

T.R.
GEBZE TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

DNA BARCODING STUDY OF TURKISH WILD CHERRY
(*Prunus avium* L.) POPULATIONS BELONGING TO
ADAPAZARI AND BOLU FORESTRY DIRECTORATES

SELİN GÜL ÜNSAL
A THESIS SUBMITTED FOR THE DEGREE OF
MASTER OF SCIENCE
DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS

GEBZE
2018

T.R.
GEBZE TECHNICAL UNIVERSITY
GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

DNA BARCODING STUDY OF TURKISH
WILD CHERRY (*Prunus avium* L.)
POPULATIONS BELONGING TO
ADAPAZARI AND BOLU FORESTRY
DIRECTORATES

SELİN GÜL ÜNSAL
A THESIS SUBMITTED FOR THE DEGREE OF
MASTER OF SCIENCE
DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS

THESIS SUPERVISOR
PROF. DR. YELDA ÖZDEN ÇİFTÇİ

GEBZE

2018

T.C.
GEBZE TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ADAPAZARI VE BOLU ORMAN
MÜDÜRLÜKLERİNE AİT YABANİ KİRAZ
(*Prunus avium* L.) POPULASYONLARININ
DNA BARKODLAMA ÇALIŞMASI

SELİN GÜL ÜNSAL
YÜKSEK LİSANS TEZİ
MOLEKÜLER BİYOLOJİ VE GENETİK ANABİLİM DALI

DANIŞMANI
PROF. DR. YELDA ÖZDEN ÇİFTÇİ

GEBZE
2018



YÜKSEK LİSANS JÜRİ ONAY FORMU

GTÜ Fen Bilimleri Enstitüsü Yönetim Kurulu'nun ...10.../...01.../2018. tarih ve 2018.../...03... sayılı kararıyla oluşturulan jüri tarafından 29/01/2018 tarihinde tez savunma sınavı yapılan Selin Gül ÜNSAL 'ın tez çalışması Moleküler Biyoloji ve Genetik Anabilim Dalında YÜKSEK LİSANS tezi olarak kabul edilmiştir.

JÜRİ

ÜYE

(TEZ DANIŞMANI) : PROF. DR. YELDA ÖZDEN ÇİFTÇİ

ÜYE

: DOÇ. DR. YILDIZ AYDIN

ÜYE

: YRD. DOÇ. DR. BAHAR YILDIZ KUTMAN

ONAY

Gebze Teknik Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulu'nun
...../...../..... tarih ve/..... sayılı kararı.

SUMMARY

Plant of the species wild cherry (*Prunus avium*) is one of the most important European hardwood species in the family Rosaceae. Wild cherry trees are of great ecological and naturalistic importance, being an essential component of sensitive and threatened ecosystems. Thus, in the frame of this thesis study, scientific investigations have been carried out with a total of 139 leaves of wild cherry (*Prunus avium* L.) trees that were collected from 7 different locations (Gölcük, Düzce, Melen, Yedigöller, Abant, Mollafenari and Kefken regions) located under the jurisdiction of Adapazarı and Bolu Forestry Directorates (Düzce, Bolu and Kocaeli). For this aim, DNA barcoding method, which is proposed as a universal tool based on DNA, was used to determine not only the most useful barcode genes but also the genetic variation among these plant samples. 4 DNA barcode genes (rbcL, trnH-psbA, matK and ITS region) of 139 individuals from 3 different populations (Düzce, Bolu and Kocaeli) were sequenced and single nucleotide polymorphism analysis was conducted. According to PCR amplification results, barcode genes of trnH-psbA gene and ITS showed 100% amplification success while matK was 89.93 %. Using these sequence information, DNA barcodes were obtained for 6 haplotypes for trnH-psbA gene, 18 haplotypes for matK gene, 29 haplotypes for ITS spacer. Also, sequence data of 139 individuals obtained from related barcode genes was analyzed by using bioinformatic tools and biostatistics programs. When all the results were evaluated together, ITS region and matK was found to be more effective than the other tested single locus barcode genes for determination of the genetic differences of tested species of *P. avium*. Moreover, DNA sequence data of haplotypes were also registrated into GenBank database.

Key Words: Plant DNA barcode, Wild cherry, *trnH-psbA*, *matK*, *ITS*.

ÖZET

P. avium (yabani kiraz) türü bitki, Rosaceae familyasından, Avrupa’da sert kereste üretiminde kullanılan en önemli türdür. Hassas ve tehdit altındaki ekosistemlerin vazgeçilmez bir parçası olan yabani kiraz ağaçları büyük bir ekolojik öneme sahiptir. Bu nedenle tezde, Adapazarı ve Bolu Orman Bölge Müdürlükleri’nin yetki alanlarındaki (Düzce, Bolu and Kocaeli) 7 ayrı lokasyondan (Gölcük, Düzce, Melen, Yedigöller, Abant, Mollafenari ve Kefken) toplanan 139 adet yabani kiraz ağacının (*Prunus avium* L.) yaprakları ile bilimsel araştırmalar gerçekleştirilmiştir. Bu amaçla DNA’ya dayalı evrensel bir araç olarak önerilen DNA barkodlama yöntemi, hem *P. avium* için en kullanışlı barkod genlerinin kullanımının belirlenmesinde, hem de bu bitki örnekleri arasındaki genetik varyasyonun saptanmasında kullanılmıştır. 3 farklı popülasyondan (Düzce, Bolu ve Kocaeli) toplamda 139 bireyin 4 DNA barkod geni (rbcL, trnH-psbA, matK ve ITS) dizilenmiş ve tek nükleotid polimorfizm analizleri gerçekleştirilmiştir. PCR amplifikasyon sonuçlarına göre, trnH-psbA ve ITS barkod genleri %100 amplifikasyon başarıları gösterirken, matK için amplifikasyon başarıları %89,93’tür. Bu sekans bilgileri kullanılarak, trnH-psbA geni için 6, matK geni için 18, ITS için ise 29 haplotip elde edilmiştir. Ayrıca, ilgili barkod genlerinden elde edilen 139 bireye ait dizi verileri biyoinformatik araçlar ve biyoistatistik programlar kullanılarak analiz edilmiştir. Bütün sonuçlar birlikte değerlendirildiğinde, *P. avium* türlerinin genetik farklılıklarının belirlenmesi için ITS bölgesi ve matK, diğer test edilen tek lokus barkod genlerinden daha etkili bulunmuştur. Bu tez çalışmasının sonucunda haplotiplerin DNA dizi verileri GenBank veritabanına kaydedilmiştir.

Anahtar Kelimeler: Bitki DNA Barkodları, Yabani Kiraz, *trnH-psbA*, *matK*, *ITS*.

ACKNOWLEDGEMENTS

I am grateful to my esteemed supervisor Prof. Dr. Yelda ÖZDEN ÇİFTÇİ who has enlightened me with her knowledge throughout my education. She has always supported me in particular during my studies in Italy and also in Turkey, with her trust in me.

I would like to thank to my precious Italian advisor Prof. Dr. Antonella CANINI who gave me a chance to be a member of her laboratory. I offer my deepest respects for all the opportunities that she has given me.

I would like to express my deepest sense to Assist. Prof. Angelo GISMONDI who is the unique chance in my academic career and also in my life. I am thankful to him for his encouragement and insightful comments, which helped me for every step of this thesis. He gave me endless possibilities in their laboratory and made this thesis possible with his valuable knowledge. During the realization of this thesis, he was my guide and he will no doubt be forever.

My very sincere thanks to my Italian family members who offered their continuous advice, sibling-like support and the home feeling throughout the course of this thesis, Dr. Gabriele DI MARCO, Dr. Lorena CANUTI, Valentina NANNI, Silvia DE ROSSI, Silvia NOVELLI and Alessia D'AGOSTINO who are the best and most unique laboratory team in Dipartimento di Biologia dell'Università degli Studi di Roma "Tor Vergata".

I would like to thank to Ercan VELİOĞLU who is the manager of General Directorate of Forestry, Poplar and Fast Growing Forest Trees Research Institute, to support me on sampling stage of cherry populations. I would like to thank to Burcu UZAN, Dr. Ezgi ÇABUK ŞAHİN, Dr. Emrah KIRDÖK and Emre BELGE for their valuable contributions to my thesis.

I am thankful to Hamit EKİNCİ for his selfless support, his infinite confidence in me and his flawless editing of my thesis. I thank him for make me feel his presence every moment.

Finally, I take this opportunity to express the profound gratitude from my deep heart to my beloved grandparents, Arif NASIR and Emel NASIR; my uncle Murat NASIR; my beloved parents, Mehmet ÜNSAL and Dilek ÜNSAL, and my brother, Alper ÜNSAL for their love and continuous support.

TABLE OF CONTENTS

	<u>Page</u>
SUMMARY	v
ÖZET	vi
ACKNOWLEDGMENTS	vii
TABLE of CONTENTS	viii
LIST of ABBREVIATIONS and ACRONYMS	x
LIST of FIGURES	xii
LIST of TABLES	xv
1. INTRODUCTION	1
1.1. Aim, Contribution and Scope of the Thesis	2
2. LITERATURE REVIEW	4
2.1. Wild Cherry (<i>Prunus avium</i> L.)	4
2.1.1. Distribution of Wild Cherry	6
2.1.2. Usage of Wild Cherry	8
2.2. DNA Barcoding	9
2.2.1. The History of DNA Barcoding	9
2.2.2. The Technique of DNA Barcoding	10
2.2.3. DNA Barcode Genes	14
2.2.3.1. DNA Barcode Genes Peculiar to Chloroplast Genome	14
2.2.3.1.1. <i>rbcL</i>	14
2.2.3.1.2. <i>trnH-psbA</i>	15
2.2.3.1.3. <i>matK</i>	16
2.2.3.2. DNA Barcode Genes Peculiar to Nuclear Genome	17
2.2.3.2.1. <i>ITS</i>	17
2.2.3.3. Other DNA Barcode Genes	18
2.3. Phylogenetic Analysis	19
2.3.1. Neighbor Joining (NJ)	21
2.2.3. Maximum Likelihood (ML)	22
2.4. Statistical Analysis	22

2.4.1. Principal Component Analysis (PCA)	22
2.4.2. Structure Analysis	23
2.5. Previous DNA Barcoding Studies	24
3. MATERIAL AND METHOD	30
3.1. Plant Material	30
3.2. The Used Equipments	32
3.3. DNA Extraction	33
3.4. Agarose Gel Electrophoresis for confirmation of DNA Extraction	34
3.5. PCR Amplifications and Agarose Gel	34
3.6. Sequencing	36
3.7. Data Analysis and Phlogenetic Tree Constraction	36
3.7.1. Principal Component Analysis (PCA)	37
3.7.2. Structure Analysis	39
4. RESULTS	42
4.1. DNA Extraction and PCR Reaction	42
4.2. DNA Sequencing (Barcoding) and Statistical Analysis	44
4.3. Phylogenetic Analysis	58
4.4. Principle Component Analysis (PCA)	68
4.5. Structure Anaysis	72
5. DISCUSSION	79
5.1. <i>trnH-psbA</i>	80
5.2. <i>matK</i>	81
5.3. <i>ITS</i>	82
5.4. Multilocus (<i>trnH-psbA+matK+ITS</i>)	83
6. CONCLUSION	85
REFERENCES	87
BIOGRAPHY	99
APPENDICES	
CD	

LIST OF ABBREVIATIONS AND ACRONYMS

<u>Abbreviations and Acronyms</u>	<u>Explanations</u>
%	: Percent
Δ	: Delta
°C	: Centigrade Degree
μL	: Microliter
μM	: Micromolar
A	: Adenine
accD	: Subunit β of carboxyl transferase acetyl coenzyme A carboxylase
AFLP	: Amplified Fragment Length Polymorphism
BOLD	: The Barcode of Life Data System
bp	: Base Pairs
BSA	: Bovine Serum Albumin
C	: Cytosine
CBOL	: The Consortium for the Barcode of Life
chDNA	: Chloroplast DNA
cm	: Centimeter
COI	: Cytochrome oxidase I
DNA	: Deoxyribonucleic acid
dNTP	: Deoxynucleotide
EDTA	: Ethylene diamine tetra acetic acid
F	: Forward
G	: Guanine
H	: Haplotype
ISSR	: Inter Simple Sequence Repeat
IUCN	: The International Union for Conservation of Nature
ITS	: Internal transcribed spacer
Kb	: Kilobase
kDa	: Kilodalton
M	: Molar
m	: Meter

matK	: Maturase K
MEGA 7	: Molecular Evolutionary Genetics Analysis version 7.0
mg	: Miligram
min	: Minute
ML	: Maximum Likelihood
mM	: Milimolar
NCBI	: The National Center for Biotechnology Information
ndhF	: NADH dehydrogenase F
ndhJ	: NADH-plastoquinone oxidoreductase subunit J
nrDNA	: Nuclear ribosomal DNA
NJ	: Neighbor Joining
PCA	: Principal component analysis
PCR	: Polymerase chain reaction
R	: Reverse
rbcL	: Ribulose-bisphosphate carboxylase
RAPD	: Random Amplification of Polymorphic DNA
RFLP	: Restriction Fragment Length Polymorphism
RNA	: Ribonucleic acid
rpm	: Revolutions per minute
rpl36-rps8	: 60S ribosomal protein L36-ribosomal Protein S8
rpl16	: Ribosomal Protein S116
rpoB	: beta subunit of RNA polymerase
rps4-trnS	: Small ribosomal protein 4
rRNA	: Ribosomal ribonucleic acid
s	: Second
SNP	: Single nucleotide polymorphisms
UPGMA	: Unweighted Pair Group Method with Arithmetic
T	: Tymine
TAE	: Tris-acetate-EDTA
U	: Unit
UV	: Ultraviolet
STR	: Short Tandem Repeat
ORF	: Open reading frame

