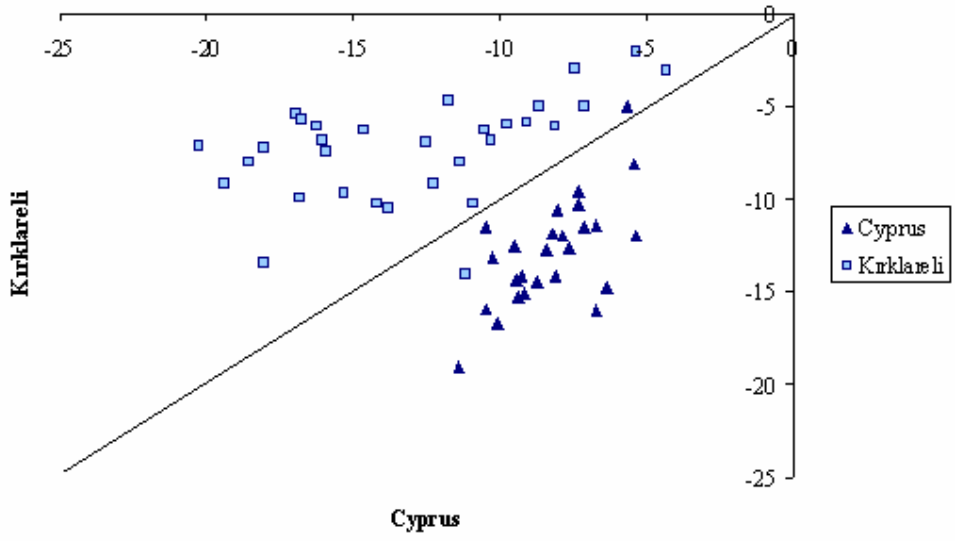

Figure 17. Log likelihood graph.

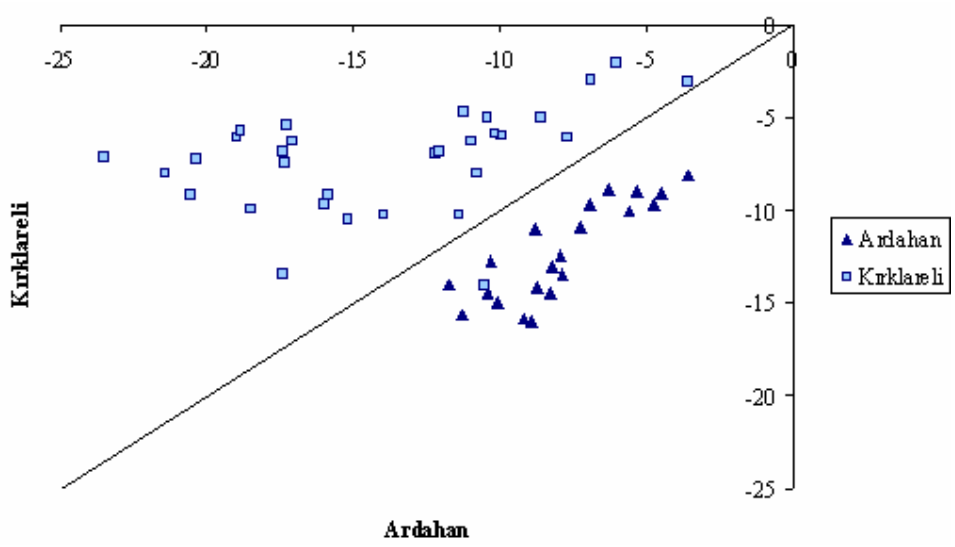

Figure 18. Log likelihood graph.

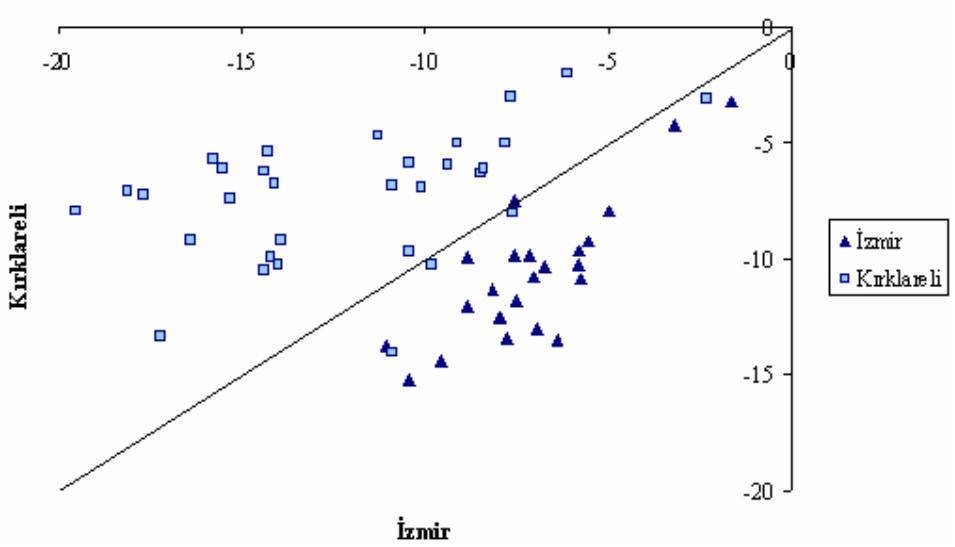

Figure 19. Log likelihood graph.

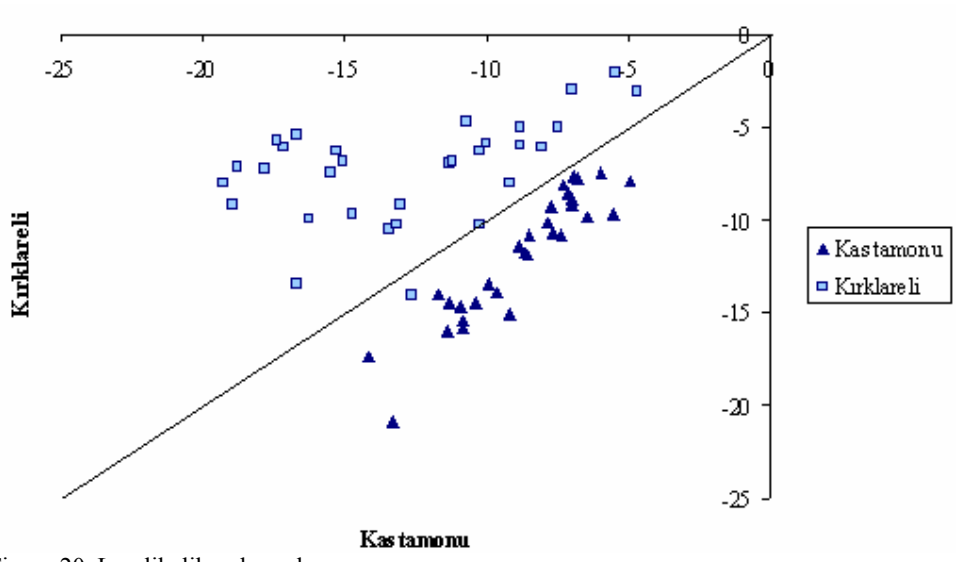

Figure 20. Log likelihood graph.

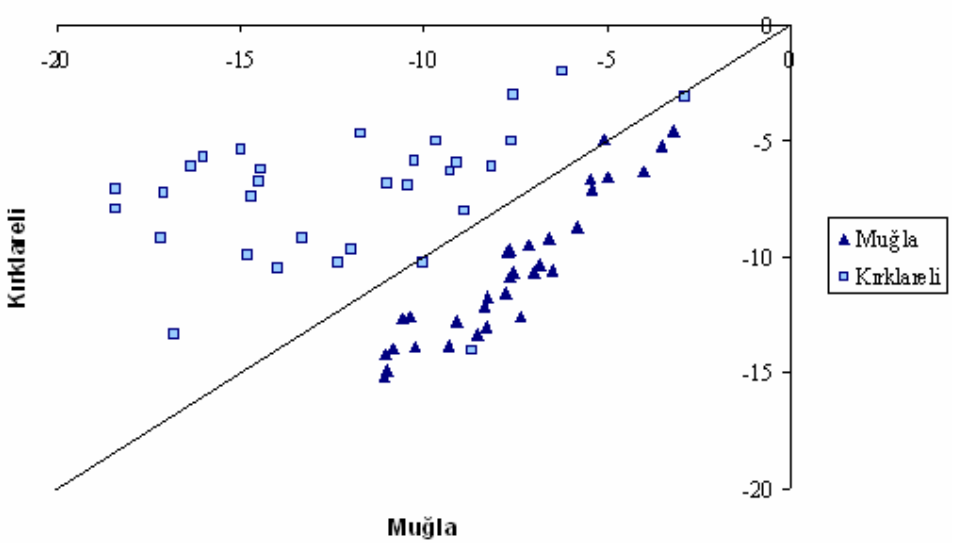

Figure 21. Log likelihood graph.

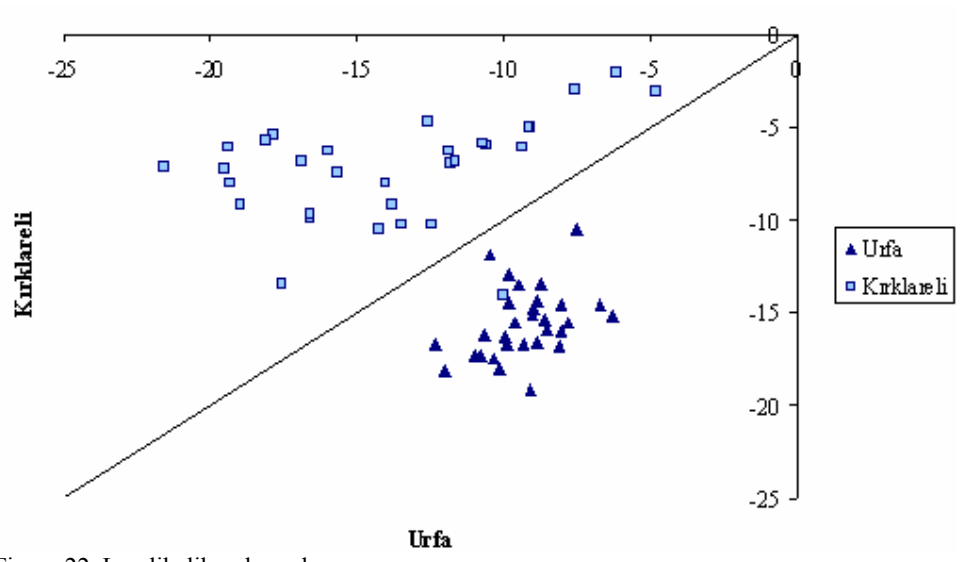

Figure 22. Log likelihood graph.

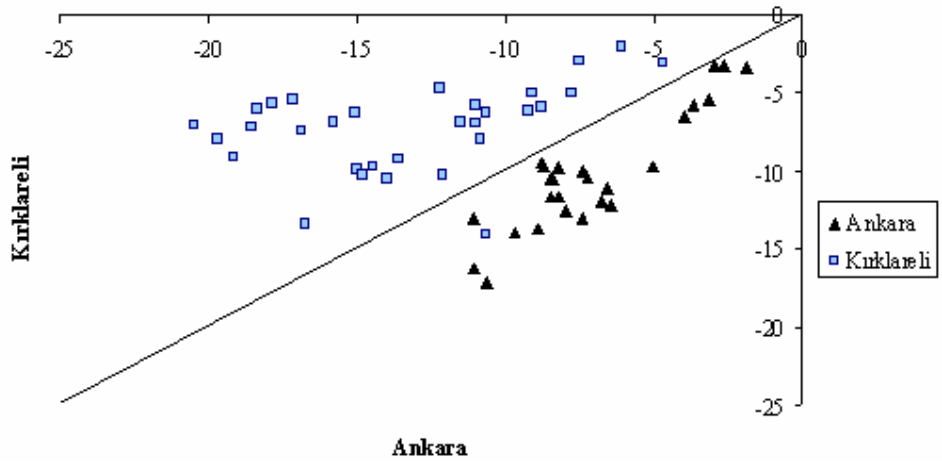

Figure 23. Log likelihood graph.

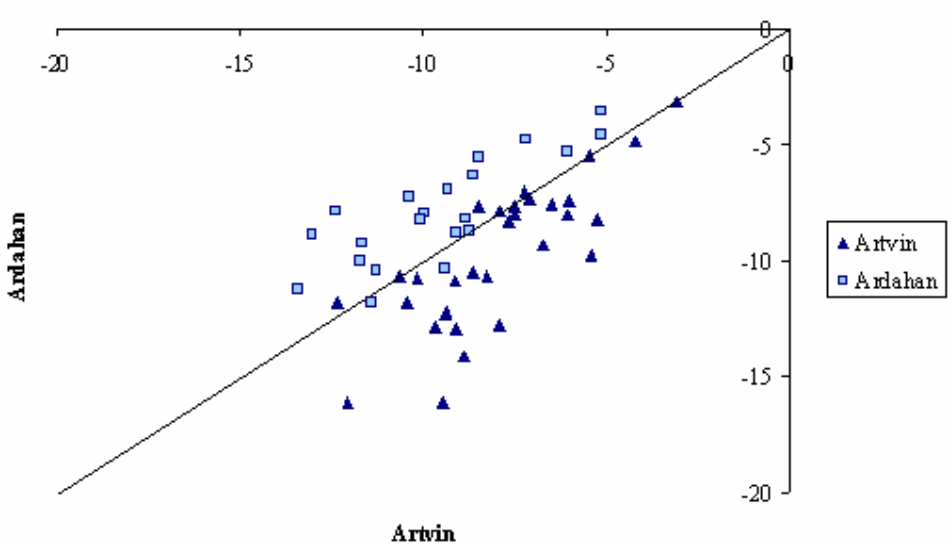

Figure 24. Log likelihood graph.

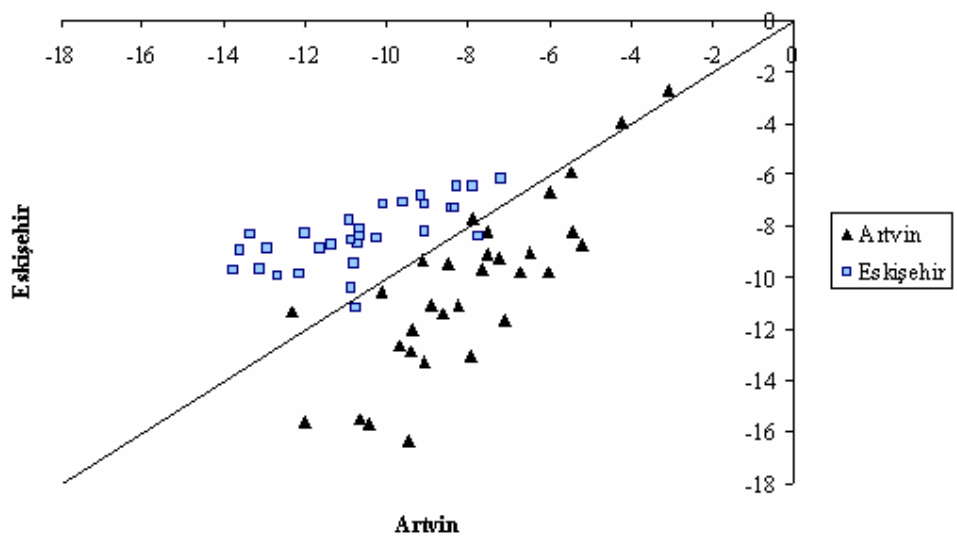

Figure 25. Log likelihood graph.

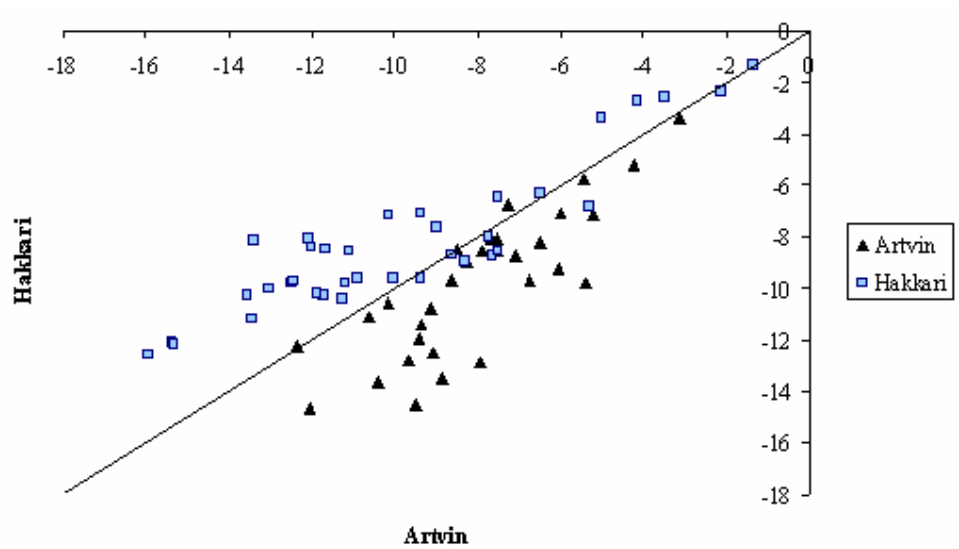

Figure 26. Log likelihood graph.

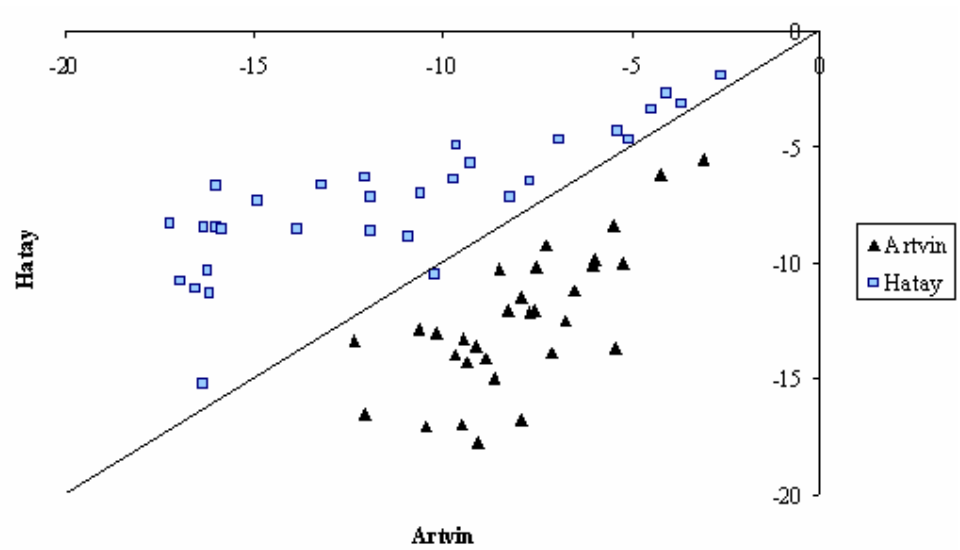

Figure 27. Log likelihood graph.

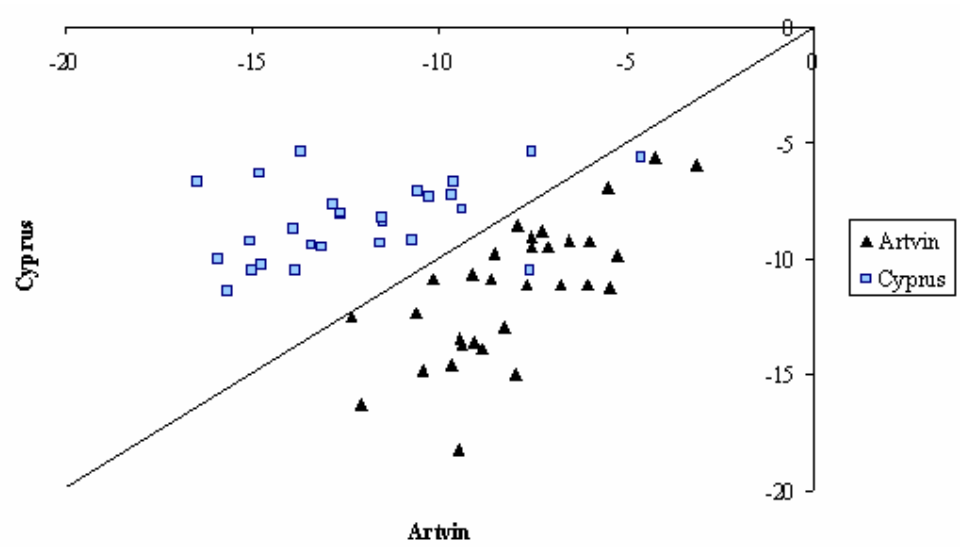

Figure 28. Log likelihood graph.

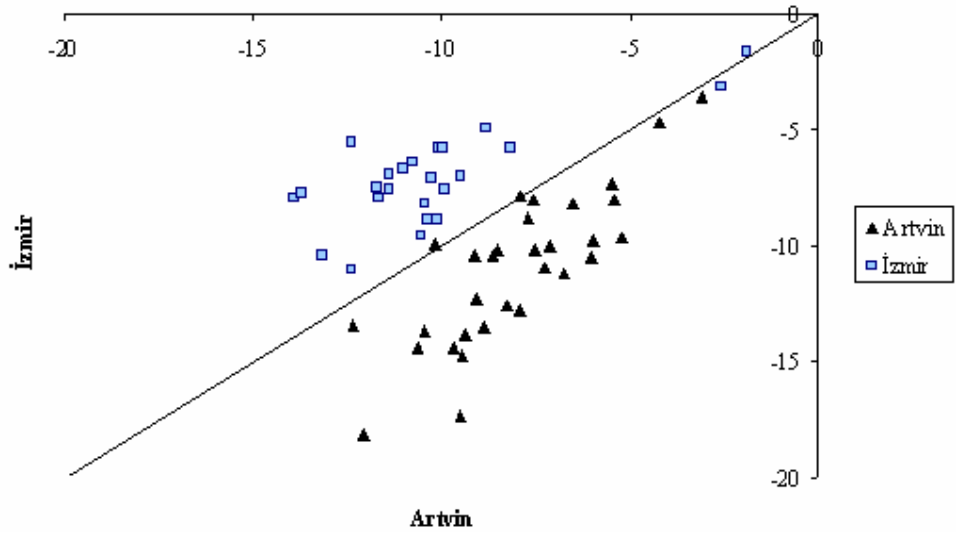

Figure 29. Log likelihood graph.

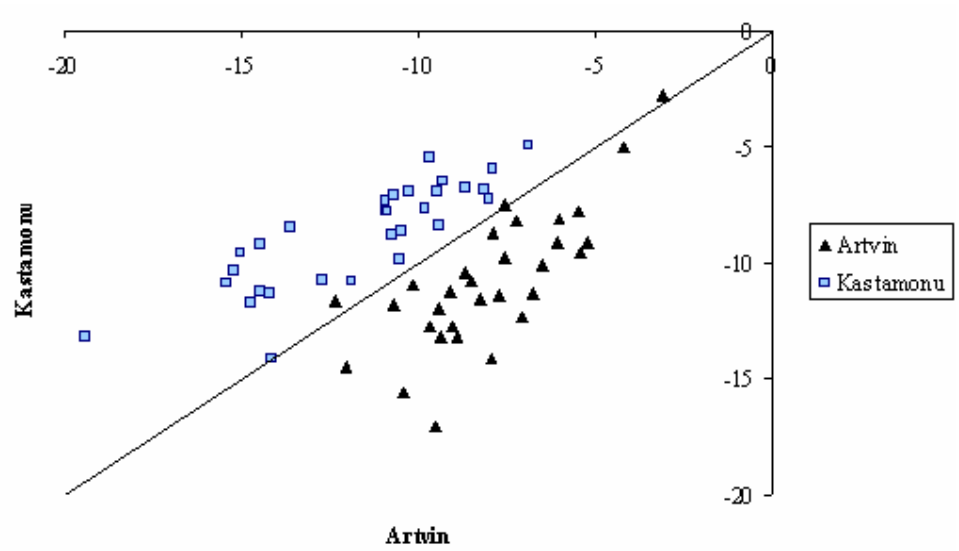

Figure 30. Log likelihood graph.

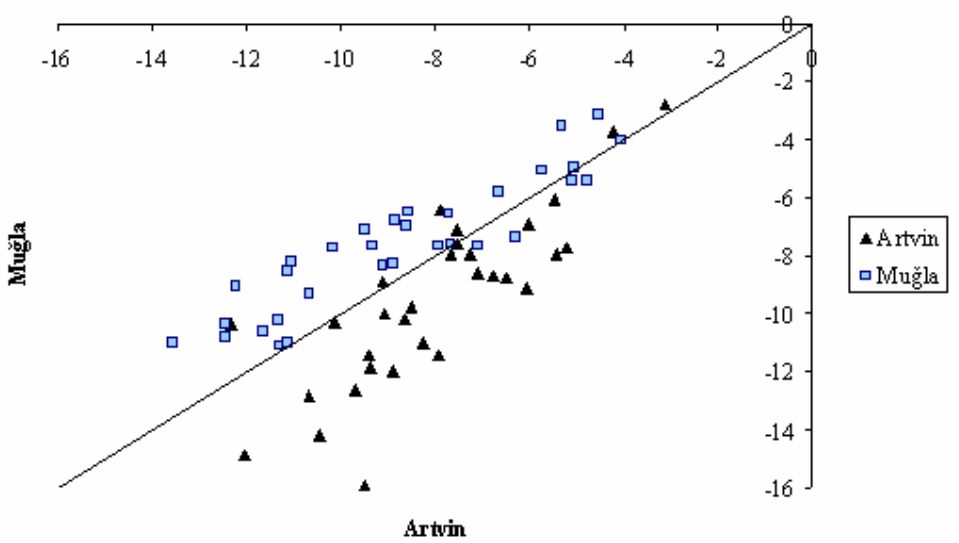

Figure 31. Log likelihood graph.

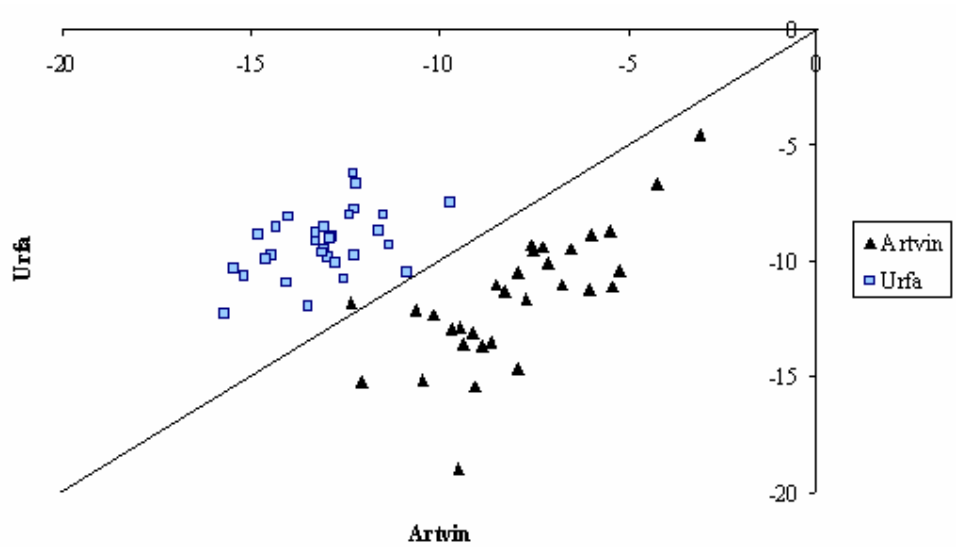

Figure 32. Log likelihood graph.