LIST OF FIGURES

<u>Figure No:</u>	<u>Page</u>
2.1: Images of <i>Prunus avium</i> .	6
2.2: Distribution map of <i>P. avium</i> (wild cherry).	7
2.3: Map estimating the maximum habitat suitability.	8
2.4: The barcoding pipeline.	11
2.6: The DNA Barcoding Locus.	13
2.7: PCR priming site within highly conserved flanking coding sequence combine with a non-coding region.	15
2.8: Diagram showing <i>matK</i> gene nested between <i>trnK</i> introns.	16
2.9: Priming site of <i>ITS</i> .	18
3.1: Distribution area of collected leaves.	31
3.2: Sampling by help of long pruning shear.	31
3.3: PCA analysis of <i>P. avium</i> haplotypes by using the data of the barcode genes (<i>trnH-psbA</i> , <i>matK</i> and <i>ITS</i>) used in the study by using “GenAlEx 6.4” programme.	38
3.4: The structure analysis of <i>P. avium</i> haplotypes by using the data of the barcode genes (<i>trnH-psbA</i> , <i>matK</i> and <i>ITS</i>) used in the study by the help of Structure 2.3.4." and "Structure Harvester" programs.	40
4.1: A representative photograph of extracted DNAs of 20 different individuals of Abant population.	42
4.2: A representative figure of amplified sequences of <i>trnH-psbA</i> gene of Düzce population.	43
4.3: A representative figure of amplified sequences of <i>matK</i> gene of Yedigöller population.	43
4.4: A representative figure of amplified sequences of <i>ITS</i> spacer of Mollafenari population.	43
4.5: A representative screen display of electropherogram obtained as a result of DNA sequence analysis with BioEdit program.	44
4.6: A representative screen display of the sequence data obtained as a result of DNA sequence analysis with BioEdit program.	45

4.7:	Graphical representation of 139 <i>P. avium</i> + <i>P. armeniaca</i> (outgroup) for DNA barcode of <i>trnH-psbA</i> .	54
4.8:	Graphical representation of 139 <i>Prunus avium</i> + <i>P. armeniaca</i> (outgroup) for DNA barcode of <i>matK</i> .	55
4.9:	Graphical representation of 139 <i>P. avium</i> + <i>P. armeniaca</i> (outgroup) for DNA barcode of <i>ITS</i> .	56
4.10:	Graphical representation of 139 <i>P. avium</i> + <i>P. armeniaca</i> (outgroup) for DNA barcode of multilocus <i>trnH-psbA+matK+ITS</i> .	57
4.11:	Bootstrap consensus tree generated by the Neighbor Joining NJ method for the <i>trnH-psbA</i> sequences obtained for tested <i>P. avium</i> populations here studied.	60
4.12:	Bootstrap consensus tree generated by the ML method for the <i>trnH-psbA</i> sequences obtained for tested <i>P. avium</i> populations here studied.	61
4.13:	Bootstrap consensus tree generated by the NJ method for the <i>matK</i> sequences obtained for tested <i>P. avium</i> populations here studied.	62
4.14:	Bootstrap consensus tree generated by the ML method for the <i>matK</i> sequences obtained for tested <i>P. avium</i> populations here studied.	63
4.15:	Bootstrap consensus tree generated by NJ method for the <i>ITS</i> sequences obtained for <i>P. avium</i> populations here studied.	64
4.16:	Bootstrap consensus tree generated by ML method for the <i>ITS</i> sequences obtained for <i>P. avium</i> populations here studied.	65
4.17:	Bootstrap consensus tree generated by NJ Method for the <i>trnH-psbA+matK+ITS</i> combined sequences obtained for <i>P. avium</i> populations here studied.	66
4.18:	Bootstrap consensus tree generated by ML method for the <i>trnH-psbA+matK+ITS</i> combined sequences obtained for <i>P. avium</i> populations here studied.	67
4.19:	PCA analysis of <i>P. avium</i> L. haplotypes of the <i>trnH-psbA</i> gene obtained from "GenAlEx 6.4" program.	68
4.20:	PCA analysis of <i>P. avium</i> L. haplotypes of the <i>matK</i> gene obtained from "GenAlEx 6.4" program.	69

4.21:	PCA analysis of <i>P. avium</i> L. haplotypes of the <i>ITS</i> gene obtained from "GenAlEx 6.4" program.	70
4.22:	PCA analysis of <i>P. avium</i> L. haplotypes of multilocus (<i>trnH-psbA+matK+ITS</i>) obtained from "GenAlEx 6.4" program.	71
4.23:	SNP analysis of representative barcode gene (<i>ITS</i>) which used in the study by "Mega 7" program.	72
4.24:	Estimation of L(K) for each K value and population structure for K=5 with samples of <i>P. avium</i> for <i>trnH-psbA</i> barcoding gene.	73
4.25:	Estimation of L(K) for each K value and population structure for K=2 with samples of <i>P. avium</i> for <i>matK</i> barcoding gene.	74
4.26:	Estimation of L(K) for each K value and population structure for K=2 with samples of <i>P. avium</i> for <i>ITS</i> barcoding gene.	75
4.27:	Estimation of L(K) for each K value and population structure for K=3 with samples of <i>P. avium</i> for multilocus (<i>trnH-psbA+matK+ITS</i>) barcoding gene.	76

LIST OF TABLES

<u>Table No:</u>	<u>Page</u>
2.1: Taxonomic classification of <i>P. avium</i> L	4
2.2: Genes from three genomes in plants that are candidate barcodes.	13
2.3: Gene/locus selected for some of the study of DNA barcoding in plants.	27
2.4: Some of the molecular marker studies with <i>Prunus avium</i> .	29
3.1: Growing regions, population and subpopulation names and altitudes on the map, coordinates and individual numbers of collected plants.	30
3.2: The list of the devices used this thesis.	32
3.3: The list of the kits and chemicals used this thesis.	33
3.4: i-Genomic Plant DNA Extraction Kit contents	34
3.5: Primer pair names, their length of nucleotide, their sequences, target gene names and Tm°C used in PCRs.	35
4.1: The success rates of DNA barcodes within the investigated data-set.	45
4.2: Molecular diversity parameters for <i>trnH-psbA</i> gene.	46
4.3: Molecular diversity parameters for <i>matK</i> gene.	46
4.4: Molecular diversity parameters for <i>ITS</i> gene.	47
4.5: Percentage of variable sites of aligned sequences.	47
4.6: Haplotype composition of the seven wild populations of <i>P. avium</i> for <i>trnH-psbA</i> gene.	48
4.7: Haplotype composition of the seven wild populations of <i>P. avium</i> for <i>matK</i> gene.	48
4.8: Haplotype composition of the seven wild populations of <i>P. avium</i> for <i>ITS</i> gene.	49
4.9: Percentages of the frequency of the individual nitrogen bases (A, G, C and T) for amplified <i>trnH-psbA</i> gene sequence.	50
4.10: Percentages of the frequency of the individual nitrogen bases (A, G, C and T) for amplified <i>matK</i> gene sequence.	51
4.11: Percentages of the frequency of the individual nitrogen bases (A, G, C and T) for amplified <i>ITS</i> spacer sequence.	52

1. INTRODUCTION

Plant taxonomy is a functional science concerned with the identification, nomenclature and classification of plants all over the world [Prajapat and Chakraborty, 2011]. There are 391,000 specified plant species in the world [IUCN, 2016] but just few of them can be identified based on traditional plant identification methods [Li et al., 2015]. In addition to traditional methods, many molecular marker methods such as RFLP (Restriction Fragment Length Polymorphism), RAPD (Random Amplified Polymorphic DNA), AFLP (Amplified Fragment Length Polymorphism), SSR (Simple Sequence Repeats), ISSR (Inter-Simple Sequence Repeats), SNP (Single Nucleotide Polymorphism) and STR (Short Tandem Repeats) are used for the identification and classification of plant species. Beside these molecular markers, the usage of DNA barcoding which is a technique for identification of species conducted by using a short DNA sequences from agreed-upon position in the genome has had a positive impact on identification of plant species and their taxonomic clarification [Gregory, 2005]. It may be one locus or several loci used together as a complementary unit.

DNA barcoding is used for a wide range of purpose including;

- contribution to taxonomy at the species level [Hajibabaei et al., 2007], [Hebert et al., 2004];
- contribution to population genetics and phylogenetic studies [Hajibabaei et al., 2007], [Hebert et al., 2004];
- identification unknown species [Hajibabaei et al., 2007], [Hebert et al., 2004],
- identification of taxonomically problematic species;
- determination of cryptic species [Hebert et al., 2004];
- contribution to ecological and environmental genomic studies [Valentiniet al., 2009];
- clarification of forensic events [Yoon, 1993], [Coyle et al., 2005], [Mildenhall, 2006].

The technique of DNA barcoding offers taxonomists the great opportunity to expand of global life's diversity [Prajapat and Chakraborty, 2011]. In literature, the classification and identification studies of many plant species such as *Cyperaceae*,

Aspalathus L., Cycadales, *Crocus sativus* L., *Picea abies*, *Olea europea* L., *Cornus mas* L., *Castanea sativa* (Mill.), *Bidens pilosa*, *Daucus carota*, *Crataegus monogyna*, *Hedera helix*, *Mahonia aquifolium*, *Brassica oleracea*, *Viburnum tinus*, *Ligustrum vulgare*, *Prunus cerasifera*, *Prunus laurocerasus* have been verified by using DNA barcoding technique.

Wild cherry which distributes mostly through Black Sea area in Turkey is a valuable broadleaved tree species especially for wild life [Praciak, et al., 2013], [Scholz, 1995]. Therefore, 7 different subpopulations (Abant, Mollafenari, Yedigöller, Düzce, Melen, Kefken and Gölcük) from 3 different populations (Düzce, Bolu and Kocaeli) of wild cherry was used as plant materials in this thesis.

1.1. Aim, Contribution and Scope of the Thesis

One of the main problems faced by biologists today is the taxonomic impediment; there are simply too many species, but too few taxonomists to discover, describe and identify all the specimens that are collected [Bleeker et al., 2008], [Chown et al., 2008]. DNA barcoding that is promoted by Hebert (2003) to identify and delimit species has been proposed as a remedy for resolving the taxonomic impediment. DNA lets to important findings about species distinction and the descent of populations within species. Today, the distance-based method has a top quality for DNA barcoding. In our country, biology and genetics of wild cherry has been neglected. Comprehensive studies with wild cherry which has valuable wood and fruit would benefit and enhance this species' productivity and its economical inputs.

The aims of the thesis are:

- to increase the genetic knowledge of wild cherry species;
- to determine the barcode genes which are present in wild cherry genome.
- to register obtained nucleotide sequences into the scientific databases;
- to promote the application of the barcoding technique for new and further studies to shed light on *Prunus avium* phylogeny, history, geographical distribution and origin.

- to demonstrate the utility of DNA barcode genes for genetic fingerprinting in wild cherry.
- to contribute to conservation and breeding studies by examining the genetics of wild cherry.

According to our knowledge, there is no scientific study has been done on DNA barcoding study of *P. avium* species in Turkey so far. Thus, this thesis is original on its subject.

This study focuses on the collection of a total 139 leaf samples from 7 different subpopulations (Abant, Mollafenari, Yedigöller, Düzce, Melen, Kefken, Golcuk) belonging to 3 populations of wild cherries (Kocaeli, Düzce, Bolu) and investigation of the sequences of three different plastidial and one nuclear barcode genes. In the scope of the thesis, the populations that are sampled will be examined, contributed to the conservation and breeding program of the wild cherry tree which is economic and ecological value, contributed to increase of the genetic knowledge of wild cherry species.

2. LITERATURE REVIEW

2.1. Wild Cherry (*Prunus avium* L.)

Wild cherry (*Prunus avium* L.) is an out breeding, self-incompatible member of the family *Rosaceae*, subfamily *Amygdaloideae* (*Prunaceae*) and it is a diploid with a chromosome number of $2n=2x=16$ (Table 2.1). This species is one of the important European tree species as it is used in woodworks [Rakonjac et al., 2014], [Hummer et al., 2009].

Table 2.1: Taxonomic classification of *P. avium* L.

TAXONOMIC CLASSIFICATION	
Kingdom	Plantae
Subkingdom	Viridiplantae
Infrakingdom	Streptophyta - land plants
Superdivision	Embryophyta
Division	Tracheophyta - vascular plants, tracheophytes
Subdivision	Spermatophytina - spermatophytes, seed plants, Phanérogames
Class	Magnoliopsida - Dicotyledons
Superorder	Rosanae
Order	Rosales
Familiy	Rosaceae - roses
Genus	<i>Prunus</i> L. - chokecherry, plum
Species	<i>Prunus avium</i> L. - sweet cherry, wild cherry

Wild cherry is a fast-growing tree, with strong apical growth. It has a distinctive smooth, reddish or purple-brown bark. The bark has large lenticels which often peels horizontally in thin strips. The leaves change color from light green in spring over dark green in summer, and to yellow, orange-red, scarlet or pink in autumn. They are alternate, pendulous, simple and elliptic-ovate to obovate acute in shape. Stem wounds produce a resin-like, amber colored odorless gum. It is one of the first trees to flower

in the spring and produces masses of white blossom. It matures at around 60-80 years, when trees are typically 20-25 m in height with trunks of 50-70 cm in diameter. Wild cherry usually lives between 70-150 years [Praciak, 2013].