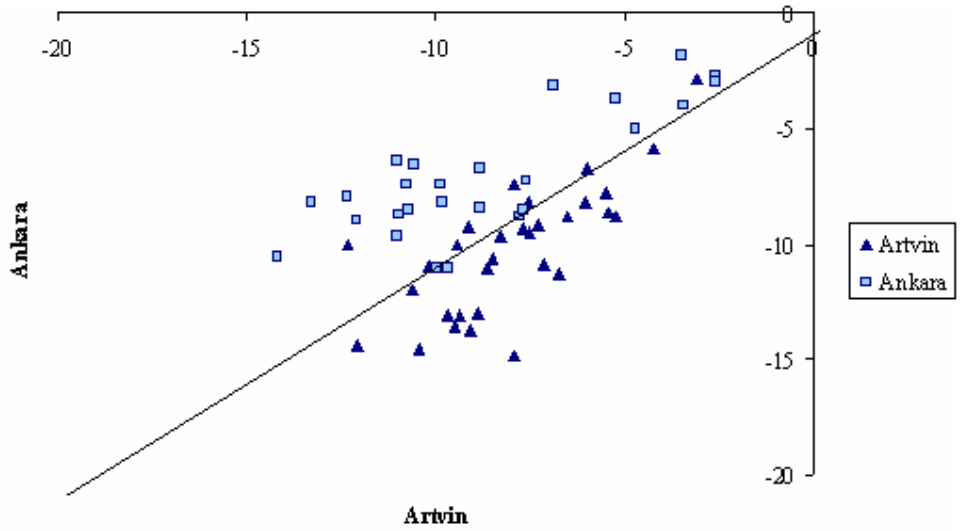

Figure 33. log likelihood graph

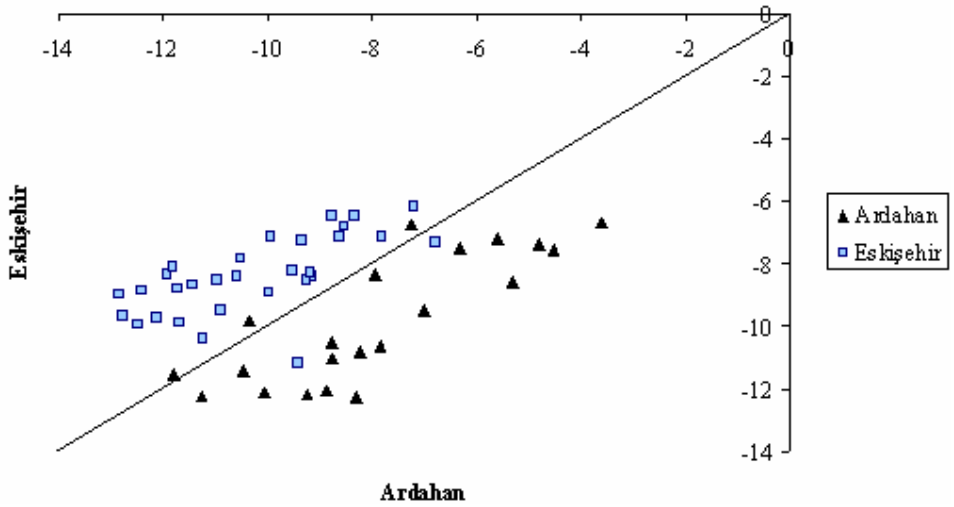

Figure 34. Log likelihood graph.

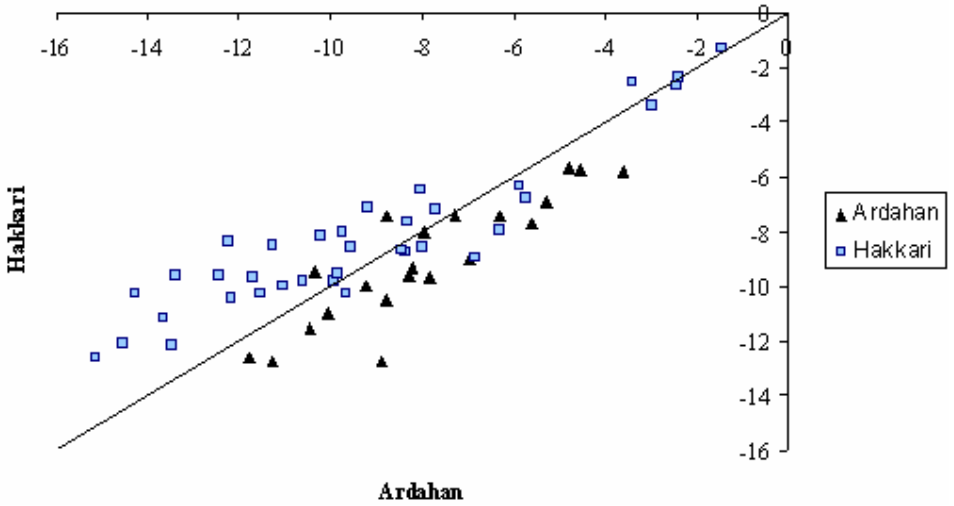

Figure 35. Log likelihood graph.

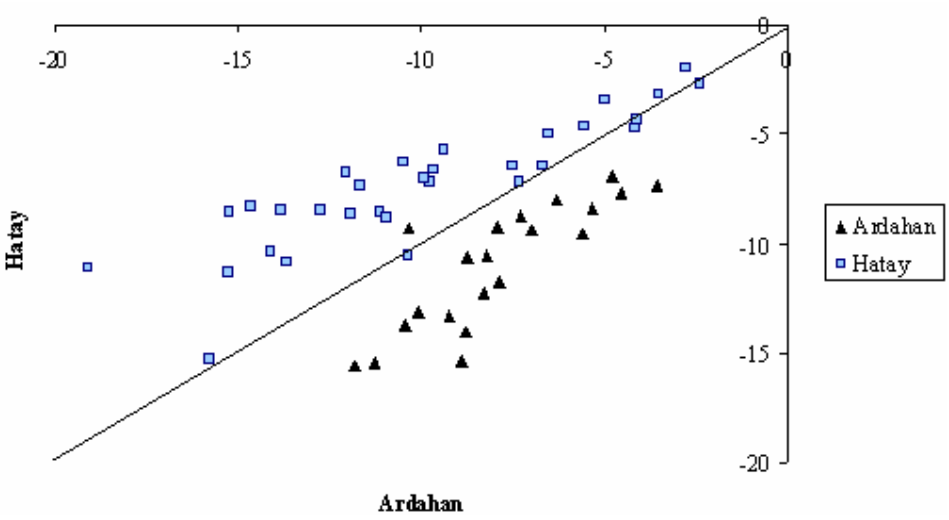

Figure 36. Log likelihood graph.

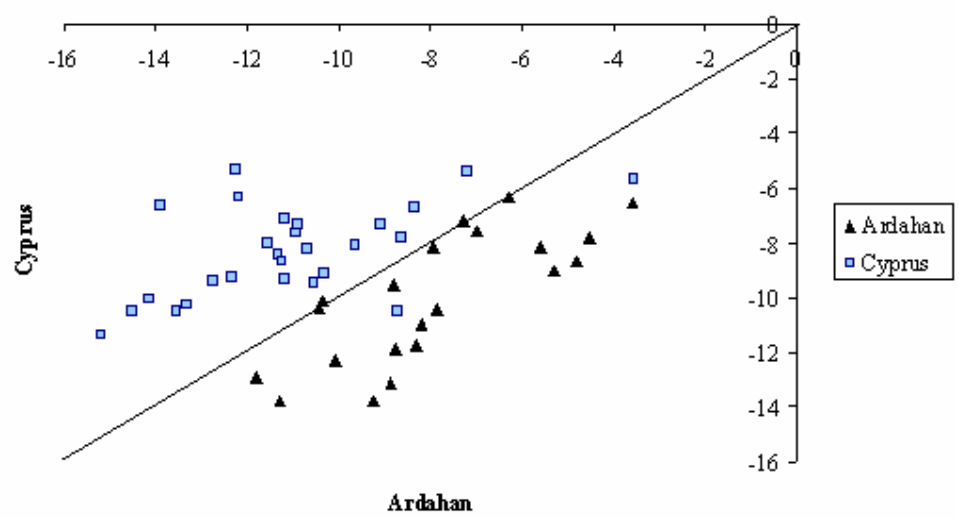

Figure 37. Log likelihood graph.

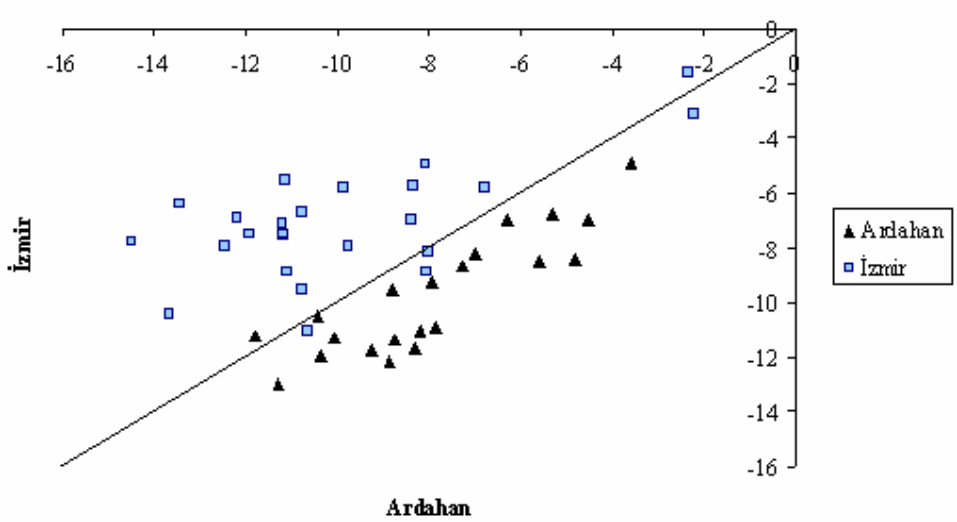

Figure 38. Log likelihood graph.

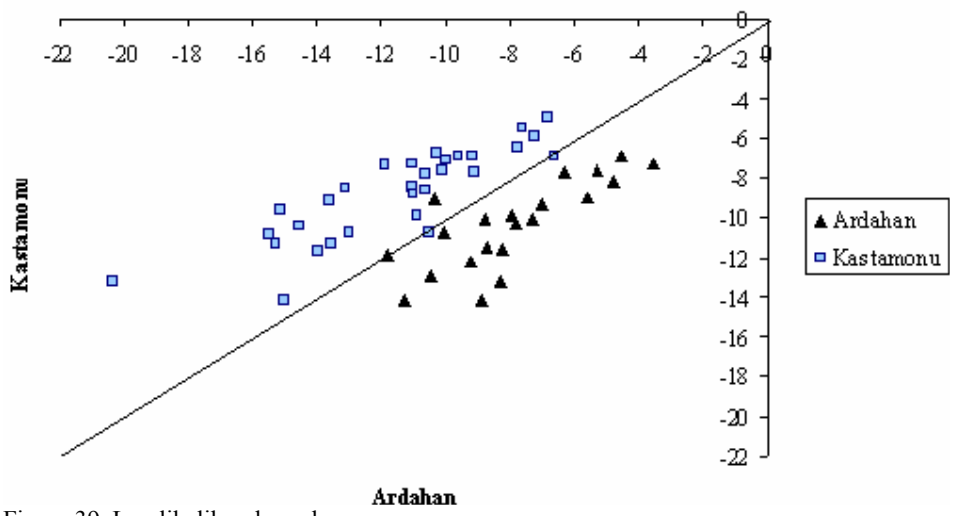

Figure 39. Log likelihood graph.

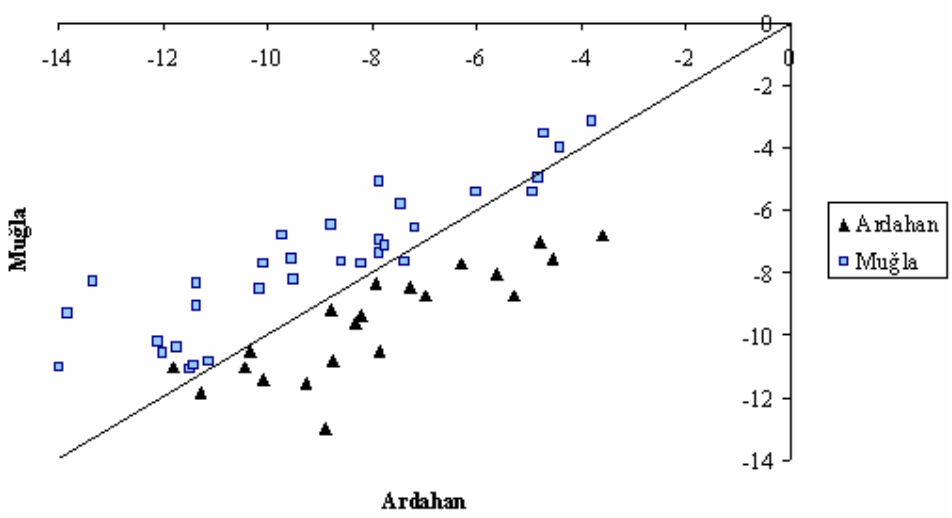

Figure 40. Log likelihood graph.

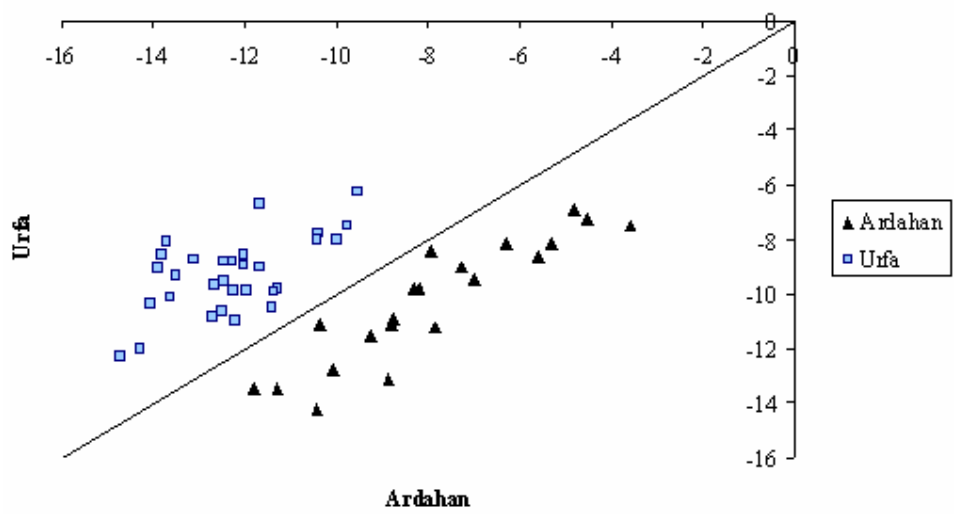

Figure 41. Log likelihood graph.

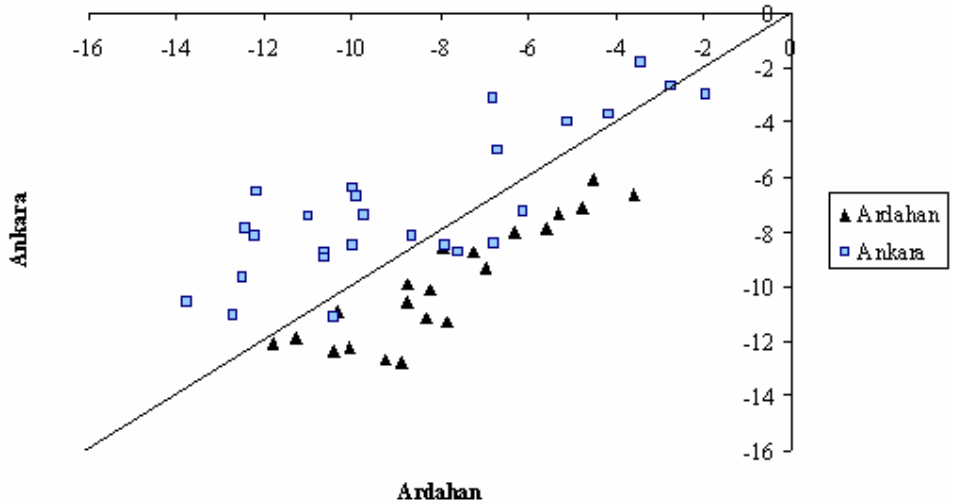

Figure 42. Log likelihood graph.

3.3.3.5. Genetic distances and population trees

Two genetic distance statistics that can be used for confirmation of the other since they treat data in different ways were used to create distance matrices among 12 honeybee populations (Table 35 and 36). These are Nei's Standard Distance (D_S) and assignment test based Likelihood Ratio Distance (D_{LR}). A correlation of 95 % is calculated between these measures for our data and this high correlation was demonstrated on a graph showing parallel changes in logarithms of these two statistics (Figure 43).

Table 36. Standard genetic distances among populations.

	ESK	ART	HAK	HAT	KIR	CYP	ARD	İZM	KAS	MUĞ	URF	ANK
ESK												
ART	0,09											
HAK	0,05	0,08										
HAT	0,17	0,20	0,05									
KIR	0,23	0,33	0,30	0,45								
CYP	0,10	0,14	0,05	0,10	0,29							
ARD	0,07	0,05	0,03	0,13	0,32	0,06						
İZM	0,06	0,10	0,11	0,28	0,25	0,14	0,08					
KAS	0,08	0,12	0,06	0,18	0,27	0,08	0,08	0,10				
MUĞ	0,01	0,05	0,04	0,15	0,25	0,08	0,05	0,04	0,06			
URF	0,15	0,17	0,04	0,05	0,44	0,12	0,11	0,24	0,13	0,12		
ANK	0,04	0,09	0,04	0,11	0,31	0,06	0,07	0,08	0,04	0,03	0,09	

Table 37. D_{LR} distances among populations.

	ESK	ART	HAK	HAT	KIR	CYP	ARD	İZM	KAS	MUĞ	URF	ANK
ESK	0,00											
ART	2,31	0,00										
HAK	1,17	1,70	0,00									
HAT	3,29	4,25	1,26	0,00								
KIR	3,55	5,51	3,73	5,73	0,00							
CYP	2,97	3,57	2,11	2,93	5,08	0,00						
ARD	1,89	1,76	1,08	2,83	5,50	2,43	0,00					
İZM	1,78	3,09	2,42	5,46	4,24	3,27	2,34	0,00				
KAS	1,87	2,82	1,88	3,39	4,29	2,35	2,60	2,50	0,00			
MUĞ	0,38	1,32	0,83	2,87	3,94	2,45	1,56	1,22	1,43	0,00		
URF	2,93	3,71	0,83	1,50	6,39	3,14	2,67	4,87	2,33	2,24	0,00	
ANK	1,11	1,93	0,84	2,06	4,57	2,13	1,80	2,40	1,16	0,86	1,51	0,00

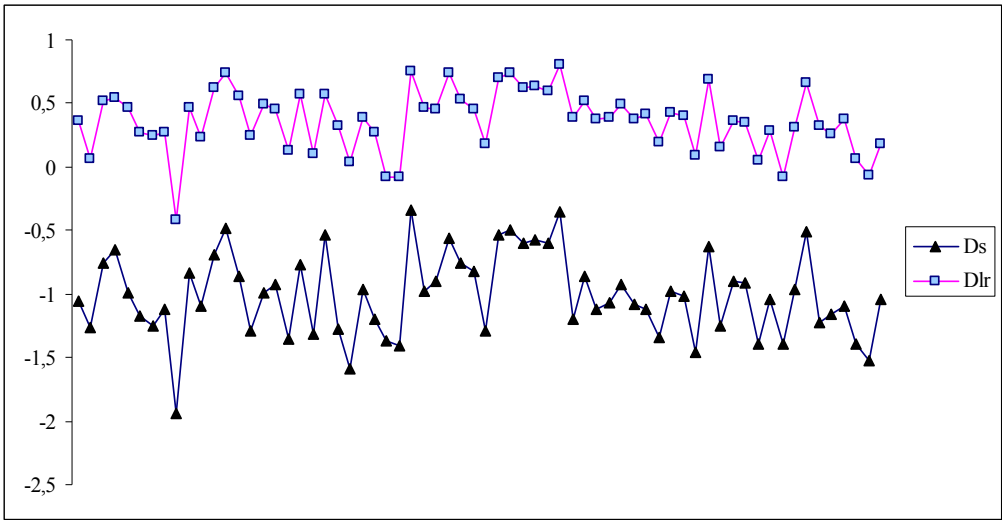

Figure 43. Parallel changes in logarithms of genetic distance measures D_S and D_{LR} for our microsatellite data.

Two different genetic distances gave very similar phylogenetic trees constructed by Neighbour Joining (NJ) method (Figures 44 and 45). In both trees Kırklareli population were separated from others completely (with 100 percent bootstrap values in D_S tree). Hatay and Urfa populations grouped together as close neighbours in both trees. Three general groups were formed in both trees containing western Türkiye (Eskişehir, Muğla and İzmir), eastern Türkiye (Hatay, Urfa and Hakkari) and northern Türkiye (Ardahan and Artvin).

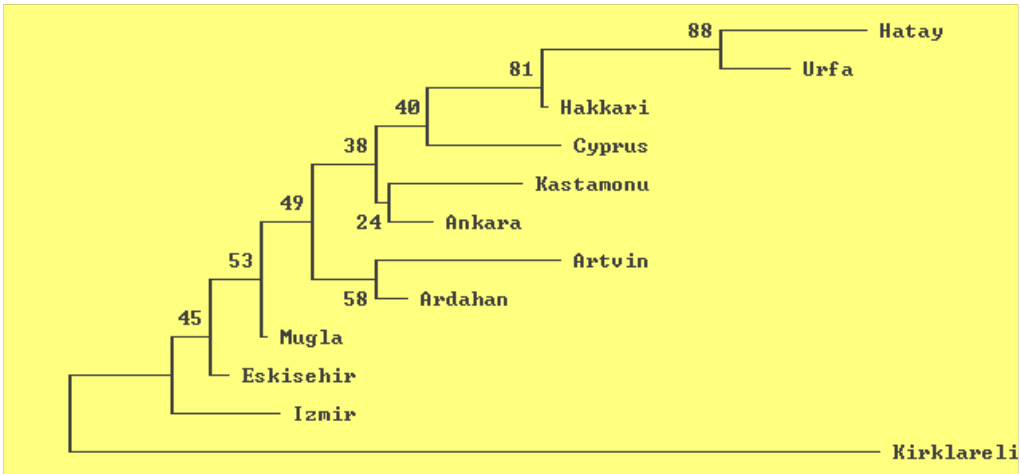

Figure 44. Neighbour Joining tree based on D_S. Numbers show bootstrap percentages for the cluster at the right to be connected to the lower nodes.

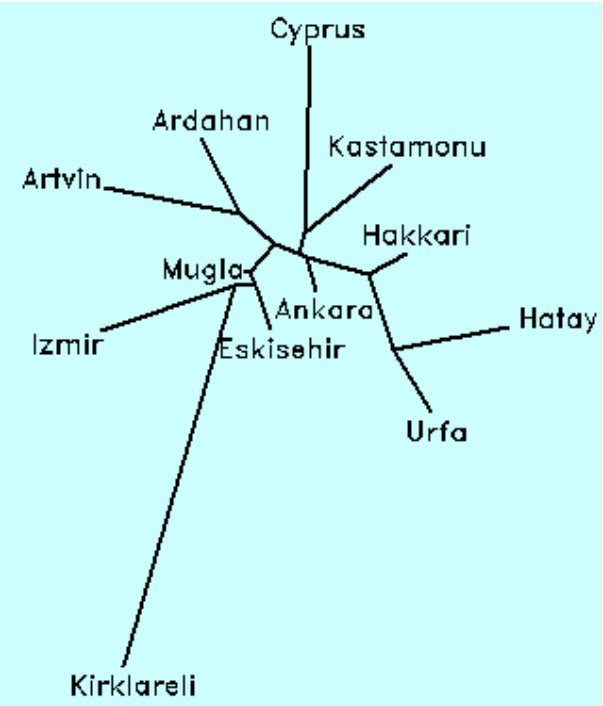

Figure 45. An unrooted tree constructed by Neighbour Joining method based on D_{LR}.

CHAPTER 4

DISCUSSION

The average gene diversities (expected heterozygosities assuming Hardy-Weinberg equilibrium) for Türkiye and Cyprus changed between 0,542 (Eskişehir) and 0,681 (Kastamonu) with a grand mean of 0,612. The average observed heterozygosities (proportion of heterozygotes) changed between 0,516 (Eskişehir) and 0,670 (Cyprus) with a grand mean of 0,609. Thus there is no any general deficit or excess of genic diversity.

So far, western honeybee (*Apis mellifera* L) microsatellite studies were generally concentrated on Europe and Africa in the Old World, the original distribution area of honeybees. Only one population from Lebanon was studied and reported from Middle East (Franck *et al*. 2000a, 2001) for which the gene diversity was recorded to be 0,65. Microsatelite studies conducted on western European honeybees (M lineage), north Mediterranean (C lineage) and African honeybees (A lineage) indicated that heterozygosity levels were highest within African honeybee populations which changed between 0,76 and 0,90 (Franck *et al*. 2001). Heterozygosity levels for C lineage were intermediate between A and M lineage levels reported to change between 0,39 and 0,68 (Franck *et al*. 2000b). Honeybee populations from M lineage which is distributed among western Europe were found to have the lowest heterozygosity levels changing between 0,26 and 0,68 (Garnery *et al*. 1998, Franck *et al* 2001).

The values for mean number of alleles for each locus in each population showed that allelic polymorphism is also changing in the order of A,C and M lineages from greatest to the lowest. The mean allelic number for Türkiye and Cyprus populations (6,95) is closest to reported numbers for C lineage (Estoup *et al*. 1995a,Garnery *et al*. 1998).