It prefers slightly acidic conditions. Wild cherry is frequently found on woodland edges and glades. It is essentially a pioneer species, and quickly colonises clearings by seeds. Vegetative regeneration from both stumps and roots after coppicing is strong. It is frequently found growing as a minor component of oak, ash and beech woodlands. Edible varieties are usually grafted onto wild cherry root stocks [Russel, 2003], [Hummer et al., 2009], [Scholz, 1995].

The species which is valued for wildlife has hermaphroditic, actinomorphic and white flowers which are about 2-2.5 cm in diameter, usually pollinated by a range of insects, mainly wild bees, honeybees and bumblebees. Outcrossing is the principal mating system, with self-fertilization usually controlled by a multi-allelic locus S, with gametophytic incompatibility [Vaughan et al., 2008], [Spiecker and Hein et al., 2009]. The images of seeds, buds, trunks, leaves, fruits and flowers of *Prunus avium* are shown in Figure 2.1. Flowering and seed production starts from about four years of age under optimal conditions. To provide the chilling requirement necessary for flower induction, the species is commonly grown in the temperate climatic zones with cooler temperatures [Russel, 2003], [Scholz, 1995], [Praciak et al., 2013]. Its edible fruits are small and red or black in color. Fruit ripening takes place from late spring until summer and are consumed or dispersed mostly by gravity. In addition to gravity, birds especially pigeons, starlings, thrushes and jays, and also larger vertebrates like foxes, badgers or wild boar play an important role in their dissemination [Ducci, 2013]. The species can also be propagated through vegetative reproduction by root suckering [Vaughan et al., 2007]. For the vegetative propagation of ornamental cultivars, fruit producing varieties, and selected genotypes; grafting techniques are commonly applied [Scaltsoyiannes et al., 2009].

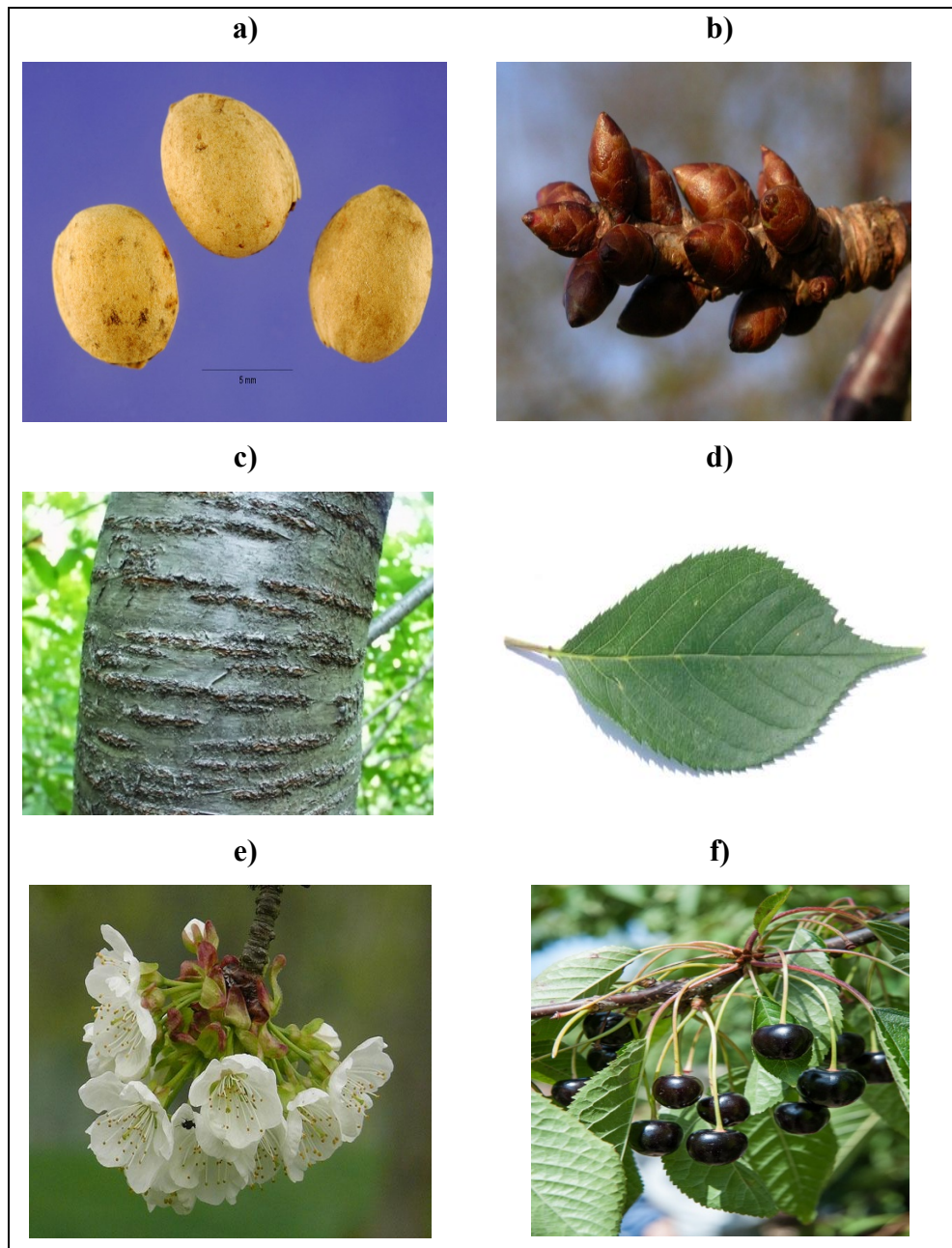


Figure 2.1: Images of *Prunus avium*. a) Seeds, b) Buds, c) Trunk, d) Leaf, e) Flowers, f) Fruits.

2.1.1. Distribution of Wild Cherry

The natural range of wild cherry, usually characterized by a scattered distribution, extends northwards to the British Islands and to Southern Scandinavia. The northern natural range limit is reached at about 55° north parallel. In the south, the range extends to North Africa, South Spain, central Italy, and the Balkans. It is strictly confined to increasingly scattered high-altitude populations in humid climate

in these regions. The eastern part of the natural range includes the Caucasus and the Elburs mountains of North Iran, where the species is reported to grow up to elevations of 2000m. It occurs up to 1700m in the Alps and reaches 1900m in South-East France. *Prunus* is naturally grown successfully in deciduous forest habitats and shrub land, especially in temperate regions of Northern Asia and North America [Russell, 2003], [Rogatis et al., 2013], [Praciak, et al., 2013].

Wild cherry species are abundant in the North, North East and North West provinces of Turkey [Ercisli, 2004]. Distribution map of *P. avium* is given in Figure 2.2 [Welk et al., 2016]. *P. avium* is distributed through Europe countries and the north side of Turkey-especially the north eastern part. Estimated map for the maximum habitat suitability is also given in Figure 2.3 [Welk et al., 2016]. While blue color represents high survivability of *P. avium*, yellow color represents low survivability.

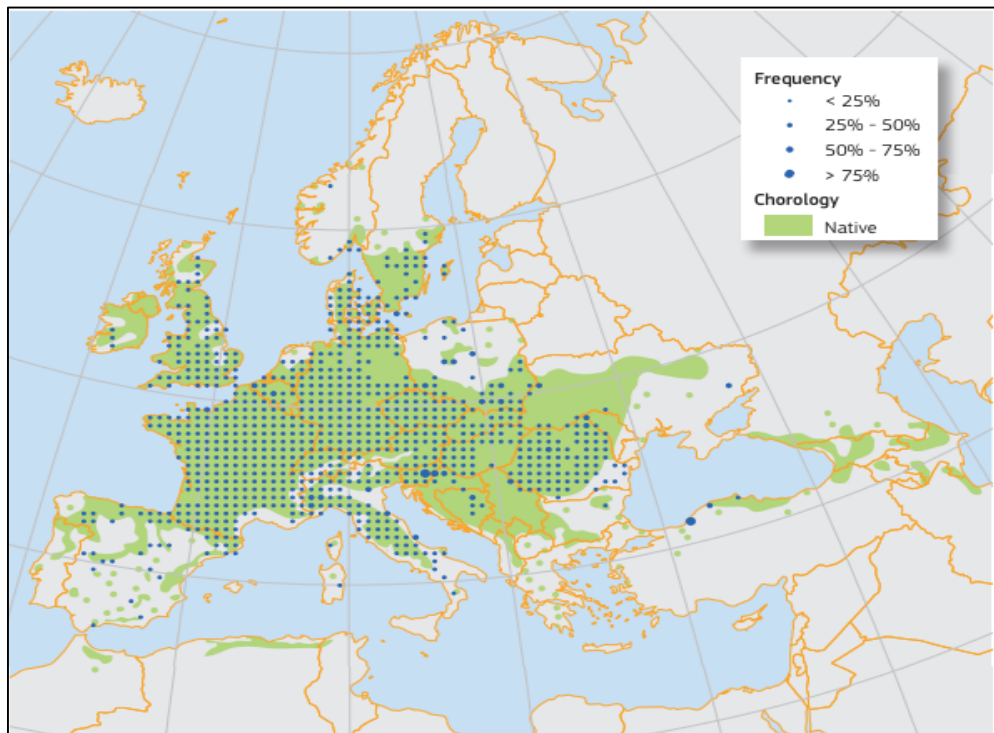


Figure 2.2: Distribution map of *P. avium* (wild cherry).

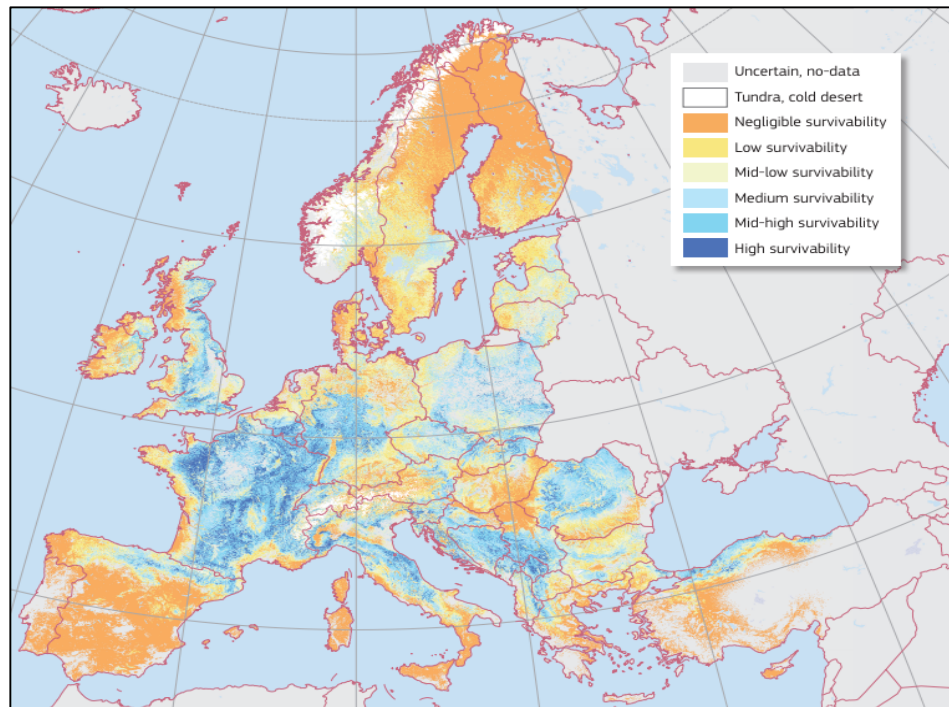


Figure 2.3: Map estimating the maximum habitat suitability.

2.1.2. Usage of Wild Cherry

Wild cherry is a valuable broadleaved tree species whose fruits have been a source of food for many species of songbirds and insects for several thousands of years [Praciak, et al., 2013], [Scholz, 1995]. Beside fruit consumption, wild cherry is the most important European hardwood species in the family Rosaceae. It has straight and fine-grained timber which makes it easier to work in wood turning. Therefore, it is sought after for panelling and cabinet-making with its valuable solid and dense wood. Wild cherry wood's reddish-brown timber is used to make decorative veneers and furniture, turnery products, decorative paneling, decorative joinery and musical instruments. It is also suitable for production of parquet floors. Its honey-colored, strong wood can be polished to a good shiny brown colour [Web 4].

The bark of wild cherry trees usually contains small amounts of water-soluble polyphenol which is tannin that is one of the most important secondary metabolites, that sometimes it rises up to 16%. From bark wounds, the aromatic gum can be obtained and it can be considered as a substitute for chewing gum [Web 1]. *P. avium* trees are also used in gardening for its foliage by adding value to the gardens during the autumn when its green leaves turn a vivid, brilliant, rich orange/red color. It is also

very valuable for medicinal purposes. The peduncles of the drupes of wild cherry boiled in water are used the treatment of cystitis, oedema, looseness of the bowels as astringent, antitussive, and diuretic [Ercisli, 2004], [Web 1]. Wild cherry trees are of great ecological and naturalistic importance, being an essential component of sensitive and threatened ecosystems. Therefore, many European countries have gene conservation and/or breeding programmes for wild cherry [Russell, 2003], [Jarni, 2012].

2.2. DNA Barcoding

2.2.1. The History of DNA Barcoding

In 1989, the World Wildlife Fund defined biodiversity as “the richness of life on Earth – millions of plants, animals and microorganisms, including the genes which they carry, and complex ecosystems that create the environment” [Primack et al., 2001]. The term biodiversity covers the variability of life on Earth. In the literature, the term of “DNA barcodes” was first used in 1993 [Arnot et al., 1993]. The mapping and protection of animal and plant biodiversity was aimed by using DNA barcoding technique in 2003 [Mendel et al., 2012]. The Consortium for the Barcoding of Life’s Plant Working (CBOL) recommended *matK* and *rbcL* coding regions of chloroplast genome as DNA barcodes for all land plants. This recommendation was promoted by the other potential gene regions [Cohen, 2011]. The aim was the creation of a public DNA barcodes library for all living organisms. DNA barcode records are stored and published in The Barcode of Life Data Systems (BOLD) which is an informatics workbench [Mendel et al., 2012].

If it is necessary to make comparisons between coding and non-coding regions, mutations that occurred in non-coding regions can accumulate with virtually no effect on phenotype. This feature makes non-coding regions as more informative nucleotide characters to distinguish between closely related species. Noncoding regions contain insertions and deletions and these mutations help resolve phylogenetic relationships at lower taxonomic levels [Cohen, 2011].

DNA barcode both provides species level identification and allows for biodiversity assessment, life history of species or populations, ecological studies and forensic analysis.

There are many potential usage areas of DNA barcoding technique such as;

- Identification of fragment of plant material,
- Identification of unknown species,
- Phylogenetic studies,
- Identification of different life stages example seed and seedling,
- Enhance forensic analysis,
- Verification of herbal medicines and food stuff,
- Inventory and ecological surveys [Prajapat and Chakraborty, 2011].

2.2.2. The Technique of DNA Barcoding

DNA barcoding is a recent scientific technology that uses the analysis of standard DNA sequences to identify both animal and plant species quickly and accurately [Kress et al., 2005], [Savolainen et al. 2005]. DNA barcodes can be used as biological barcodes for identification of species at the species level of any living organisms. DNA sequence differences in small parts of the genomes of organisms and they can be used as a universal diagnostic dataset to identify such species by matching DNA sequences of unidentified species with DNA sequences in DNA barcode databases [Sass et al., 2007]. The main feature distinguishing DNA barcoding from other molecular markers is that it has a universal biological identification system for all species using a single gene region [Erickson et al., 2008].

The goal of DNA barcoding is to find a few DNA regions that can distinguish plant and animal species in the world and to create a large-scale reference library of life by sorting these regions from different sample groups. The main purpose of this approach is to make a major contribution to taxonomy at the species level and to help to identify unknown species. In addition to taxonomic research, DNA barcoding also contributes to population genetics and phylogenetic studies [Hajibabaei et al., 2007], [Hebert et al., 2004].

Because of the relatively low evolutionary rates of chloroplast genes in comparison to nuclear genome, plant molecular systematics and DNA barcoding techniques which are used for studying the molecular phylogeny of plants most rely on the use of chloroplast genome sequence data. The haploid chloroplast genome which generally uniparentally transmitted has a simple and stable genetic structure as there is very rare recombination in the chloroplast genome. PCR amplification and sequencing of chloroplast genes are comparatively easy due to the usage of universal primers that amplify target chloroplast gene sequences [Hurst et al., 2005].

As it is schematized in Figure 2.4, the DNA barcoding method comprises the steps of performing DNA isolation from tissue samples of the investigated plant line, amplification of the resulting DNA with the targeted barcode gene region using the PCR technique, performing DNA sequence analysis of the amplified region, and accessing the barcodes to identify the species with the obtained DNA sequences. At the end of this process, DNA barcodes are recorded in data banks [Hebert et al., 2003], [Web 2].

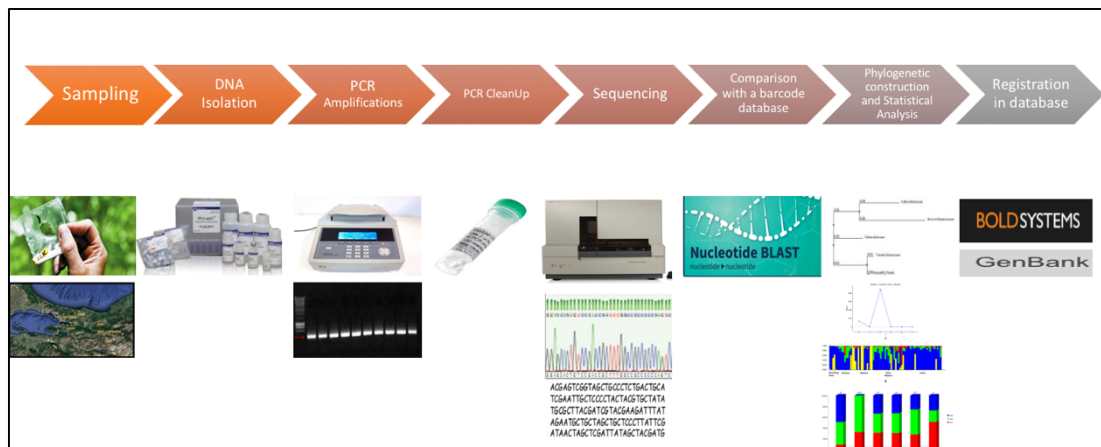


Figure 2.4: The barcoding pipeline.

A genetic barcode is a region of DNA that has an unchanged nucleotide sequence at intraspecific level with sufficient interspecific variability [Shneer, 2009]. There are three criteria for an excellent DNA barcode must contain i) significant genetic variation and divergence at the species level (discrimination); ii) possessing flanking sites the highly conserved amplification region for the development of universally applicable universal primers (universality); iii) a short sequence length to facilitate the template extraction process and amplification of the gene of interest (quality). Species

identification by DNA barcoding is accomplished by a short DNA sequence (barcode) from a standard segment of the genome. Each unknown barcode sequence is then compared to a reference barcode sequence library generated from known individuals. If the sample sequence is significantly similar to any sequence in the barcode library, it is identified. If no resemblance can be found, a new haplotype leads to a new barcode sequence for the given species [Kress and Erickson, 2008], [Ausubel, 2009].

In plants, candidate barcode genes belonging to the three genomes (chloroplast, mitochondria and nuclear) and their potential for identification of plant species are shown in the Table 2.2 (modified figure from Chen et al., 2010) and Figure 2.6 [Purty and Chatterjee 2016].

Although the gene encoding subunit 1 of the cytochrome-c oxidase (*COI*) is previously used as an animal DNA barcode, it is not a suitable plant barcode due to lack of sequence variation among plant species. *COI* gene region, is widely used as an animal DNA barcode. Plants rapidly change their mitochondrial genome structure. Many strategies have been proposed for plant species, either based on a single chloroplast region or on a combination of different regions. For plant species, due to its lower evolution rate than isomorphs present in animals, *COI* is not valid [Kress et al., 2005].

Table 2.2: Genes from three genomes in plants that are candidate barcodes.
 ***markers are potential barcodes, **markers are pending to be investigated and
 *markers are poor candidates.

Chloroplast	Mitochondria	Nucleus				
<i>trnH-psbA</i> ***	<i>COI</i> *	<i>ITS</i>				
<i>rpoC1</i> ***		<i>ITS 1</i> **	<i>ITS 2</i> **	5.8 <i>S</i> *		
<i>rbcL</i> ***						
<i>matK</i> ***						
<i>rpoB</i> *	Other mitochondrial genes*					
<i>ndhJ</i> *						
<i>accD</i> *						
<i>atpF-atpH</i> **						
<i>psbK-psbI</i> **						
<i>ycf5</i> **						
<i>trnL</i> **						
Other plastid genes or their introns**						

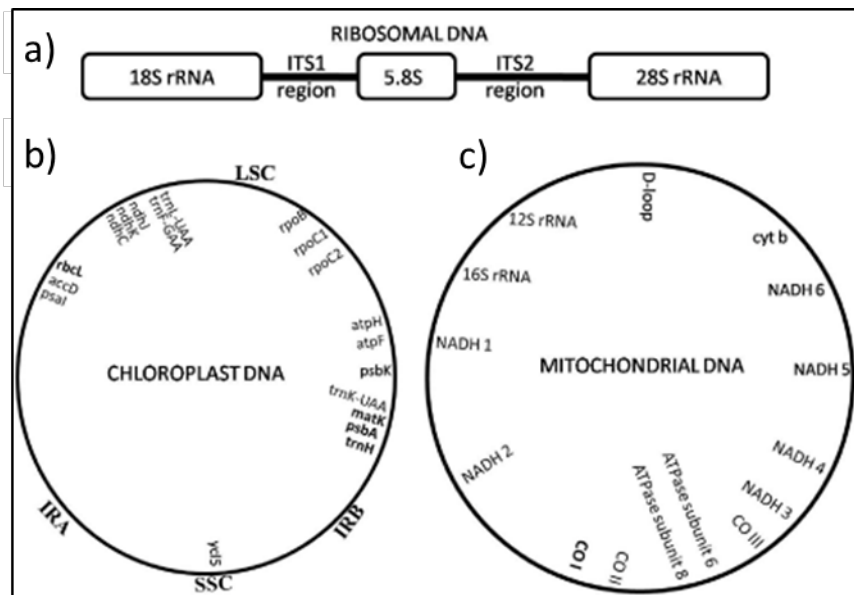


Figure 2.6: The DNA Barcoding Locus a) Ribosomal DNA showing ITS region b) Chloroplast DNA c) Mitochondrial DNA.

Based on the CBOL-The Consortium for the Barcode of Life studies, four coding regions (*matK*, *rbcL*, *rpoB* and *rpoCl*) and three non-coding regions (*atpF-atpH*, *trnH-psbA* and *psbK-psbI*) were identified as active barcode regions [Dong et al., 2012]. There is no agreement on this issue among the research groups. DNA barcoding aims to find a single sequence to identify all species, despite no single-locus barcode can achieve that goal yet [Li et al., 2015].

2.2.3. DNA Barcode Genes

2.2.3.1. DNA Barcode Genes Peculiar to Chloroplast Genome

The genotype of the chloroplast varies in size between 120 and 170kb pairs. The chloroplast genome in land plants is a genome that is suitable for evolutionary and phylogenetic studies, especially at the species level, because it has a high degree of dimensional order, structure, gene content and linear order of genes. DNA barcoding technology is based on the chloroplast genome, which has a relatively low evolutionary rate compared to the nuclear locus [Dong et al., 2012].

2.2.3.1.1. *rbcL*

The gene coding for *rbcL* (RuBisCO large subunit) is located in the large single-copy region of the plastid genome [Savolainen et al., 1997]. *rbcL* gene codes for distinct enzymatic functions: large subunit of the ribulose-1,5-bisphosphate-carboxylase/oxygenase (RuBisCO), a free enzyme in the stroma [Zurawski et al., 1982]. *rbcL* is the best characterized gene among all the plastid regions. *rbcL* provides high quality of bidirectional sequence [Prajapat and Chakraborty; 2011]. *rbcL* is one of a slow-evolving genes. It has high sequence conservation that allows a good resolution that has been confined to the family level among plant groups [Hilu and Liang, 1997]. It is not the most variable region nevertheless it is a component of the best performing multi-locus combination for plant species discrimination. CBOL Plant Working group recommended a core-barcode consisting of portions of two plastid coding regions, *rbcL+matK* [CBOL Plant Working group, 2009]. The usage of the combination of *rbcL+matK* as a recommended barcode was based upon the

discriminatory power of the *matK* gene and the straightforward recovery of the *rbcL* region [Hollingsworth, 2011].

2.2.3.1.2. *trnH-psbA*

trnH-psbA is a polygenic intergenic spacer, located between the coding genes for the tRNA^{His}GUG and the 32-KDa photosystem II tilacoidal membrane protein. The spacer between *PsbA* which is coding for the photosystem II protein D1 and *trnH* which is responsible for the transfer RNA for histidine, is rapidly evolving regions within the chloroplast genome [Kress et al., 2007]. *trnH-psbA* is the most variable region in the chloroplast genome [Whitlock et al., 2010]. *trnH-psbA* gene, which is the non-coding intergenic region, is used strictly as a plastid barcode. *trnH-psbA* barcode genome provides non-coding sequences of the chloroplast genome, the main source of data for molecular systematic and population genetic studies in plants [Shaw et al., 2005]. As Kress et al. (2007) demonstrated, this sequence has good amplification capacity among terrestrial plants (93% in Angiosperm) and a high level of interspecific divergence.

Priming site of *trnH-psbA* exhibits high sequence divergence among species so, it is highly suitable as a plant barcode [Goremy et al., 2005].