In an 8 microsatellite study on 7 populations from western Europe (Spain, Portugal and France) honeybees and 4 populations from eastern Africa (Morocco and Guinea) , Franck *et al*. (1998) found gene diversities (expected heterozygosities) changing between 0,29 and 0,38 for M lineage and between 0,77 and 0,88 for A lineage populations. When we calculated the average gene diversities (expected heterozygosities) for 5 microsatellite loci (A24,A113,A7,A43,A28) that we used in common with Franck *et al*. (1998) these values changed between 0,230 and 0,395 for western European populations and between 0,764 and 0,896 for African populations. Our results for these common 5 loci show that heterozygosity of populations from Türkiye and Cyprus changes between 0,524 and 0,693 (mean:0,636).

A general deficit for genic diversity for M populations and their allelic range being within the range for African lineage A, supported the hypothesis of colonization of western European bees by African honeybees. But mtDNA studies rejects this hypothesis since there is not any detected M haplotype in Africa. Mitochondrial DNA studies suggested an ancient divergence between A and M lineages (Franck *et al*. 1998). Nuclear and mitochondrial markers often show discordant patterns of differentiation in the honeybee (Franck *et al*. 2001).

In a 6 microsatellite study (4 loci are same with the ones used in our study) on 8 African honeybee populations, Franck *et al*. (2001) found that gene diversities were changing between 0,756 and 0,896. They also studied 3 C lineage populations which gave gene diversities ranging between 0,406 and 0,663 and 3 M lineage populations that gave gene diversities between 0,259 and 0,356. A gene diversity of 0,636 was detected in a *syriaca* population from Lebanon as a representative of O lineage.

Garnery *et al*. (1998) found in a 11 microsatellite study that gene diversities were changing between 0,339 and 0,678 among 15 M lineage (western European) populations. When we calculated the average gene diversities (expected heterozygosities) for 6 microsatellite loci (A24, A113, A7, A43, A28 and Ap43) that we used in common with Garnery *et al*. (1998) these values changed between 0,200 and 0,659 for western European populations. Our results for these common 6 loci show that heterozygosity of populations from Türkiye and Cyprus changes between 0,563 and 0,724 (mean:0,668). Mean number of alleles per population per locus was calculated as 6,55 for M lineage populations (Garnery *et al*.

1998). For two populations each representing C and A alleles mean number of alleles per locus was 7,82 and 10,82 respectively.

Garnery *et al*. (1998) reported some alleles that are present in considerably high frequencies in a C lineage population but either absent or present in lower frequencies in 15 M lineage populations, as diagnostic alleles between M and C lineages. Among these diagnostic alleles 108 allele of A24 locus, 127 and 141 alleles of A43 locus, 116, 118, 120, 122, 123, 126, 128, 130, 132, 135, 137, 142, and 156 alleles of A7 locus, 214 allele of A113 locus, 143, 145 and 147 alleles of Ap43 locus and 138 allele of A28 locus are relevant to our study.

The diagnostic alleles for A24, A113, A7, A43, and A28 loci (Garnery *et al*. 1998) are also supported with frequecies reported in a 7 microsatellite study on 9 populations representing M,C and A lineages (Estoup *et al*. 1995a). In this study gene diversities were reported to change between 0,291 and 0,410 for M lineage, 0,464 and 0,612 for C lineage and 0,788 and 0,872 for A lineage populations. When we calculated the average gene diversities for 5 microsatellite loci (A24, A113, A7, A43, and A28) that we used in common with Estoup *et al*. (1998) these values changed between 0,232 and 0,400 for western European populations, between 0,410 and 0,564 for northern Mediterranean populations and between 0,764 and 0,869 for African populations. Our results for these common 5 loci show that heterozygosity of populations from Türkiye and Cyprus changes between 0,524 and 0,693 (mean: 0,636). Mean number of alleles per population per locus were 4,83 for M, 5,67 for C and 9,3 for A lineages in this study.

Gene diversities between 0,39 and 0,68 were reported for C lineage populations in an 8 microsatellite study on 6 honeybee populations from Italy and Sicily (Franck *et al*. 2000b). The gene diversity for a Lebanon population (O lineage) was reported to be 0,647 (Franck *et al*. 2000a).

When we look at A24 locus range is very similar to the range of C lineage except that a 96 allele which is not reported for any of C,M and A lineages seems to have an increasing frequency trend going through the eastern Türkiye. This allele has zero frequencies in Kırklareli, Cyprus and İzmir but its frequency increases up to 0,367 in Urfa. Another allele, 108 which was stated as a diagnostic allele between C and M lineages were found in 6 out

of 11 populations but with much lower frequencies than it was detected in C lineage. Another allele 102 that was reported only for African populations (Franck *et al*. 1998, Estoup *et al*. 1995a and Garnery *et al*. 1998) was found in 6 out of 11 populations in Türkiye and Cyprus (not in Hatay and Urfa).

When A113 allele frequencies is analysed, a very similar range is the case with previously studied C lineage populations and 214 allele which was reported as a diagnostic allele between C and M lineages was found to be present in a very high frequency (0,438) in Kırklareli population (Thrace) very similar to C lineage frequencies. The alleles 226, 228 and 230 are present in most of the populations in frequencies higher than 0,1 but these alleles were reported to be absent or lower than 0,1 in C lineage or A and M lineages (Garnery *et al*. 1998, Estoup *et al*. 1995a, Franck *et al*. 1998). Another allele 212 were found in considerable frequencies (0,205 and 0,117) in only two populations Hatay and Urfa which was only detected in African populations in such high frequencies.

A7 locus was found to be the most polymorphic microsatellite locus with a 68 alleles detected in honeybee populations from Türkiye and Cyprus. Range of this locus is between 99 and 200 which was recorded as 103-160 for M, 98-150 for C and 98-177 for A lineages (Estoup *et al*. 1995a,Franck *et al*. 1998 and Garnery *et al*. 1998). This level of polymorphism was not reported previously in any study among different lineages of honeybees. The highest reported number of alleles detected at this locus was 33 (Garnery *et al*. 1998). Except 122 allele all the 13 reported diagnostic alleles (Garnery *et al*. 1998) for C lineage were present in some populations in differing proportions.

Size range of A43 locus (117-148) is also higher than the ranges of M and C lineages (126-148). Both 127 and 141 alleles that were reported as diagnostic ones for C lineage were detected in low frequencies only in Hakkari and Hatay populations respectively at A43 locus. A novel A43 allele, 119, which was not reported for any lineage was found to be in a considerable frequency (0,175) in Hatay population. Allele 142 that was not reported for C lineage previously was found to be in considerable frequencies in all 11 populations from Türkiye and Cyprus.

Range of A28 alleles are very similar to previously recorded ranges for A, M, and C lineages. Allelic findings in A28 locus is very interesting. The most common allele by far

for this locus is 138 with frequencies changing between 0,688 and 1. This allele was reported as the only diagnostic allele for A28 locus for C lineage (Garnery *et al*. 1998). This allele was reported only in low frequencies for African and western European honeybee populations but in very high frequencies (0,870 and 0,967) in C lineage populations (Estoup *et al*. 1995a, Garnery *et al*. 1998, Franck *et al*. 1998). Thus this allele seems to be a good indicator that supports that honeybee populations of Türkiye and Cyprus are among C lineage. However the presence of 133 allele in considerable frequncies in 10 out of 11 populations and the presence of 129 allele in 4 out of 11 populations in considerable frequencies indicate a distinctness since these alleles were only reported for African populations previously.

The reported diagnostic alleles of C lineage at Ap43 locus (143, 145, and 147) were also detected in populations of Türkiye and Cyprus in relatively high frequencies supporting the presence of C lineage within Türkiye and Cyprus. However another allele showed a distinct feature of these populations. Allele 175 which was not reported previously in any lineage was found to be in considerabe frequencies (0,017-0,161) in 10 out of 11 populations.

Especially the presence and frequencies of diagnostic alleles at A28, Ap43, and A113 loci strongly supports that the honeybees from all over the Türkiye and Cyprus belong to C lineage. However the presence and frequencies of novel alleles that was not detected previously in any lineage at A24, A113, and Ap43 loci, and the alleles that were only reported in African populations at A24, A113, and A28 calls the idea that these populations could be distinct from all three lineages A, M, and C.

When F_{ST} measures are analysed per locus, it is observed that Ap226, A43, Ac306, and A113 performed best in terms of differentiating honeybee populations in Türkiye and Cyprus as indicated by their high values (Table 15). Ap226, Ap68, and Ac306 loci were not used in widespread honeybee population genetic studies so far. Performances of these new loci in differentiating honeybee populations of Türkiye and Cyprus in our study showed that especially Ap226 and Ac306 loci have great potential as divergence markers for honeybee populations.

Pairwise F_{ST} values that are reported between lineages are generally higher than 0,1 (0,06-0,61) (Franck *et al* 2000a, 2000b, 2001, Garnery *et al*. 1998). Within lineage pairwise F_{ST}

levels are generally lower than 0,1 level for M and A lineages but there are F_{ST} values up to 0,19 within M lineage (Garnery *et al*. 1998). This could be different for C lineage populations among which F_{ST} values up to 0,24 was reported (Franck *et al* 2000a, 2000a, 2001, Garnery *et al*. 1998). We observed a very high differentiation among 11 populations from Türkiye and Cyprus when compared to previous studies in Europe and Africa. Pairwise F_{ST} values are changing between zero and 0,2. Heterogeneity as much as found in this study was not reported for such a limited geographical region before. In a study among honeybee populations from Italy and Sicily (C lineage) pairwise F_{ST} values reported to change between 0,004 and 0,051. In Garnery *et al*.'s study (1998) among western European populations (Spain, Portugal, France, Belgium, and Sweden) pairwise F_{ST} values changed between 0,002 and 0,185 in a wide geographical area. And pairwise F_{ST} values were recorded to change between 0,01 and 0,12 in Franck *et al*.'s study (2001) among African populations. We found that 52 population pairs out of 66 are significantly different at 0,05 significance level which is an indicator of an extraordinary differentiation. In Garnery *et al*.'s (1998) work on M lineage populations only 19 out of 105 population pairs were recorded to differentiate at this significance level.

When the pairwise F_{ST} values are analysed closely we see that honeybee populations sampled from Kırklareli, the European part of Türkiye and İzmir, west end of Türkiye seem to genetically differentiate from all others significantly. The population pairs that did not differentiate are generally geographically close populations with the exceptions of Kastamonu-Cyprus (≈ 650 km), Kastamonu-Muğla (≈ 600 km), Kastamonu-Hakkari (≈ 900 km), Ankara-Hakkari (≈ 1000 km), Ankara-Cyprus (≈ 500 km), Ankara-Artvin (≈ 800 km), and Ankara-Posof (≈ 850 km) population pairs which are not close. Kastamonu population failed to differentiate from 4 populations (Artvin, Hakkari, Muğla, and Cyprus). Least pairwise differentiation was observed for Ankara population which is located at Central Anatolia region. This population genetically diverged from only 4 of the 11 other populations (Kırklareli, Hatay, İzmir, and Urfa).

Gene flow is known to decrease genetic divergence. Pairwise number of migrants (Nm) values (35 values higher than 2 out of 66 values) indicate that there is considerable potential for genetic divergence among most of the populations although overall Nm values is higher than 2. The highest numbers of migrants per generation seem to be for Hakkari, Eskişehir, Ankara and Muğla populations. Gene flow due to migratory beekeeping

activities seems to be able to decrease the distinctnesses of these remote populations seriously (Eskişehir and Ankara are close provinces). Kırklareli, Cyprus, Artvin, İzmir, and Kastamonu populations have especially low amount of migration rates according to Nm values. These Nm values are also in agreement with high genetic differentiation of these populations as indicated by pairwise F_{ST} values.

One interesting discordance between F_{ST} and Nm values is between Hakkari and Kırklareli populations. Although this population pair is significantly different according to F_{ST} value (0,139), the Nm value (2,587) indicates considerable gene flow among them. The region that we collected Kırklareli samples are fairly isolated region where honeybees are not travelled long. Thus this discrepancy may be a result of the difference in the ways that these two measures treat the data. Although Nm statistic considers private allele frequencies that are present in only one individuals F_{ST} statistic is primarily affected from intermediate frequency alleles.

High level of genetic divergence among honeybee populations of Türkiye and Cyprus was also confirmed by high pecentages of correct assignments. Assignment tests gave percentages differing between 57 and 87 percent for 9 out of 12 populations analysed. Three populations seems extremely heterogeneous and affected by high migration rates. These populations are sampled from Hakkari, Muğla and Ankara for which correct assignment proportions are 0,25, 0,34, and 0,35 respectively. These results together with F_{ST} and Nm measures indicate that Hakkari, Muğla, and Ankara populations are seriously affected by migratory beekeeping and their gene pools are being contaminated with introgression of foreign bees. Gene pool of so called "Muğla Bee", if exists, is seriously under danger. The highest assignment scores were obtained in honeybees of Kırklareli, Cyprus, Ardahan, Artvin, and İzmir which indicates high level of genetic differentiation at these populations in agreement with F_{ST} scores.

Among the populations that show very low differentiation, Hakkari is a region where migratory beekeeping activities with Black Sea, Mediterranean, and Southeastern Regions of Türkiye and Iran are frequent. Muğla region receives foreign honeybees from Eastern Anatolia and Thrace in winter and honeybees of this region are carried to Central Anatolia and Marmara region in summer during migratory beekeeping activities. Ankara region is again seriously affected from migratory beekeeping between Eastern Anatolia and Aegean

Regions, between Central anatolia and Marmara Regions and between Black Sea and Marmara regions. There are only one diagnostic alleles for each of Muğla and Hakkari populations.

Twelve Türkiye and Cyprus honeybee populations are determined as separate populations in Hardy-Weinberg equilibrium with data randomization tests we performed during assignment tests. This result further strengthened our other results indicating significant population differentiation among honeybees of Türkiye and Cyprus.

Camili (Artvin) and Posof (Ardahan) honeybees are under conservation in order to prevent gene flow. It is forbidden to import stranger honeybees into the region. Camili honeybees (tagged as Artvin in this study) proved to remain distinct as indicated by pairwise F_{ST} and Nm measures. These two populations represent two ecotypes of the same subspecies, *A.m.caucasica*.

Camili (Artvin) population seems divergent and well conserved as visible by a high percentage of correct assignment number, 71 %. This population diverged from all populations except from Ardahan, Kastamonu and Ankara populations based on pairwise F_{ST} values. Posof is about 60 km, Kastamonu is about 600 km and Ankara is 800 km away from Camili. Pairwise Nm values that are under 2 level with 9 out of 11 also supports genetic divergence of Camili (Artvin) population. The 2 Nm value that are over 2 are the ones with Hakkari and Ankara populations which seem to be extremely heterogeneous populations. We detected 4 diagnostic alleles for this conserved population located at the border with Georgia.

Another conserved population, Posof (Ardahan) failed to diverge from only Artvin, Hakkari, and Ankara populations based on pairwise F_{ST} values. Very high proportion of correct assignment (76%) further marks the genetic distinctness in this population. We marked 7 diagnostic alleles for this population. But pairwise Nm values with Eskişehir, Hakkari, İzmir, Urfa, and Ankara were over 2 level indicating a considerable migration rate among Ardahan and these populations that may homogenize Posof Bee in the future.

Caucasica honeybees which are well adapted to cool climate of Caucasus Mountains and humid regions among sea level of Black Sea are seen as a hybridizing agent and queens of

this subspecies are sold to several regions of Türkiye without serious consideration of climate and habitat adaptation. This is especially obvious for Ardahan bees in our analyses. We also know that Camili (Artvin) queenbees were recently introduced to Edirne which is located at Thrace region. This region is not humid and hot summers are predominant which will certainly cause adaptation problems and bad performances of hybrid bees and more importantly loss of gene sources in this European region of Türkiye.

Honeybees of Kırklareli are differing from all other populations based on high F_{ST} values, low number of migrants (Nm), a very high correct assignment percentage (87 %) and population trees based on genetic distances. However pairwise F_{ST} values are differing between 0,076 and 0,200. Since this much F_{ST} values were previously reported within M and C lineages we don't have basis to assume that honeybees of Anatolia and Cyprus belong to another lineage instead of C lineage that Kırklareli population is known to belong. Moreover there are other pairwise F_{ST} values among Anatolian haneybee populations that exceeds 0,1. Thus together with the presence of diagnostic alleles our results supports the mtDNA results in suggesting that Anatolian and Cypriot honeybees belong to north Mediterranean (C) lineage. In addition to this high allelic ranges, high number of alleles, great amount of genetic differentiation detected by F_{ST} values and assignment tests indicate that Anatolia could be thought as a gene center for C lineage.

We found 5 diagnostic alleles that are in high frequencies in Kırklareli and absent or in very low frequencies within Anatolian and Cypriot honeybee populations. Honeybees of Trace region were naturally isolated after the formation of Bosphorus about 7.000 years ago. We also detected 5 diagnostic alleles for Anatolia which is in relatively high frequencies and absent or in very low frequencies within Kırklareli population.

Hatay samples which were found to have a unique mtDNA haplotype were argued to represent the fourth evolutionary branch (Smith *et al*. 1997). In another mtDNA study Kandemir *et al*. (submitted) found African elements in some colonies sampled from this region. In our study we found that Hatay samples not genetically diverged from Urfa samples according to pairwise F_{ST} values and Hatay samples are clustered with Urfa and Hakkari samples as eastern Türkiye group in phylogenetic trees. However we detected 4 diagnostic alleles for this region which are in relatively high frequencies in Hatay population and either absent or in very low frequencies in other populations.

İzmir population sampled from Karaburun town was found to be highly differentiated as indicated by 71 % correct assignment score, high pairwise F_{ST} scores (Significantly different from all other populations) and low pairwise Nm values. This location is at the west end of Anatolia and experiences low level of migratory beekeeping movements in short distances. Furthermore we detected 3 diagnostic alleles for this region. These alleles are present in relatively high frequencies in İzmir population and either absent or in very low frequencies in other populations.

Urfa population was found to be among the populations that show the highest genetic differentiation. A correct assignment score of 73 % and high pairwise F_{ST} values (significantly different from all populations except Hatay population). Two alleles were detected that are diagnostic for this population.

Cyprus population is relatively isolated from migratory beekeeping activities and thus is the second highly differentiated population as indicated by 81 % correct assignment score, high pairwise F_{ST} values (significantly different from all populations except Kastamonu and Ankara) and low pairwise Nm values. Moreover we detected 8 diagnostic alleles which have relatively high frequencies in Cyprus and either absent or have very low frequencies in other populations.

Kastamonu honeybee population which has famous "delibal" honey gave a relatively good assignment score of 57 %. Nm values are not high pairwise F_{ST} values showed that this population is not differentiated from Artvin, Cyprus, Hakkari, and Muğla populations. We detected 3 diagnostic alleles for this population.

When the population trees are analysed, what we could see in common is first of all separation of Kırklareli population (100 % bootstrap value in D_S). A general separation among trees may be simplified by stating the main clusters are western (Kırklareli, Eskişehir, Muğla, and İzmir), northern (Artvin, Ardahan, and Kastamonu) and eastern (Hakkari, Urfa, and Hatay) Türkiye groups. Ankara and Cyprus populations are placed somewhere almost equally distant to these 3 main clusters.

In a study conducted in Mexico, the seasonal frequencies of European honeybee drones and African derived honeybee drones have been shown to vary in mating areas according to different peaks in male production in these types of honeybees (Quezada-Euan *et al*. 2001). This phenomenon may be a partial genetic barrier between different types of honeybees (Quezada-Euan *et al*. 2001). This kind of genetic barrier could also be effective in preservation of the high genetic differentiation among Türkiye and Cyprus honeybees. Another study that was conducted to analyze the *A.m.ligustica* introgression in *A.m. mellifera* populations, showed that the admixture among these subspecies was either zero or at most 10 % (Jensen *et al*. 2005).

Microsatellites are fast evolving markers which are very suitable for intraspecific population genetic studies. The microsatellite loci have advantages of being mostly neutral, having high mutation rates and exhibiting codominant inheritance as population genetic study markers for closely related species and populations over the morphometric and electrophoretic markers which are subject to selection pressure (Freeman and Herron 1998). Polygenic determinism is the major drawback of morphometrical characters which proved to be useful in discriminating honeybee populations (Ruttner 1988). Mitochondrial DNA is another high resolution marker in population genetic studies as microsatellites. However this uniparentally inherited marker has drawbacks such as inheritance as a single allele without recombination.

Dynamics of microsatellite evolution is not completely resolved yet. Although main mutation mechanism is DNA slippage, point mutations, insertions and deletions, recombinational events are also involved in these processes. Length constraints, mutation biases, differences in mismatch repair mechanisms, differences depending on age, sex and organism further complicates the evolution of these markers. In case of microsatellite loci analyses; infinite allele model (IAM) and stepwise mutation models (SMM) are basic models. IAM assumes that a mutation occurs in a microsatellite locus with an addition or loss of repeat unit or units regardless of the number resulting in a novel allele that was not present previously in the population. However stepwise mutation model states that a mutation for a microsatellite allele occurs by addition or loss of only one repeat unit and the new allele may be present previously in the population. There are several other models suggested as explained in introduction section of this thesis. It seems that the ideal model should include a mutational bias and a balance should be assumed between DNA slippage

and point mutations that break large microsatellite alleles. Simulations and direct observations to test these models showed that the mutational mechanisms are alternating depending on microsatellite loci and organisms. Repeat types and whether a locus is perfect or interrupted may cause serious changes in mutational processes. Hence it is not possible to talk about a perfect mutation model for all kind of microsatellites. Several microsatellite loci were shown not to follow SMM and it is suggested to be sure about the mutation mechanism of the locus if one is using SMM based models.

The difference between allelic polymorphism among 9 microsatellite loci we used is actually a result of different rates of mutations and mechanisms in different types of loci. Rates of mutation and mutational mechanisms depend on length constraints, selection, point mutations, repeat numbers, repeat types, flanking regions, and recombination rates (Schlötterer 2000). Interrupted microsatellites are believed to be less variable than uninterrupted ones since interruptions seem to stabilize the tract in core region (Estoup *et al.* 1995b).

Among the microsatellite loci that we have used in this study A28 is a compound locus that contain both di and tri nucleotide repeats. Other 2 loci, A113 and Ap43 are interrupted loci which contain several interruptions among dinucleotide repeats. Among these loci A113 was previously studied for mutation mechanisms (Estoup *et al*. 1995b). In this study A113 locus was reported not to follow SMM but be suitable for IAM. Point mutations are thought to be involved in evolutionary process especially for interrupted loci (Estoup *et al*. 1995b). In our study we used IAM based genetic distances (F_{ST} and D_S) since we have 2 interrupted and one compound microsatellites for which it is impossible to calculate SMM based statistics.

In population genetic studies increasing the number of microsatellite loci was reported to be more important than choosing the mutation model or focusing on size homoplasy (Estoup *et al*. 2002). Increased number of microsatellites could compensate for the decreased polymorphism beacause of homoplasy. Moreover in honeybees within lineage microsatellite polymorphism were shown to be not affected by size homoplasy by molecularly accessible size homoplasy (MASH) studies (Estoup *et al*. 1995b).

We have chosen Nei's standard genetic distance measure D_S and genotype likelihood ratio distance D_{LR} in order to construct phylogenetic trees. We used the novel D_{LR} which treats the data in radically different ways than D_S in order to test D_S measure and we obtained a 95 % correlation among them which strengthens our results. Among several genetic distances D_S is among the classical ones which was shown to increase linearly with time under IAM if mutation drift equilibrium is maintained in the evolution of populations (Takezaki and Nei 1996). Genetic distance statistics based on SMM use variance in repeat numbers, however the statistics based on IAM use variance in allelic frequencies (Richard and Thorpe 2001). Many simulation that detect the linearity of genetic distances with time and their variances indicated that IAM based classical distances performed better than SMM based measures which had high variances (Takezaki and Nei 1996, Paetkau *et al*. 1997, Gaggiotti *et al*. 1999). Especially D_S, D_A and D_{LR} distances was found to be best performers in these studies. Phylogeny construction studies also supported the success of IAM based measures over SMM based distances in giving correct tree topology (Richard and Thorpe 2001).

CHAPTER 5

CONCLUSION

Our microsatellite analyzes on honeybee populations of Türkiye and Cyprus support the mtDNA findings that Anatolian honeybees belong to C lineage. Our analyses further revealed that Anatolia is a genetic center for north Mediterranean (C) lineage. Characterisation of Anatolian honeybees by microsatellites in addition to mtDNA was an essential task in understanding of honeybee evolution. In order to understand the evolution and distribution of honeybees better we believe that molecular characterization of Iranian honeybees by mtDNA and microsatellites is needed. It is still not clear where the honeybees firstly speciated and spreaded through the original distribution areas.

We detected an extraordinary genetic differentiation of honeybee populations within Türkiye based on pairwise F_{ST} values. This level of differentiation among populations was not reported before for European or African populations. Correct assignment scores indicated very high genetic structures for most of the populations.

In recent years there are several attempts to introduce, rear and sell Italian (*ligustica*) and Carniolan (*carnica*) honeybees in Türkiye. In a few locations in Aegean and Mediterranean Regions of Türkiye these honeybees have been reared and sold throughout the country for a couple of years. People who support this introduction have arguments stating that countries such as Australia, China, and USA, Italian and Carniolan honeybees were introduced and flourished successfully. But these countries are not original distribution areas of *Apis mellifera*, and there was not any western honeybee before this introductions in those regions. However local honeybee subspecies in Türkiye and Cyprus have been adapted to the extremely divergent climate and habitat conditions for thousands of years. Hence attempts to replace these local honeybees by foreign honeybees will spoil the adaptations and genetic diversity attained by local honeybees. Reduced genetic diversity will probably lead to inability to adapt when environmental conditions change.

There are ideas about replacing the Urfa honeybees with Italian honeybees. We found that this population is one of the most conserved populations. The local honeybees of this region has their adaptation to the local climate and has defensive tactics against local predators evolved in thousands of years. We think that this idea of replacement is unacceptable since agriculture in this area have been dependent on pollination by these honeybees.

Our genetic analyzes indicated that the isolated areas that are formed in Artvin-Camili, Ardahan-Posof and Kırklareli areas proved very efficient in conservation of gene pools of these honeybee populations. It is forbidden to carry foreign honeybees to these regions. Furthermore based on high genetic differentiation indicated by high assignment scores, diagnostic alleles and high F_{ST} values we suggest that İzmir-Karaburun and Cyprus honeybee populations should be next conservation areas for the preservation of differentiation. To forbid carrying foreign queenbees to these regions seems logical when we consider the geographical locations of these populations. It is relatively easy to control the entrance to these regions since Cyprus population is a naturally isolated one and İzmir population is very close to the west end of Anatolia.

Other than populations of these regions Hatay and Kastamonu populations are also genetically diverged. Introduction and trade of Italian and Carniolan queenbees to Türkiye and Cyprus should be forbidden in order to preserve these enormous genetic differentiation among honeybee populations. If these precautions are not taken legally, genetic pollution of honeybee populations of Türkiye may lead loss of the rich genetic resources of Middle East and C lineage honeybee populations.

REFERENCES

Adam Brother (1983).In search of the best strains of bees. Dadant sons, Hamilton Illinois.