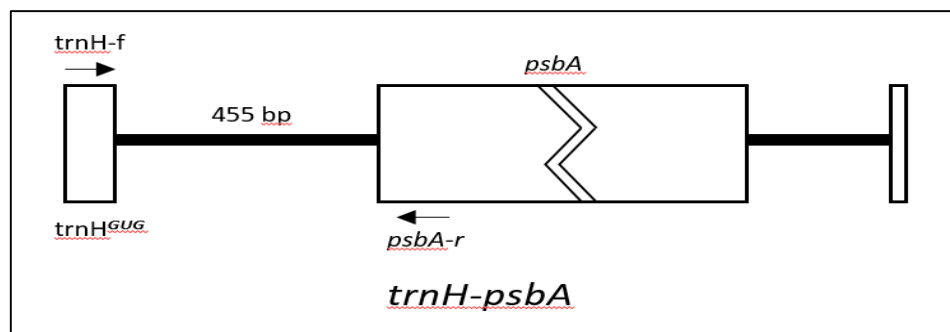


Figure 2.7: PCR priming site within highly conserved flanking coding sequence combine with a non-coding region for *trnH-psbA* region. Large white boxes indicate coding areas, small white boxes indicate intergenic spacers, and small black boxes indicate introns.

2.2.3.1.3. *matK*

matK (maturase K) is one of the most rapidly evolving plastidic regions in plants. It is one of the most promising candidates for barcode analysis in land plants. For most land plants, *matK* gene whose ORF is approximately 1570 bp in most angiosperms is nested between the two exons of *trnK*, tRNA-lysine [Hilu et al., 1999]. A fragment of 600-800 bp for the DNA barcoding *matK* gene is usually sufficient. *matK* exhibits relatively high insertion/deletions at the 3' end, thus evolves more rapidly than other chloroplast genes [Hilu et al., 1999]. It encodes for maturase K, the enzyme necessary for the processing of the plasidogenous ribosomal RNA, and its sequence exhibits high levels of discrimination among species in the group of Angiosperms [Fazekas AJ et al., 2008], [Lahaye et al., 2008]. The maturase K (*matK*) is one of the most variable coding genes of angiosperms and it is useful for identification at family, genus, and even species levels [Dunning et al., 2010], [Yu et al., 2011]. It is characterized by its fast evolution rate. *matK* has the rate of nucleotide substitution is three times higher than the large subunit of Rubisco. [Hilu and Liang, 1997]. As a consequence of its high substitution rate and its more variable sites, the *matK* gene is considered to be fast-evolving comparing to the other plastidial genes. It provides enough revenues for evolutionary analysis at the family level and below [Hilu and Liang, 1997]. In the DNA barcoding studies, it was found that the chloroplast (*matK*) gene region was the fastest growing plastid coding region; it has a high degree of discrimination as it has high evolutionary proportions and separation between species [Lahaye et al., 2008]. Unfortunately, *matK* is difficult to amplify universally using currently available primer sets. Therefore, additional universal primers for optimization of PCR reactions can be required [Dunning et al., 2010], [Yu et al., 2011]. As a consequence of the findings, the *matK* barcode should be used in combination with other barcode genes, since it has a high discrimination power but is universally difficult to amplify.

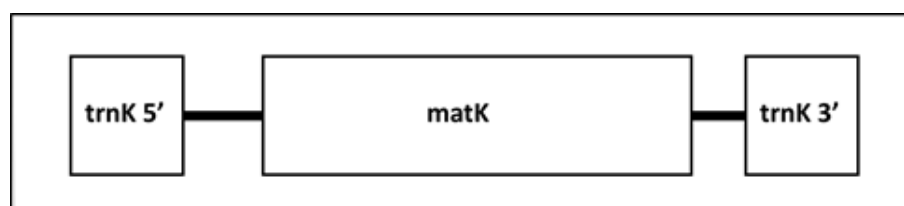


Figure 2.8: Diagram showing *matK* gene nested between *trnK* introns.

2.2.3.2. DNA Barcode Genes Peculiar to Nuclear Genome

The complexity of the nuclear genome and the high number of introns and repetitive sequences, are the reasons of their less frequently usage in systematic botanical studies.

2.2.3.2.1. *ITS*

ITS (Internal transcribed spacer) refers to the spacer DNA situated between the small-subunit ribosomal RNA (rRNA) and large-subunit rRNA genes in the chromosome or the corresponding transcribed region in the polycistronic rRNA precursor transcript. Especially 5.8S, *ITS* regions, which are internal transcribed regions of nuclear ribosomal DNA, can be used as potential barcodes (with the exception of ferns) and suggested as a possible plant barcode locus [Stoeckle, 2004]. *ITS* is the evolutionary conserved region which is present on nrDNA (nuclear ribosomal DNA) sequence. *ITS* locus sequences are important markers in systematic investigations such as identification of plant species, taxonomic clarification and phylogenetic analysis at the molecular level. An advantage of this region, particularly for degraded sample analysis, is the possible amplification in two smaller fragments (*ITS1* and *ITS2*) adjoining the 5.8 locus. The entire *ITS* region (*ITS1*-5.8S-*ITS2*) varies between 600-1000bp depending on plant species while in most of the angiosperms it ranges between 565-700bp. *ITS* spacer has much higher divergence value than any of the plastid regions [Prajapat and Chakraborty; 2011]. *ITS* spacer, which is a repetitive region in the ribosomal RNA, is a powerful phylogenetic marker at the species level showing high levels of interspecific divergence. Because of the high discriminatory power on the plastid regions in low taxonomic levels, it is proposed as a plant barcode [Alvarez and Wendel, 2003], [Sahare, 2013].

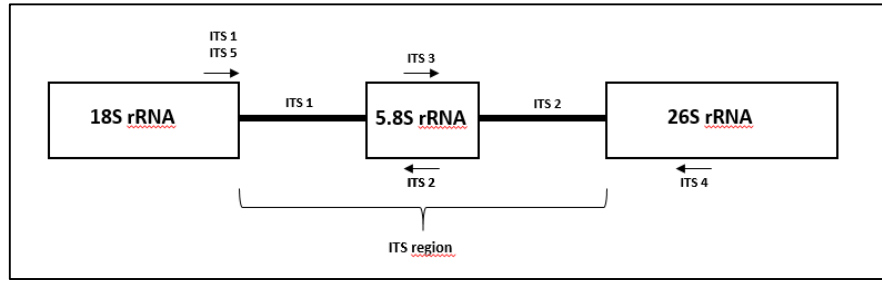


Figure 2.9: Priming site of *ITS*. Large white boxes indicate coding areas, small white boxes indicate intergenic spacers, and small black boxes indicate introns.

2.2.3.3. Other DNA Barcode Genes

rps16, *trnL-F*, *trnS-G*, *rPOB*, *rPOCl*, *aTPF-aTPH*, *psbK-psbI*, *yCF5*, *atpB*, *atpB-rbcL*, *ndhF*, *rpl16*, *rps4-trnS*, and *trnL* which are the fragments of coding regions, introns, and intergenic spacers, are used for phylogenetic reconstructions [Gao et al., 2008], [Li, 2008], [Wilson, 2009], [Peterson et al., 2010]. These chloroplast regions are valuable for phylogenetic analyzes and barcoding studies at high taxonomic levels, but unfortunately, in species which have diverged recently in evolution, these regions often lack variations. [Li et al., 2015], [Kress et al., 2005], [Chase et al., 2007], [Newmaster et al., 2006], [Hollingsworth et al., 2001], [Seberg et al., 2009]. By the means of identified combined-variable locations and their universal primers which were designed, such extra investments could be avoided. Therefore, a combination of primers which amplifies barcoding genes must be used for improving the phylogenetic analysis [Dong et al., 2012].

Some regions such as *rbcL*, *rpoB*, and *rpoCl* are useful for identification at the family rather than species level. Therefore, at present, *matK* and *trnH-psbA* are accepted as two most promising choices of chloroplast regions [Dunning et al., 2010], [Yuet al., 2011]. However, many investigators have suggested that a multi-locus method is necessary to obtain adequate species discrimination [Hebert et al., 2004], [Kress and Erickson, 2007], [Kress and Erickson et al., 2008], [Lahaye et al., 2008]. A single locus may not differentiate more than 80% of plant species. Several plastid locus combinations including *trnh-psbA* have been proposed. *rpCL* + *trnH-psbA* [Kress and Erickson, 2007] *rPoCl* + *rPoB* + *matK* or *rPoC* + *matK* + *trnH-psbA* [Chase et al., 2007] combined barcodes exhibit higher species discrimination than single-locus approaches. Researchers concluded that a combined analysis of two or

three of these genes could constitute a standard DNA barcode protocol for plant kingdom [Kress et al., 2005], [Li et al., 2015].

Identification and greater availability of a larger number of variable loci with very high evolutionary rates in chloroplast genomes has been advocated for the best discrimination in identifying species, estimation of biodiversity, elevation of the ability to solve phylogenetic problems at the species level and for molecular systematics. Thus, both plant phylogenetic analyses and DNA barcoding studies, investigation of more regions with high evolutionary rates is extremely important [Dong et al., 2012], [Kress et al., 2007].

2.3. Phylogenetic Analysis

Evolution is defined as the development of a biological form, from its pre-existing forms or its natural origins through various natural selection and modifications to reach its present form. The answer to the question of how a family develops can be found in the evolutionary process [Patwardhan et al., 2014]. Based on the nucleotide differences identified in the same barcode regions sequenced by DNA of different taxa, it is possible to estimate the phylogenetic relationships and evolutionary divergence times that exist between them. Phylogenetic trees are used to show the evolutionary relationships of various biological species and phylogenetic trees are important in terms of showing the patterns and times of branching, specifying the sorting order, and showing the closeness of groups of organisms (taxa) on the same tree [Patwardhan et al., 2014], [Peng, 2007]. Along with the geographical and ecological characteristics of the species from the phylogenetic trees, information about the processes or causes of the genesis in the group can be provided [Barracough and Nee, 2001]. The phylogenetic tree, consisting of nodes and branches, shows the time evolution of ancestral populations of species. In addition, nodes correspond to the point at which a species divides into two or more derivative populations [Penny et al., 1993]. The root represents a common order, and the tree can be rooted anywhere. In the case of rootless trees, only the interspecific relationship is shown as a network without a common ancestor [Mount, 2001].

In the evolutionary process, each taxonomic group coincides with a node of the phylogenetic tree. Species closely related to each other are located in adjacent allows

in the tree. The pattern of tree branches is called the topology, and the branch length usually specifies the number of changes that occur in the branch [Mount, 2001].

The general principle, which is the basis of the phylogenetic method, is to draw a tree chart, which groups the most similar individuals among them, trying to find a logical explanation of the genetic differences identified in them and minimizing theoretical evolutionary changes which allowed such divergences [Hedges, 2002]. Especially in the past, the construction of such phylogenetic trees was based on molecular clock theory, according to which random mutations, through which genes evolve, occur at almost constant frequencies over time. Based on this principle, it was possible to estimate the time elapsed since two species, descending from the same common ancestor, separated, evaluating the amount of nucleotide differences present in related DNA sequences. This theory has proved to be true but not universal: not all genes mutate with the same frequency and not all species show the same ability to change [Wilson et al., 1977], [Morgan et al., 1998]. Therefore, various probabilistic computer programs have been designed and developed to generate phylogenetic trees based on different evolutionary principles [Nei et al., 2000]. The reliability and significance of the results obtained in this field can be tested in several ways, the most used being the Bootstrap Method [Felsenstein et al., 1985]: randomly sampled nucleotides with their substitutions and a new tree; this procedure is repeated many times and the occurrence frequency of particular nodes between Bootstrap trees determines a value of support (trust) in deciding the significance of that node in the tree. The construction of phylogenetic trees, therefore, has the purpose of trying to explain and interpret, in evolutionary terms, the molecular differences identified between different organisms. This type of analysis allows to determine the most likely forms of common ancestor together with when the various evolutionary branches have started to diverge and what are the characteristics of the most recently never individuals.

Such methods for creating phylogenetic trees are based on clustering or optimization algorithms. The former takes into account the genetic distances calculated by comparison matrices; such as the Neighbor Joining (method that selects among the various hypothesized trees the one with the overall length, ie the sum of the lengths of all branches, shorter) [Saitou and Nei, 1987] and the UPGMA (assuming instead, as the true and absolute theory of the molecular clock). The latter, on the other hand, try to produce trees based on objective quality criteria. In this latter group, we

find methods such as Maximum Likelihood [or maximum likelihood, it selects the tree with the highest probability of explaining the data obtained] [Felsenstein, 1981] or Maximum Parsimony (which privileges, amongst the various possibilities, the tree that requires the lowest number of replacements). These methods are flexible enough to include a priori different information about the biological nature of the molecular sequence change, such as the different rates of variation between the sites. The greatest advantage of the distance method over character-based methods is that it is much faster and less computer-centric. The disadvantage is that the real character data is ignored.

Models are a tool to reduce the complexity of the mutation process to pattern that can be described by a small number of parameters [Strimmer, 1997].

2.3.1. Neighbor Joining (NJ)

Neighbor joining which is a clustering method is an algorithm to optimize a tree according to the 'balanced minimum evolution' criterion. It is based on Kimura-2-parameter for the creation of phylogenetic trees [Saitou and Nei, 1987]. The Kimura model distinguishes between two types of substitutions among sequences: transitions which are interchanges of two-ring purines (A to G or visa versa) or of one-ring pyrimidines (C to T or visa versa) and transversions where a purine is replaced by a pyrimidine or vice versa [Kimura, 1980]. Owing to the fact that transversions are more genotypically meaningful for an organism, the model assumes that the rate of transversions is more major than transitions. The method is used for constructing phylogenetic trees by using evolutionary distance data without making the assumption of molecular clock. The steps of the calculation of the algorithm were proposed by Saitou and Nei, 1987 are explained below:

- At first, unresolved star-like tree is constructed.
- The sum of all branches length is calculated of the resultant tree.
- The pair that yields the smallest sum is considered the closest neighbors and they are joined.
- A new branch is inserted between them.
- The branches of the rest of the tree length is recalculated.
- The process is repeated until only one terminal is left [Saitou and Nei, 1987].

It is known that the Neighbor Joining method gives statistically more stable results. The NJ tree model [Saitou and Nei, 1987] is a useful clustering method for large data sets [Yang, 2012].

2.2.3. Maximum Likelihood (ML)

The algorithm which is first presented by Felsenstein (1973) is an efficient technique that is used for computing phylogenetic likelihoods. The method is a statistical inference framework. This estimation method is used to estimate the parameters in a probabilistic data generating model, besides this, it is the most computationally intensive method. The method requires that evolution at different sites must be statistically independent. According to Felsenstein (2004) there are two crucial assumptions for this method:

- On a given tree, evolution at different sites is independent.
- Evolution proceeds independently on different branches of the phylogenetic tree [Dhar and Minin, 2015].

Time is flowing from left to right. In this tree, the horizontal axis is measured as genetic change and to convert this into actual time it is needed to make some assumptions-molecular clock- about the relationship between genetic change and time. Branch lengths are related to the time evolutionary decomposition of subspecies mutually. The horizontal lines are branches which provide a scale of change and gives the amount of genetic change that means represent evolutionary lineages changing over time. As the branch size in the horizontal dimension increases, the amount of change become to increase [Web3].

2.4. Statistical Analysis

2.4.1. Principal Component Analysis (PCA)

Principal component analysis (PCA) is a mathematical procedure that transforms a number of possibly correlated variables into a smaller number of uncorrelated

variables called principal components. This statistical method is used in population genetics to identify structure in the distribution of genetic variation across geographical location by projecting samples onto a series of orthogonal axes; each of which is made up of a linear combination of allelic values across SNPs or other types of variant [McVean, 2009].

2.4.2. Structure Analysis

The method uses a systematic Bayesian clustering approach which identifies populations from the data and assigns individuals to that population representing the best fit for the variation patterns found for population analysis. Structure applies a model to the data of K assumed populations [Hurtado et al., 2013]. ΔK estimates of the posterior probability distribution of the data for a given K . ($|L''(K)|$) is calculated by the formula of $[|L(K+1) - 2L(K) + L(K-1)|]$. Accordingly, ΔK is calculated as $\Delta K = m(|L''(K)|) / s[L(K)]$. It has been reported that the calculated ΔK value enables a correct K value to be determined in the studied populations [Evanno et al 2005].

The steps of this calculation method proposed by Evanno et al (2005) are explained below:

- The average probability $[L(K)]$ for the number of repetitions for each K is calculated,
- The average difference between successive probability values of K values is calculated using the formula,

$$L'(K) = L(K) - L(K - 1) \quad (2.1)$$

- The absolute value between consecutive $L'(K)$ values is calculated using the formula,

$$|L''(K)| = |L'(K + 1) - L'(K)| \quad (2.2)$$

- Finally, the value of

$$\Delta K = m(|L''(K)|)/s[L(K)] \quad (2.3)$$

is calculated by dividing by the standard deviation of the L (K) value calculated for the number of repetitions of the average value of the absolute value of L''(K).

The lowest stable K value represents the optimum value. 100,000 times repeated method analyses differences in the distribution of genetic variants amongst populations by placing samples into groups whose members share similar patterns of variation [Hurtado et al., 2013]. A Python program based ‘Structure Harvester’ is used for the aim of summerizing the data received from ‘Structure 2.3.4’.

2.5. Previous DNA Barcoding Studies

Two species of Angiosperms (*Atropa belladonna* and *Nicotiana tabacum*) were used by Kress and co-workers (2005) in the aim of DNA barcoding study bu using the gene regions of *trnK-rps16*, *trnH-psbA*, *rpl136-rps8*, *atpB-rbcL*, *ycf6-psbM*, *trnV-atpE*, *trnC-ycf6*, *psbM-trnD* and *trnL-F-trnh-psbA*, ITS and *rbcL*. Their study showed that the *trnH-psbA* gene region is the best plastid option for determining intra-species variation of these two species as a DNA barcode sequence.

Newmaster and co-workers (2008), aimed the availability of the universal plastids of *UPA*, *rpoB*, *rPoCl*, *accD*, *rBcL*, *matK* and non-coding spacer of *trnH-psbA* chloroplast loci in Componeura. They have shown that 5 regions (*UPA*, *rpoB*, *rpoCl*, *accD*, *rbcL*) which were tested are not variable with respect to the species. However, *matK*, *trnH-psbA* regions show significant variation. Based on this evidence, the Authors maintain DNA barcoding studies are promising in coconut.

By the application of the DNA barcode technique, Gismondi and co-workers (2013) analyzed different *Crocus* species genomes and partially clarified some aspects of the phylogeny of this genus: in particular, *C. sativus* possible genetic derivation was elucidated. Hispanic *C. sativus* (ES), *C. carpetanus* and *C. serotinus* plants and Italian *C. cartwrightianus*, *C. hadriaticus*, *C. thomasii* and *C. sativus* (IT) species were subjected to genetic characterization bu usibg DNA barcodes of *matK*, *rbcL*, *trnHpsbA*

and *ITS*. Phylogenetic tree clearly showed the sharp separation of *C. hadriaticus* and *C. thomasi* from the other specimens, as suggesting a very ancestral origin of these two species. *RbcL* and *matK* sequences proved to be 100% conserved. *TrnH-psbA* and *ITS* regions presented genetic divergences, respectively, of 1 and 10 %. Obtained results showed that different *C. sativus* species might have evolved by independent events. They demonstrated that barcoding method which usually adopted for interspecific taxonomic identification, could be also applied to intraspecific and population studies.

Gismondi and co-workers (2012) suggested and emphasized the use of DNA barcode method for ancient plant sample analysis. In order to identify ancient seeds, aDNAs were used as templates to amplify selected barcodes which are *rbcL*, *matK* and *trnH-psbA* genes. Final outcomes showed that studied fossil remains could be classified as two different plant species seeds: *Olea europaea* L. and *Cornus mas* L. This study demonstrated the importance of the DNA barcode technique, with respect to morphological approach that is usually not completely informative, for archaeobotanical sample scientific identification. It also permitted the botanical classification of “Mora Cavorso” ancient seeds.

Ancient vegetal remains' chloroplast DNAs from a sealed amphora were amplified by using *rbcL*, *matK* and *trnH-psbA* genes by Gismondi and co-workers (2013). The intersection of the out-comes allowed to scientifically reveal the botanical origin of all the samples: *Olea europaea* L., *Crataegus monogyna* Jacq. eeds., *Pinus* sp. pollen cone., *Quercus* genus (by showing high sequence identity (95%) of the sample), *Quercus petraea* (Mattuschka) Liebl. fruit. However, data suggested that one of the sample was not a vegetable fiber (probably it had an animal origin).

Gismondi and co-workers (2015) sequenced five barcode genes of *rbcL*, *matK*, *trnH-psbA*, *ITS* and 18S rDNA from *C. sativa* samples. They observed the nucleotide sequence of *rbcL* of the samples possessed 2–3% of variability with respect to those deposited in the scientific database. *matK* region revealed the percentage of identity 100% of the matches between sample and *C. sativa* registered accessions. It is clear that *matK* gene was identified as well-conserved plastidial region. *trnH-psbA* sequence results suggested that the nucleotide variability in this gene could be associated with the geographical distance of the specimens compare. *C. sativa* 18S rDNA, was sequenced and submitted in GenBank for the first time. *rbcL*, *trnH-psbA* and *ITS* markers were suggested as good candidates for intraspecific and population

analyses. According to the final results of their study, best barcoding candidates were *rbcL* as a plastid genome and *ITS* as a nuclear genome for *C. sativa*.

It was suggested that the genetic polymorphisms can be used when comparing botanical remains to identify the variety of *Bidens pilosa* present at a crime scene by Tsai and co-workers (2008). *B. pilosa* is a plant producing barbed fruits is commonly found during forensic investigations and until now little is known of their genetic composition. Thus, the aim of the study concerned to discover genetic polymorphisms among *B. pilosa* L. varieties. Five loci within the nuclear and chloroplast genomes were screened. DNA fragments of all samples at the *trnL* intron and *trnL-trnF* IGS loci of the chloroplast genome, internal transcribed spacer (*ITS1* and *ITS2*) and the 5.8S of nuclear ribosomal DNA (nrDNA) were amplified and sequenced. According to the results, *TrnL* intron, *ITS1*, *ITS2* and 5.8S were found to be useful markers to identify *B. biternata* and *B. pilosa*.

Gonzalez and co-workers (2009) conducted a study of the tropical rain forest sampling a total of 1073 trees. The coding plastid regions *rbcLa* (first part of the *rbcL* gene), *rpoC1*, *rpoB*, *matK*, *ycf5*, the noncoding regions *trnL* and *psbA-trnH*, and the nuclear region *ITS* was used. None of the plastid markers achieved more than 70% correct plant identification. *ycf5* and *rpoC1* had the worst performance as a DNA barcode. Sequencing success was very low in *matK* and *ITS*. The *trnL*-UAA intron ranked second in the 'best close match' test. Compared with the all results, *psbA-trnH* had the best performance as a DNA barcode.

DNA barcodes were tested by Sass et al. (2007) for members of the Cycadales by using six regions (*ndhJ*, *rpoB* and *matK*, *accD*, *ycf 5* and *rpoC1*) initially. However, three of them (*ndhJ*, *rpoB* and *matK*) did not amplify across all Cycadales. In the next stage, *nrITS* showed promise as a barcoding region because it had enough variability to identify many species. *PsbA-trnH* showed promise as a barcoding locus for cycads. Authors suggested a set of primers designed for each of the major clades of land plants could be used simultaneously if universal tails were added to the primers.

Some of the other DNA barcoding studies published in plant and of the molecular marker studies with *Prunus avium* are shown in Table 2.3 and Table 2.4, respectively.

Table 2.3: Gene/locus selected for some of the study of DNA barcoding in plants.

Organism	Gene/Locus	Reference
<i>Phyllanthus spp.</i>	<i>trnH-psbA</i>	Srirama et al., 2010
<i>Senna spp.</i>		Monkheang et al., 2011
<i>Cistaceae spp.</i>		Han et al., 2010
753 genera		Chen et al., 2010
<i>Angelica spp.</i>		He et al., 2012
<i>Rhododendron spp.</i>		Liu et al., 2012
<i>Lonicera spp.</i>		Sun et al., 2011
<i>Solanum spp.</i>		Li et al., 2012
<i>Ginseng genus</i>		Zuo et al., 2011
<i>Lamiaceae</i>		Theodoridis et al., 2012
<i>Scutellaria spp. Astragalus spp.</i>		Guo et al., 2011
<i>Smilax spp.</i>		Kritpetcharat et al., 2011
<i>Vitex spp.</i>		Phoolcharoen et al., 2013
<i>Paris spp.</i>		Yang et al., 2011
<i>Sideritis spp.</i>		Tezcan et al., 2010
<i>Astragalus spp.</i>	<i>matK</i>	Guo et al., 2010
<i>Ginseng genus</i>		Zuo et al., 2011
<i>Galpemia spp.</i>		Sharma et al., 2012
<i>Solanum spp.</i>		Li et al., 2012
<i>Rhubarb</i>		Xu et al., 2013
<i>Lonicera spp.</i>		Sun et al., 2011
<i>Lamiaceae</i>		Theodoridis et al., 2012
<i>Cistaceae spp.</i>		Han et al., 2010
<i>Sideritis spp.</i>		Tezcan et al., 2010
<i>Scutellaria spp. Astragalus spp.</i>		Guo et al., 2011
753 genera		Chen et al., 2010
<i>Puerariacandollei, Butea seberb, Mucunacolletti</i>		Wiriyakarin et al., 2013
<i>Rhododendron spp.</i>		Liu et al., 2012
<i>Angelica spp.</i>		He et al., 2012
<i>Dendrobium spp.</i>		Asahina et al., 2010
<i>Boerhavia spp., Astragalus spp.</i>	<i>ITS</i>	Selvaraj et al., 2012
<i>Taraxacum and adulterants</i>		Chiang et al., 2011
<i>Orobancha spp.</i>		Sun et al., 2012
<i>Meconopsis spp.</i>		Li et al., 2011
<i>Astragalus spp.</i>		Xiao et al., 2011
753 genera		Chen et al., 2010
<i>Angelica spp.</i>		He et al., 2012
<i>Rhododendron spp.</i>		Liu et al., 2012
<i>Lonicera spp.</i>		Sun et al., 2011
<i>Solanum spp.</i>		Li et al., 2012
<i>Ginseng genus</i>		Zuo et al., 2011
<i>Astragalus spp.</i>		Guo et al., 2010
<i>Citrus spp.</i>		Wang et al., 2012
<i>Ochradenus spp.</i>		Khan et al., 2012
<i>Dendrobium spp.</i>		Chiang et al., 2012
<i>Paris spp.</i>		Liu et al., 2012
<i>Dendrobium spp.</i>		Takamiya et al., 2011

Table 2.3: continued.