Amos W, Sawcer SJ, Feakes RW, *et al*. (1996). Microsatellites show mutational bias and heterozygote instability. Nature Genetıcs 13 (4): 390-391.

Anderson TJC, Su XZ, Roddam A, *et al*. (2000). Complex mutations in a high proportion of microsatellite loci from the protozoan parasite *Plasmodium falciparum*. Molecular Ecology 9 (10): 1599-1608.

Angers B, Bernatchez L (1997). Complex evolution of a salmonid microsatellite locus and its consequences in inferring allelic divergence from size information. Molecular Biology and Evolution 14 (3): 230-238.

Arias MC and Sheppard WS (1996) Molecular phylogenetics of honeybee subspecies (*Apis mellifera* L) inferred from mitochondrial DNA sequence. Mol. Phylogen. Evol. 5: 557-566.

Asal Ş, Kocabaş C, Elmacı C, Yıldız MA (1995) Enzyme polymorphism in honeybee (*Apis mellifera* L) from Anatolia. Turk. J. Zool. 19: 153-156.

Berlocher SH (1984) Insect molecular systematics. Annu. Rev. Entomol. 29: 403-433.

Berube M, Aguilar A, Dendanto D, *et al*. (1998). Population genetic structure of North Atlantic, Mediterranean Sea and Sea of Cortez fin whales, *Balaenoptera physalus* (Linnaeus 1758): analysis of mitochondrial and nuclear loci. Molecular Ecology 7 (5): 585-599.

Bodur Ç (2001). Microsatellite analysis in honeybee populations of Turkey. M.Sc. thesis, Middle East Technical University, Ankara, Türkiye.

Bodur Ç, Kence M, Kence A (2004). Genetic structure and origin determination in honeybee populations of Anatolia. Proceedings of the First European Conference of Apidologie: 40. Udine, Italy.

Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL (1994) High resolution of human evolutionary trees with polymorphic microsatellites. Nature 368: 455-457.

Braaten DC, Thomas JR, Little RD, Dickson KR, Goldberg I, Schlessinger D, Ciccodicola A, d'Urso M (1988) Locations and contexts of sequences that hybridize to poly(dG-dT).(dC-dA) in mammalian ribosomal DNAs and two X-linked genes. Nuc. Ac. Res. 16: 865-881.

Calabrese P, Durrett R (2003). Dinucleotide repeats in the *Drosophila* and human genomes have complex, length-dependent mutation processes. Molecular Biology and Evolution 20 (5): 715-725.

Colson I and Goldstein DB (1999). Evidence for complex mutations at microsatellite loci in *Drosophila*
Genetıcs 152 (2): 617-627.

Cornuet JM (1986). Bee Genetics and Breeding (Edited by Rinderer TE). Academic Press Inc.
426 p.

Hartl DL and Clark AG (1997). Principles of population genetics. Sinauer Associates Inc. Publishers. Sunderland, Massachussettes. 542 p.

Cornuet JM (1982). The MDH polymorphism in some West Mediterranean honeybee populations. In M. D. Breed, C. D. Michener and H. E. Evans (eds.), The Biology of Social Insects. Proceedings, 9th Congress of the IUSSI. Westview Press, Boulder, CO.

Darendelioğlu Y and Kence A (1992) Morphometric study on population structure on honeybee, *Apis mellifera* L. (Hymenoptera: Apidae). Türkiye 2. Entomoloji Kongresi Bildirileri: 387-396.

De la Rua P, Galian J, Serrano J, Moritz RFA (2001). Genetic structure and distinctness of *Apis mellifera* L. populations from the Canary Islands. Molecular Ecology 10 (7): 1733-1742.

De la Rua P, Galian J, Serrano J, Moritz RFA (2002). Molecular characterization and population structure of the honeybees from the Balearic islands (Spain). Apidologie 32 (5): 417-427.

De La Rua P, Galian J, Serrano J, Moritz RFA (2003). Genetic structure of Balearic honeybee populations based on microsatellite polymorphism. Genetıcs selectıon evolutıon 35 (3): 339-350.

Di Rienzo A, Donnelly P, Toomajian C, et al. (1998). Heterogeneity of microsatellite mutations within and between loci, and implications for human demographic histories. Genetics 148 (3): 1269-1284.

Ellegren H (2004). Microsatellites: Simple sequences with complex evolution. Nature Reviews Genetics 5 (6): 435-445.

Estoup A, Solignac M, Harry M, Cornuet JM (1993) Characterization of $(GT)_n$ and $(CT)_n$ microsatellites in two insect species: *Apis mellifera* and *Bombus terrestris*. Nuc. Ac. Res. 21: 1427-1431.

Estoup A, Solignac M, Cornuet JM (1994). Precıse assessment of the number of patrılınes and of genetıc relatedness ın honeybee colonıes. Proceedıngs of the royal society of London series b-biologıcal sciences 258 (1351): 1-7.

Estoup A, Garnery L, Solignac M, Cornuet JM (1995a) Microsatellite variation in (*Apis mellifera* L) populations: hierarchical genetic structure and test of the infinite allele and stepwise mutation models. Genetics 140: 679-695.

Estoup A, Tailliez C, Cornuet JM, *et al*. (1995). Size homoplasy and mutational processes of interrupted microsatellites ın 2 bee species, *apis-mellifera* and *bombus-terrestris* (apıdae). Molecular Biology and Evolution 12 (6): 1074-1084.

Estoup A, Jarne P, Cornuet JM (2002). Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. Molecular Ecology 11 (9): 1591-1604.

Feldman MW, Bergman A, Pollock DD, *et al*. (1997). Microsatellite genetic distances with range constraints: Analytic description and problems of estimation. Genetics 145 (1): 207-216.

Felsenstein J (1988). Phylogenies from molecular sequences - inference and reliability. Annual Review of Genetics 22: 521-565.

Forbes SH, Hogg JT, Buchanan FC, *et al*. Microsatellite evolution in congeneric mammals - domestic and bighorn sheep. Molecular Biology and Evolution 12 (6): 1106-1113.

Franck P, Garnery L, Solignac M, Cornuet JM (1998) The origin of west European subspecies of honey bees (*Apis mellifera*): new insights from microsatellite and mitochondrial data. Evolution 52: 1119-1134.

Franck P, Garnery L, Solignac M, Cornuet JM (2000a) Molecular confirmation of a fourth lineage in honeybees from the Near East. Apidologie 31: 167-180.

Franck P, Garnery L, Celebrano G, Solignac M, Cornuet JM (2000b). Hybrid origins of honeybees from Italy (*Apis mellifera ligustica*) and Sicily (*A-m. sicula*). Molecular Ecology 9 (7): 907-921.

Franck Garnery L, Loiseau A (2001). Genetic diversity of the honeybee in Africa: microsatellite and mitochondrial data. Heredıty 86: 420-430.

Freeman S, Herron JC (1998) Evolutionary Analysis. Prentice-Hall, Inc., Simon & Schuster/ A Viacom Company, New Jersey, 786 p.

Gaggiotti OE, Lange O, Rassmann K, *et al*. (1999). A comparison of two indirect methods for estimating average levels of gene flow using microsatellite data. Molecular Ecology 8 (9): 1513-1520.

Garnery L, Cornuet JM, Solignac M (1992) Evolutionary history of the honeybee *Apis mellifera* inferred from mitochondrial DNA analysis. Mol. Ecol. 1: 145-154.

Garnery L, Solignac M, Celebrano G, Cornuet JM (1993) A simple test using restricted PCR-amplified mitochondrial DNA to study the genetic structure of *Apis mellifera* L. Experientia 49: 1016-1021.

Garnery L, Mosshine EH, Cornuet JM (1995) Mitochondrial DNA variation in Moraccan and Spanish honey bee populations. Mol. Ecol. 4: 465-471.

Garnery L, Franck P, Baudry E, Vautrin D, Cornuet JM, Solignac M (1998) Genetic diversity of the west European honey bee (*Apis mellifera mellifera* and *A.m. iberica*). II. Microsatellite loci. Genet. Sel. Evol. 30 (Suppl. 1): S49-S74.

Garza JC, Slatkin M, Freimer NB (1995). Microsatellite allele frequencies in humans and chimpanzees, with implications for constraıntsonallele size. Molecular Biology and Evolution 12 (4): 594-603.

Garza JC, Freimer NB (1996). Homoplasy for size at microsatellite loci in humans and chimpanzees. Genome Research 6 (3): 211-217.

Goldstein DB, Linares AR, Cavalli-Sforza LL, *et al*. An evaluation of genetic distances for use with microsatellite loci. Genetics 139 (1): 463-471.

Goldstein DB, Pollock DD (1997). Launching microsatellites: A review of mutation processes and methods of phylogenetic inference. Journal Of Heredity 88 (5): 335-342.

Goudet J, Raymond M, De Meeüs T, Rousset F (1996) Testing differentiation in diploid populations. Genetics 144: 1933-1940.

Guo SW and Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportions for multiple alleles. Biometrics 48: 361-372.

Haldane JBS (1954) An exact test for randomness of mating. J. Genet. 52: 631-635.

Hall HG (1990) Parental analysis of introgressive hyridization between African and European honeybees using nuclear DNA RFLPs . Genetics 125: 611-621.

Harr B, Kauer M, Schlotterer C (2002). Hitchhiking mapping: A population-based fine-mapping strategy for adaptive mutations in *Drosophila melanogaster*. Proceedings of the National Academy of Sciences of the United States of America 99 (20): 12949-12954.

Henderson STand Petes TD (1992) Instability of simple sequence DNA in *Saccharomyces cerevisiae*. Mol. Cell Biol. 12: 2749-2757.

Huang QY, Xu FH, Shen H, *et al*. (2002). Mutation patterns at dinucleotide microsatellite loci in humans. American Journal of Human Genetics 70 (3): 625-634.

Jensen AB, Palmer KA, Boomsma JJ, Pedersen BV (2005). Varying degrees of *Apis mellifera ligustica* introgression in protected populations of the black honeybee*, Apis mellifera mellifera*, in northwest Europe. Molecular Ecology 14 (1): 93-106.

Jones AG, Rosenqvist E, Berglund A, *et al*. (1999). Clustered microsatellite mutations in the pipefish *Syngnathus typhle*. Genetics 152 (3): 1057-1063.

Kandemir İ and Kence A (1995) Allozyme variability in a central Anatolian honeybee (*Apis mellifera* L) population. Apidologie 26: 503-510.

Kandemir İ (1999) Genetic and morphometric variation in honeybee (*Apis mellifera* L) populations in Türkiye. Ph.D. dissertation, Middle East Technical University, Ankara, Türkiye.

Kandemir İ, Kence M, Kence A (2000) Genetic and morphometric variation in honeybee (*A. mellifera* L.) populations of Türkiye. Apidologie 31: 343-356.

Kruglyak S, Durrett RT, Schug MD, *et al*. (1998). Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. Proceedings of the National Academy of Sciences of the United States of America 95 (18): 10774-10778.

Kyle CJ, Strobeck C (2001). Genetic structure of North American wolverine (*Gulo gulo*) populations. Molecular Ecology 10 (2): 337-347.

Levinson G and Gutman GA (1987) High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. Nuc. Ac. Res. 15: 5323-5338.

Litt M, Luty JA (1989) A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am. J. Hu. Gen. 44: 397-401.

Morrison A, Johnson AL, Johnson LH, Sugino A (1993) Pathway correcting DNA replication errors in *Saccharomyces cerevisiae*. EMBO J. 12:1467-1473.

Nei M (1972) Genetic distances between populations. Am. Nat. 106: 283-292.

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci., USA 70: 3321-3323.

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89: 583-590.

Nei M (1983). Genetic diversity and the neutral mutation theory. Heredity 51 (Oct): 531-531.

Nunamaker RA, Wilson WT and Haley BE (1984) Electrophoretic detection of Africanized honeybees (*Apis mellifera scutellata*) in Guatemala and Mexico based on malete dehydrogenase allozyme patterns. J. Kans. Entomol. Soc. 57: 622-631.

Ostrander EA, Sprague GF, Rine J (1993). Identification and characterization of dinucleotide repeat (CA)n markers for genetic-mapping in dog. Genomics 16 (1): 207-213.

Ota T (1993) DISPAN (Genetic Distance and Phylogenetic Analysis) software. Pennsylvania State University.

Packer GJ, and Owen RJ (1992) Intra-articular dislocation of the patella, Arc. Emer. Med. 9: 244-245.

Paetkau D, Calvert W, Stirling I, Strobeck C (1995) Microsatellite analysis of population structure in Canadian polar bears. Mol. Ecol. 4: 347-354.

Paetkau D, Waits LP, Clarkson PL, Craighead L, Strobeck C (1997) An empirical evaluation of genetic distance statistics using microsatellite data from bear (Ursidae) populations. Genetics 147: 1943-1957.

Paetkau D, Waits LP, Clarkson PL, *et al*. (1998). Variation in genetic diversity across the range of North American brown bears. Conservation Biology 12 (2): 418-429.

Paetkau D, Shields GF, Strobeck C (1998). Gene flow between insular, coastal and interior populations of brown bears in Alaska. Molecular Ecology 7 (10): 1283-1292.

Palmer MR, Smith DR, Kaftanoğlu O (2000) Turkish honeybees: genetic variation and evidence for a fourth lineage of *Apis mellifera* mtDNA. The Journal of Heredity 91: 42-46.

Pamilo P, Varvio-Aho SL, Pekkarinen A (1978) Low levels of heterozygosity in Hymenoptera as a consequence of haplodiploidy. Hereditas 88: 93-99.

Primmer CR, Moller AP, Ellegren H (1996). New microsatellites from the pied flycatcher *Ficedula hypoleuca* and the swallow *Hirundo rustica* genomes. Hereditas 124 (3): 281-283.