<i>Rehmannia spp.</i>		Kim et al., 2012
<i>Ruta spp.</i>		Al-Qurainy et al., 2011
<i>Dipsacus spp.</i>		Li et al., 2012
<i>Hypericum spp.</i>		Kazi et al., 2013
<i>Rubus spp.</i>		Yang et al., 2012
<i>Sedum spp. Astragalus spp.</i>		Liu et al., 2011
<i>Pteridophytes</i>	<i>rbcL</i>	Ma et al., 2010
<i>Galpemia spp.</i>		Sharma et al., 2012
<i>Dendrobium spp.</i>		Asahina et al., 2010
<i>Rhododendron spp.</i>		Liu et al., 2012
<i>Angelica spp.</i>		He et al., 2012
753 genera		Chen et al., 2010
<i>Lonerica spp.</i>		Sun et al., 2011
<i>Solanum spp.</i>		Li et al., 2012
<i>Astragalus spp.</i>		Guo et al., 2010
<i>Ginseng genus</i>		Zuo et al., 2011
<i>Sabia spp.</i>		Sui et al., 2011
<i>Lamiaceae</i>		Theodoridis et al., 2012
<i>Scutellaria spp. Astragalus spp.</i>		Guo et al., 2011

Table 2.4: Some of the molecular marker studies with *Prunus avium*.

Plant species	Number of Individuals	Molecular Marker Techniques	Sampling Region	References
<i>Prunus avium</i> L. (Wild cherry)	120	SSR	Slovenia	Jarni et al., 2012
<i>Prunus avium</i> L. (Wild cherry)	278	SSR	Italy	Rogatis et al., 2013
<i>Prunus avium</i> L. (Wild cherry)	131	SSR	Spain	Fernandez-Cruz et al., 2014
<i>Prunus avium</i> L. (Wild cherry)	21	ISSR / SSR	Northern Greece	Ganopoulos et al., 2011
<i>Prunus avium</i> L. (Wild cherry)	551	SSR	England	Vaughan et al., 2007
<i>Prunus avium</i> L. (Wild cherry)	24	SSR	North-West Croatia	Tanceva-Crmaric et al 2011
<i>Prunus avium</i> (wild cherry), <i>Prunus cerasus</i> , <i>Prunus mahaleb</i>	37 8 7	SSR	Turkey	Türkoğlu et al., 2012
<i>Prunus avium</i> cultuvar, <i>Prunus avium</i> (wild herry)	63 165	AFLP	France	Tavaud et al., 2001
<i>Prunus avium</i> L. (Wild cherry)	44	PCR-AFLP	Spain, France, Italy, Germany, Scotland	Mohanty et al., [2001]

3. MATERIAL AND METHOD

3.1. Plant Material

The first part of this thesis study was carried out in Turkey (Gebze Technical University) leaf sampling: leaves were collected from 7 different natural sites (Abant, Mollafenari, Yedigöller, Düzce, Melen, Kefken and Gölcük) of Black Sea Region of Turkey under the jurisdiction of Adapazarı and Bolu Forestry Directorates (Table 3.1 and Figure 3.1). Rest of the thesis were carried out in Italy (University of Rome “Tor Vergata”). To increase the information level of this analysis, one other *Prunus* species as considered outgroup of *P. avium* have been added to the Turkish samples at Rome Tor Vergata region: *P. armeniaca*, in order to make a direct comparison with the Turkish specimen.

Table 3.1: Growing regions, population and subpopulation names and altitudes on the map, coordinates and individual numbers of collected plants.

Region	Population	Subpopulation	Altitude	North	East	Number of collected individual
Adapazarı	Kocaeli	Gölcük	670	40°39' 43.96"	29°41'59.59"	20
Adapazarı	Kocaeli	Kefken	150	41°7'59.51"	30°11'33.62"	20
Adapazarı	Kocaeli	Mollafenari	300	40°58'28.87"	29°32'29.88"	20
Bolu	Bolu	Abant	950	40°39'30.50"	31°24'19.97"	20
Bolu	Bolu	Yedigöller	960	40° 47'24.03"	31°38'17.75"	19
Bolu	Düzce	Düzce	250	40° 47'42.87"	31°15'21.42"	20
Bolu	Düzce	Melen	320	40°47' 32.06"	30°53'26.52"	20



Figure 3.1: Distribution area of collected leaves.

About 20 individuals per each of 7 different Turkish wild cherry subpopulations' leaves from adult productive trees were collected with the help of long pruning shear (Figure 3.2) from Abant, Mollafenari, Yedigöller, Düzce, Melen, Kefken and Gölcük regions by paying attention to be at least 100 m distance between the trees and to be absence of altitude differences of more than 300 m between trees. Samples were stored at -80°C.



Figure 3.2: Sampling with the help of long pruning shear.

3.2. The Used Equipments

The list of the devices used in this thesis is given in the Table 3.2.

Table 3.2: The list of the devices used this thesis.

DEVICES	CATALOG NUMBER	BRAND
Microbalances	TP3030	Denver Instrument
Autoclave	HV-110L	Hirayama
Drying oven	DIN 12880	Binder
Magnetic stirrer	MR3001	Heidolph
Deep-freeze (-20 °C)	GS26DN13NE	Siemens
Refrigerator (4 °C)	RG 1380	Indesit
Ultra low temperature (ULT) freezer (-80 °C)	DW-86L628AH	Symphony
Hot water bath	D-77960	Julabo
Vortex	AC230	VNR International
Microwave oven	MW71E	Samsung
Automatic pipette	3862243	Eppendorf Research
Centrifuge	5430R	Eppendorf
Gel electrophoresis tank and power adaptor	MP300	Claver Scientific
Gel imaging machine	Universal Hood II	BIO-RAD
Thermal cycler	IQ5	BIORAD
Genetic sequencer	ABI 3130	Applied Biosystems

The list of the kits and chemicals used in this thesis is given in the Table 3.3.

Tablo 3.3: The list of the kits and chemicals used this thesis.

KIT AND CHEMICALS' NAME	CATALOG NUMBER	BRAND
i-genomic plant DNA extraction mini kit	17371	Intron biotech
Agarose	T6066	Sigma Aldrich
Trizma base	S3014	Sigma Aldrich
Acetic acide	324503	Calbiochem
KAPA Taq PCR Kit	BK1001	Kapa Biosystems
DMSO	D9170	Sigma Aldrich
HyperLadder™ 1kb	BIO-33053	Bioline
Loading dye	372978	Sigma Aldrich
λ DNA-Hind III	A3461.1000	AppliChem
BigDye™ Terminator v3.1 Cycle Sequencing Kit	4337456	Applied Biosystems
MassRuler DNA Ladder Mix,	SM0403	Fermentas
Hi-Di Formamide	4311320	Applied Biosystems
ExoSAP-IT™ PCR Product Cleanup Reagent	78200.200.UL	Applied Biosystems
Ethyl alcohol, Pure 95.0%	493511	Sigma Aldrich
Sodium acetate anhydrous, for molecular biology, ≥99%	S2889-250G	Sigma Aldrich

3.3. DNA Extraction

The genetic material was extracted from leaf portions of each sample by using i-genomic plant DNA extraction mini kit (Intron Biotechnology) (Table 3.4).

Briefly, the leaves (50 mg) were crushed with liquid nitrogen with the aid of homogenizator until they became powder. The samples were re-suspended in 422 µL lysis buffer (390 µL Buffer PG, 7 µL enhancer solution, 20 µL proteins K, 5 µL RNase A) which was provided by the kit and incubated at 65°C for 30 min. 100 µL PPT

solution t was added to each sample; then, tubes were incubated on ice for 5 min and centrifuged at 13.000 rpm/min for 5 min at room temperature. 650 µL PB buffer which is provided by the kit was added to 200 µL of supernatant of each sample. 650 µL of prepared mix was transferred to filtered tubes and centrifuged at 13.000 rpm/min for 1 min. Supernatants were discarded. Nucleic acid that was bound to a specific membrane surface were washed, by adding solution (PWA buffer) and centrifuged at 13.000 rpm/min for 1 min. Filtered tubes were carried into new collection tubes and were precipitated by adding a washing solution (PWB buffer). Pellets were eluted in 100 µL PE buffer which was provided by the kit and incubated at room temperature for 1 min; then, a final centrifugation at 13.000 rpm/min for 1 min was conducted. All samples were stored at -20°C.

Table 3.4: i-Genomic Plant DNA Extraction Kit contents.

Label	Description
Buffer PG	Lysis buffer
Buffer PPT	Precipitation buffer
Buffer PB	Binding buffer
Buffer PWA	Washing buffer A
Buffer PWB	Washing buffer B
Buffer PE	Elution buffer
Encancer Solution	Binding enhancer buffer
RNase A	Enzyme
Proteinase K	Enzyme

3.4. Agarose Gel Electrophoresis for confirmation of DNA Extraction

Extracted DNA samples were run on 1.5% agarose gel using 1X TAE buffer (40mM Tris, 1mM EDTA, 20mM acetic acid, pH 8.5) containing 10 mg / ml of ethidium bromide visualized and photographed under UV rays (Gel Doc 2000 BIO-RAD).

3.5. PCR Amplifications and Agarose Gel Electrophoresis

PCR amplification of the purified DNA was conducted in a 50 µL solution that contained 50-200 ng of template DNA, 2.5 U KAPA Taq DNA Polymerase (Kapa

Biosystems), 20 µM of each of the two primers (Sigma-Aldrich), 0, 2 mM for each dNTP, 1X Taq DNA Polymerase Buffer, 3 mM MgCl₂, and 5% DMSO. DNA amplification was performed using BIO-RAD iCycler according to the following parameters: initial denaturation at 95°C for 4 min (hot start); 45 cycles of: denaturation at 95°C for 20 sc, annealing at the appropriate temperature for 20 s and extension at 72°C for 30 s; final extension at 72°C for 10 min and storage at 4°C. The chosen barcode genes were: *trnH-psbA*, *matK*, *ITS*. Primers are described in detail in Table 3.5.

Table 3.5: Primer pair names, their length of nucleotide, their sequences, target gene names and Tm°C used in PCRs.

Primer pair name	Lenght of nucleotide (bp)	Sequence (5' - 3')	Target gene	Tm°C
<i>rbcL</i> F2	600	ATGTCACCACAAAC AGAGACTAAAGC	<i>rbcL</i>	57.2 °C
<i>rbcL</i> R3	600	CTTCTGCTACAAAT AAGAATCGATCTC	<i>rbcL</i>	57.2 °C
<i>trnH-psbA</i> F3	525	CGCGCATGGTGGAT TCACAATCC	<i>trnH-psbA</i>	54 °C
<i>trnH-psbA</i> R3	525	GTTATGCATGAACG TAATGCTC	<i>trnH-psbA</i>	54 °C
<i>matK</i> F1	834	GTTCTAGCACAAGA AAGTCGA	<i>matK</i>	54 °C (for Abant, Mollafenari and Gölcük populations) 52 °C (for Yedigöller, Düzce and Kefken populations)
<i>matK</i> R3	834	CCTATCCATCTGGA AATCTTG	<i>matK</i>	54 °C (for Abant, Mollafenari and Gölcük populations) 52 °C (for Yedigöller, Düzce and Kefken populations)
<i>ITS</i> F	650	TCCTCCGCTTATTGA TATGC	<i>ITS</i>	51 °C
<i>ITS</i> R	650	CCTTATCATTTAGA GGAAGGAG	<i>ITS</i>	51 °C

PCR products were run on 1% agarose gel using 1X TAE buffer (40mM Tris, 1mM EDTA, 20mM acetic acid, pH 8.5) containing 10 mg / ml of ethidium bromide, visualized and photographed under UV rays (Gel Doc 2000 BIO-RAD)].

3.6. Sequencing

As no amplification obtained with *rbcL* primers, the study was continued with other tested three primers (*trnH-psbA*, *matK* and *ITS*). PCR products (1 μ L) were treated with 0.5 μ L ExoSAP-IT and 5.5 μ L sterile distilled water for 15 minutes at 37 °C to remove unused nucleotides and primers. To the samples were added 1.5 μ L of Bright Dye Terminator (iCloning) were added and 1.5 μ L of the corresponding primer forward and then subjected to a new PCR process according to the following parameters: 25 cycles of denaturation at 95 ° C for 10 sc, annealing at 50 ° C for 5 s and final extension at 64 ° C for 4 min. The DNA was precipitated by adding 50 μ L of 95% ethanol and 2 μ L of 3M Sodium Acetate (pH 5.2). After incubation at room temperature for 15 minutes, the samples were centrifuged for 30 minutes at 2500 rpm / min. The pellets were resuspended in 150 μ L of 70% ethanol and then centrifuged for 30 minutes at 13000 rpm / min. Finally, DNA was resuspended in 20 μ L of 100% formamide. Sequencing was performed using 3100 Avant Genetic Analyzer (HITACHI, Applied Biosystem).