Primmer CR, Saino N, Moller AP, *et al*. (1998). Unraveling the processes of microsatellite evolution through analysis of germ line mutations in barn swallows *Hirundo rustica*. Molecular Biology and Evolution 15 (8): 1047-1054.

Quezada-Euan JJG, May-Itza WD (2001). Partial seasonal isolation of African and European-derived *Apis mellifera* (Hymenoptera : Apidae) drones at congregation areas from subtropical Mexico. Annals of the Entomological Society Of America 94 (4): 540-544.

Raymond M and Rousset F (1994) GenePop ver. 3.0. Institut des Sciences de l'Evolution. Universite de Montpellier, France.

Raymond M and Rousset F (1995) An exact test for population differentiation. Evolution 49: 1280-1283.

Raymond M, Rousset F (1995). Genepop (Version-1.2) – Population genetics software for exact tests and ecumenicism. Journal of Heredity 86 (3): 248-249.

Reynolds J, Weir BS, Cockerham CC (1983) Estimation for the coancestry coefficient: basis for a short term genetic distance. Genetics 105: 767-779.

Richard M, Thorpe RS (2001). Can microsatellites be used to infer phylogenies? Evidence from population affinities of the Western Canary Island lizard (*Gallotia galloti*). Molecular Phylogenetics and Evolution 20 (3): 351-360.

Richardson BJ, Baverstock PR, Adams M (1986) Allozyme electrophoresis. Academic, New York.

Roderick GK (1996) Geographic structure of insect populations: gene flow, phylogeography, and their uses. Annu. Rev. Entomol. 41: 325-352.

Ruttner F (1988) Biogeography and Taxonomy of Honeybees. Springer-Verlag, Berlin Heidelberg, 284 p.

Ruttner F (1992) Naturgeschichte der honigbienen. München: Ehrenwirth, 357 p.

Sainudiin R, Durrett RT, Aquadro CF, *et al*. (2004). Microsatellite mutation models: Insights from a comparison of humans and chimpanzees. Genetics 168 (1): 383-395.

Saitou N and Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic tree. Mol. Biol. Evol. 4: 406-425.

Samadi S, Erard F, Estoup A, *et al*. (1988). The influence of mutation, selection and reproductive systems on microsatellite variability: a simulation approach. Genetical Research 71 (3): 213-222.

Savatier P, Trabuchet G, Faure C, Chebloune Y, Gouy M, Verdier G, Nigon VM (1985) Evolution of the primate beta-globin gene region. High rate of variation in CpG dinucleotides and in short repeated sequences between man and chimpanzee. J. Mol. Biol. 182: 21-29.

Schug MD, Wetterstrand KA, Gaudette MS, *et al*. (1998). The distribution and frequency of microsatellite loci in *Drosophila melanogaster*. Molecular Ecology 7 (1): 57-70.

Schlotterer C (2000). Evolutionary dynamics of microsatellite DNA. Chromosoma 109 (6): 365-371.

Schneider S, Roessli D, Excoffier L (2000) Arlequin ver. 2.000: A software for population genetics data analysis. Genetics and Biometry Laboratory, University of Geneva, Switzerland.

Sheppard WS and Berlocher SH (1984) Enzyme polymorphisms in *Apis mellifera mellifera* from Norway. J. Apic. Res. 23: 64-69.

Sheppard WS (1986) Genetic variation and differentiation in honeybees (*Apis*) Ph. D. Dissertation. University of Illinois at Urbana-Champaigni Illinois.

Sheppard WS, Arias MC, Meixner MD, Grech A (1997) *Apis mellifera ruttnerii*, a new honeybee subspecies from Malta. Apidologie 28: 287-293.

Sheppard WS and Smith DR (2000) Identification of African-derived bees in the Americas: a survey of methods. Ann. Entomol. Soc. Am. 93: 159-176.

Shrier MD, Jın L, Chakraborty R, *et al*. (1993). VNTR allele frequency-distributions under the stepwise mutation model - A computer-simulation approach. Genetics 134 (3): 983-993.

Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. Genetics 139: 457-467.

Slatkin M and Excoffier L (1996) Testing for linkage disequilibrium in genotypic data using the Expectation-Maximization algorithm. Heredity 76: 376-383.

Smith DR, Taylor OR, Brown WM (1989) Neotropical Africanized honey bees have African mitochondrial DNA. Nature 339: 213-215.

Smith DR (1991) Mitochondrial DNA and honeybee biogeography, 131-176. In DR Smith (eds.), Diversity in the genus *Apis*. Westview, Boulder, CO.

Smith DR, Slaymaker A, Palmer M, Kaftanoğlu O (1997) Turkish honeybees belong to the east Mediterranean lineage. Apidologie 28: 269-274.

Solignac M, Vautrin D, Loiseau A, Mougel F, Baudry E, Estoup A, Garnery L, Haberl M, Cornuet JM (2003). Five hundred and fifty microsatellite markers for the study of the honeybee (Apis mellifera L.) genome. Molecular Ecology Notes 3 (2): 307-311.

Strand M, Prolla TA, Liskay RM, Petes TD (1993) Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature 365: 274-276.

Suazo A, McTiernan R, Hall HG (1998) Differences between African and European honey bees (*Apis mellifera*) in random amplified polymorphic DNA (RAPD). J. Hered. 89: 32-36.

Suazo A and Hall HG (1999) Modification of the AFLP protocol applied to honey bee (*Apis mellifera*) DNA. Biotechniques 26: 704-705, 708-709.

Susnik S, Kozmus P, Poklukar J, *et al*. (2004). Molecular characterisation of indigenous *Apis mellifera carnica* in Slovenia. Apidologie 35 (6): 623-636.

Takezaki N and Nei M (1996). Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. Genetics 144 (1): 389-399.

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nuc. Ac. Res. 17: 6463-6471.

Tautz D (1993) Notes on the definition and nomenclature of tandemly repetitive DNA sequences. EXS 67: 21-28.

Tautz D, Wolff C, Schroeder R, *et al*. (1999). Evolution of insect segmentation. Developmental Biology 210 (1): 377.

Valdes AM, Slatkin M, Freimer NB (1993). Allele frequencies at microsatellite loci - The stepwise mutation model revisıted. Genetics 133 (3): 737-749.

Van Oppen MJH, Rico C, Turner GF, *et al*. (2000). Extensive homoplasy, nonstepwise mutations, and shared ancestral polymorphism at a complex microsatellite locus in Lake Malawi cichlids. Molecular Biology and Evolution 17 (4): 489-498.

Viard F, Franck P, Dubois MP, *et al*. (1998). Variation of microsatellite size homoplasy across electromorphs, loci, and populations in three invertebrate species. Journal of Molecular Evolution 47 (1): 42-51.

Vos P, Hogers R, Bleeker M, Reijans M, Van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nuc. Ac. Res. 23: 4407-4414.

Waser PM and Strobeck C (1998) Genetic signatures of interpopulation dispersal. TREE 13: 43-44.

Weber JL and May PE (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hu. Gen. 44: 388-396.

Weber JL and Wong C (1993). Mutation of human short tandem repeats. Human Molecular Genetics 2 (8): 1123-1128.

Weir BS and Cockerham CC (1984) Estimatinf F-statistics for the analysis of population structure. Evolution 38: 1358-1370.

Weir BS (1990) Genetic data analysis. Sinauer Publ., Sunderland, MA.

Xu X, Peng M, Fang Z, *et al*. (2000). The direction of microsatellite mutations is dependent upon allele length. Nature Genetics 24 (4): 396-399.

Zhivotovsky LA, Feldman MW, Grishechkin SA (1997). Biased mutations and microsatellite variation. Molecular Biology and Evolution 14 (9): 926-933.

Zhivotovsky LA (1999). A new genetic distance with application to constrained variation at microsatellite loci. Molecular Biology and Evolution 16 (4): 467-471.

APPENDIX A

SAMPLING LOCATIONS

PROVINCE	LOCATION	NUMBER OF SAMPLES
Eskişehir	Çifteler/Osmaniye	6
Eskişehir	Çifteler/Orhaniye	4
Eskişehir	Seyitgazi/Bardaklı	6
Eskişehir	Merkez	14
Artvin	Kayalar	6
Artvin	Efeler	6
Artvin	Düzenli	6
Artvin	Camili	13
Hakkari	Çengel	16
Hakkari	Merkez	20
Hatay	Yayladağı	16
Hatay	Reyhanlı	8
Hatay	Samandağ	7
Kırklareli	Çağlayık	31
Kastamonu	Evrenye	11
Kastamonu	Ahlat	5
Kastamonu	Benli Sultan	5
Kastamonu	Azdavay	9
Cyprus	Güzelyurt	15
Cyprus	Karaağaç	12
Ardahan/Posof	Süngülü	10
Ardahan/Posof	Yeniköy	5
Ardahan/posof	Alköy	6
Izmir/Karaburun	Merkez	24
Muğla	Merkez/İkizce	3
Muğla	Merkez/Yaraş	2
Muğla	Bodrum/Dereköy	3
Muğla	Bodrum/Gümüşlük	2
Muğla	Marmaris/Aspiran	2
Muğla	Marmaris/Bayır	3
Muğla	Marmaris/Çamlı	2
Muğla	Milas/Akçalı	2
Muğla	Milas/Bafa	3
Muğla	Milas/Derince	3
Muğla	Milas/Karakuyu	2
Muğla	Ula/Elmalı	2
Muğla	Ula/Karabörtlen	3
Urfa	Akçakale	15
Urfa	Halfeti	8
Urfa	Bozova	7
Ankara/Beypazarı	Merkez (5 different villages)	26

APPENDIX B

LIST OF REAGENTS

REAGENT	BRAND NAME	CATALOGUE NUMBER
Acrylamide/bis-Acrylamide	Sigma	A-2917
Ammonium Persulfate	Sigma	A-9164
Autoradiography film	Kodak Biomax MR-2	Z35
Bovine Serum Albumin	MBI Fermentas	B14
Chloroform	Merck	
Dithiothreitol	Sigma	D-9779
EDTA	AppliChem	A2937
DNTP set	MBI Fermentas	RO181
Formamide	AppliChem	A2156
Isoamyl alcohol	Sigma	I9392
Lauryl Sulfate	Sigma	L4390
$MgCl_2$	MBI Fermentas	
PCR Buffer	MBI Fermentas	
Phenol-Cloroform-Isoamyl alc.	AppliChem	A0889
Primers	IDT	
Sigmacote	Sigma	SL2
Sodium Chloride	Sigma	S3014
Taq DNA Polymerase	MBI Fermentas	EP0402
TBE Buffer	Sigma	T-4415
TEMED	Sigma	T-7024
Tris	Sigma	T1378
Urea	AppliChem	A1049,5000

APPENDIX C

LIST OF EQUIPMENT

EQUIPMENT	BRAND NAME	MODEL
Centrifuge	Eppendorf	5415R
Exposure cassette	Sigma	E9510
Gel drying system	E-C	EC356
Ph meter	Eutech	Cyberscan 500
Sequencing apparatus	Owl	S4S
Thermocycler	Techne	HL-1

APPENDIX D

COMPOSITIONS OF SOLUTIONS

Table 1. Preparation of Wilson buffer

Add the followings:
10 ml from 1 M Tris.Cl pH 8 stock solution
200 µl from 0.5 M Ethylenediaminetetraacetic acid (EDTA) stock solution
1 ml from 10% (w/v) Lauryl sulfate (SDS) stock solution
0.771 g of Dithiothreitol (DTT)
0.584 g of Sodiumchloride (NaCl)
Add distilled water to complete to 100 ml.

Table 2. Six percent acrylamide / urea solution

75 ml from 40% acrylamide solution
50 ml from 10x TBE
240 g from urea
Adjust the volume to 500 ml by distilled water.

Table 3. Loading buffer for polyacrylamide gel electrophoresis

Formamid	10 ml
Xylene cyanol FF	10 mg
Bromophenol blue	10 mg
0.5 M EDTA (pH=8)	200 µl

APPENDIX E

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Bodur, Çağrı
Nationality: Turk (TC)
Date and Place of Birth: 13 December 1976 , İzmir
Marital Status: Single
Phone: +90 312 210 51 87
Fax: +90 312 210 12 89
email: cagribodur@gmail.com

EDUCATION

Degree	Institution	Year of Graduation
MS	METU Biology	2001
BS	METU Biology	1999
High School	Selçuk High School, İzmir	1993

WORK EXPERIENCE

Year	Place	Enrollment
1999- Present	METU Department of Biological Sciences	Research Assistant
1998 August	Hıfzısıhha Enstitüsü	Intern Biology Student

FOREIGN LANGUAGES

Advanced English

PUBLICATIONS

Bodur Ç (2001). Microsatellite analysis in honeybee populations of Turkey. M.Sc. thesis, Middle East Technical University, Ankara, Türkiye.

Bodur Ç, Kence M, Akkaya M, Kence A (2002). Türkiye'deki balarısı populasyonları arasında mikrosatelit lokusları bakımından farklılaşmalar. XVI. Ulusal Biyoloji Kongresi, Malatya, Türkiye. Özetler: 76.

Bodur Ç, Kence M, Akkaya M, Kence A (2003). Microsatellite variation in honeybee (*Apis mellifera* L.) populations of Turkey. XIX. International Congress of Genetics, Melbourne, Australia. Abstracts: 150.

Bodur Ç, Kence M, Kence A (2004). Genetic structure and origin determination in honeybee populations of Anatolia. Proceedings of the First European Conference of Apidologie: 40. Udine, Italy.

HOBBIES

Tennis, philosophy, movies, photography, swimming