3.7. Data Analysis and Phylogenetic Tree Constraction

Obtained nucleotide sequences were visualized by BioEdit v7.0.5 program. They were compared with known sequences present in the GenBank, and matched with Nucleotide Basic Local Alignment Search Tool programme (BLASTN).

Graphical DNA barcoding representation of all individuals of 139 *P. avium* for studied DNA barcode genes were completed by using apex library that works on the basis of R statistical program.

Phylogenetic analyzes were conducted with phylip-3.69 software with its updated version of Felsenstein (1989). In particular, DNAdist program which is based on nucleotide evolution model of Kimura's 2-parameter (K2P) [Kimura, 1980] were used for Neighbor Joining method (NJ) [Saitou and Nei, 1987]. DNAmI program

which is a data specific model used for Maximum Likelihood method (ML) [Nei and Kumar, 2000]. The Bootstrap technique (SEQBOOT program, phylip-3.69) was used to validate tree topologies obtained from ML and NJ. Node support was assessed via 100 bootstrap replicates.

The analyzes were conducted either separately or by combining all the genes together. Phylogenetic trees were designed in phylip-3.69 according to a narrow consensus and then displayed through the TreeView32 program.

Obtained nucleotide sequences were also aligned by Mega 7 program (ClustalW) and all SNPs were detected. Data which is used for Structure analysis and PCA analysis were used on the basis of Mega7 program results.

3.7.1. Principal Component Analysis (PCA)

Principle component analysis (PCA) which allows grouping haplotypes in the spatial plane using sequence information of *trnH-psbA*, *matK* and *ITS* barcode genes, was carried out by using GenAlEx 6.4. programme. For the PCA analysis, the genetic distance (pairwise distance) between the samples was determined according to the Kimura 2-parameter model using the Mega 7 program. The obtained distance data was loaded into the programme GenAlEx 6.4 and the distance between the haplotypes was plotted graphically on the spatial plane (Figure 3.3).

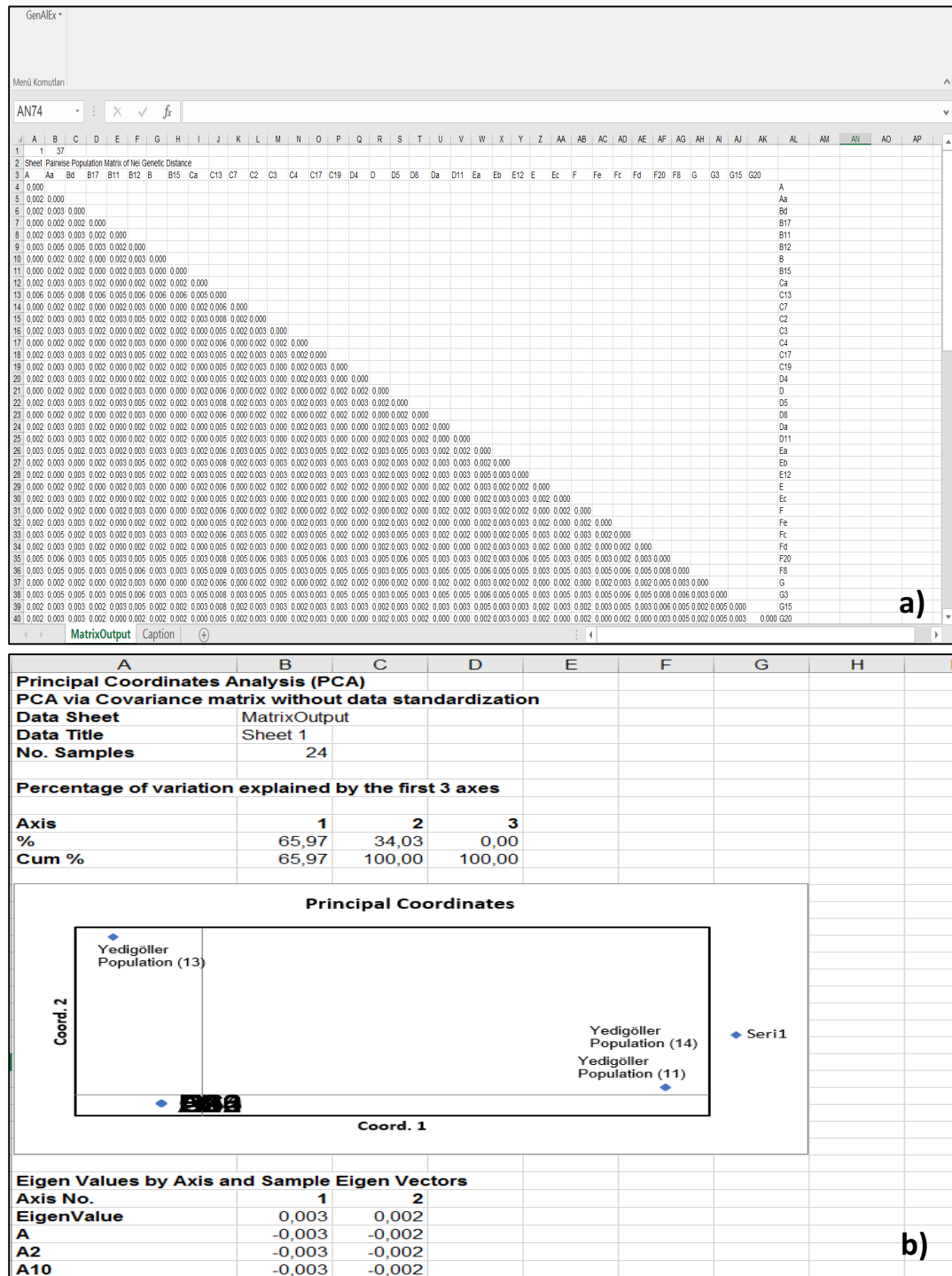


Figure 3.3: PCA analysis of *P. avium* haplotypes by using the data of the barcode genes (*trnH-psbA*, *matK* and *ITS*) used in the study by using “GenAlEx 6.4” programme. a) Performing PCA analysis of the data of pairwise distances which were calculated by “Kimura-2-parameter” by entering into the “GenAlEx 6.4” programme. b) At the end of PCA analysis, determination of the distances among the samples on the coordinate plane.

3.7.2. Structure Analysis

Sequences of *trnH-psbA*, *matK* and *ITS* DNA barcode genes from *P. avium* individuals for Structure analysis were aligned in the Mega 7 programme for identification of the *P. avium* haplotypes. Then, to be used as data in Structure analysis, SNP analyzes were performed by deleting columns with more than 5% spacing from aligned sequences, less than 5% SNPs and deleting all the columns without any SNPs. Specified SNPs were exported to the excel file and each bases were scored as 0 and 1 for non SNPs and SNPs, respectively. Structure analysis were carried out by using Structure 2.3.4. and Structure Harvester programmes (Figure 3.4).

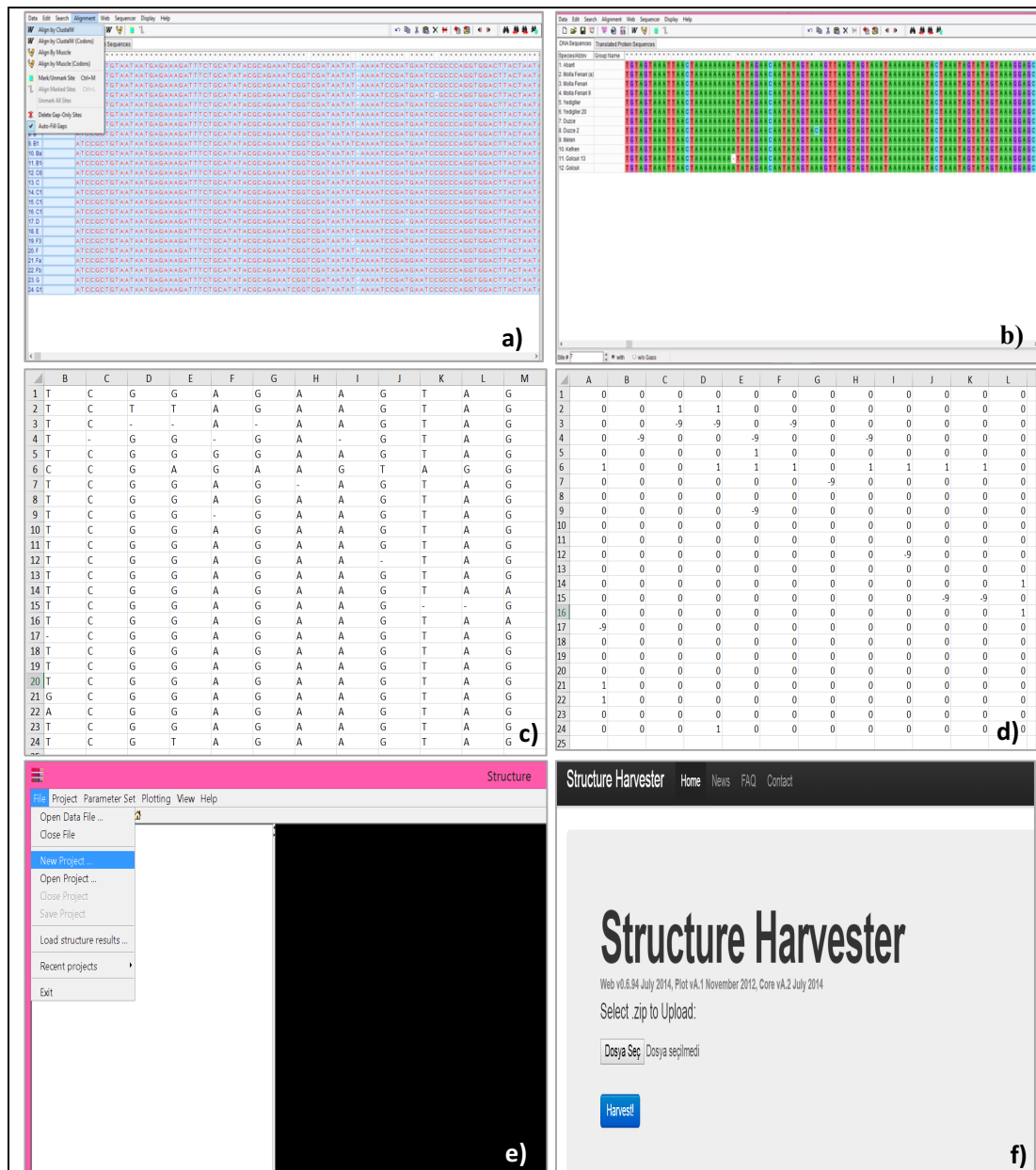


Figure 3.4: The structure analysis of *P. avium* haplotypes by using the data of the barcode genes (*trnH-psbA*, *matK* and *ITS*) used in the study by the help of Structure 2.3.4." and "Structure Harvester" programs. a) Alignment with "Mega 7" programme by ClustalW method. b) The overview of aligned in "Mega 7" programme. c) Transferring of the specified SNPs to the excel file. d) Scoring of "0", "1" and "-9" for each base as required data to be able to be entered into the programme "Structure 2.3.4". e) Entering the SNP data into the Structure programme. f) Finalizing the data in the "Structure Harvester" programme.

The sequences registered in the NCBI database and used in this work can be traced through their sequence code (ID).

- Accession numbers of *trnH-psbA* gene sequence which belongs to representative individuals for each subpopulation:

- *P. avium* Abant population (13) sequence code: MF431607;
- *P. avium* Mollafenari population (13) sequence code: MF590194;
- *P. avium* Yedigöller population (10) sequence code: MF590195;
- *P. avium* Düzce population (1) sequence code: MF590197;
- *P. avium* Melen population (1) sequence code: MF590196;
- *P. avium* Kefken population (4) sequence code: MF590198;
- *P. avium* Gölcük population (12) sequence code: MF590199

- Accession numbers of *matK* gene sequence which belongs to representative individuals for each subpopulation except Abant population:

- *P. avium* Abant population (3) sequence code: MF431608;
- *P. avium* Mollafenari population (10) sequence code: MF590188;
- *P. avium* Yedigöller population (10) sequence code: MF590189;
- *P. avium* Düzce population (16) sequence code: MF590190;
- *P. avium* Melen population (15) sequence code: MF590191;
- *P. avium* Kefken population (7) sequence code: MF590192;
- *P. avium* Gölcük population (6) sequence code: MF590193

- Accession numbers of *ITS* gene sequence which belongs to representative individuals for each subpopulation except Abant population:

- *P. avium* Abant population (17) sequence code: MF431606;
- *P. avium* Mollafenari population (18) sequence code: MF590200;
- *P. avium* Yedigöller population (9) sequence code: MF590201;
- *P. avium* Düzce population (17) sequence code: MF590202;
- *P. avium* Melen population (20) sequence code: MF590203;
- *P. avium* Kefken population (7) sequence code: MF590204;
- *P. avium* Gölcük population (10) sequence code: MF590205

4. RESULTS

4.1. DNA Extraction and PCR Reaction

Total genomic DNAs were extracted by using kit from the leaves of all individuals of collected Turkish *P. avium* (wild cherry) populations and one sample of *P. armeniaca* (apricot). According to the image of 1.5% agarose gel, it is shown (Figure 4.1) that all extracted DNAs were pure and adequate to carry on DNA barcoding studies.

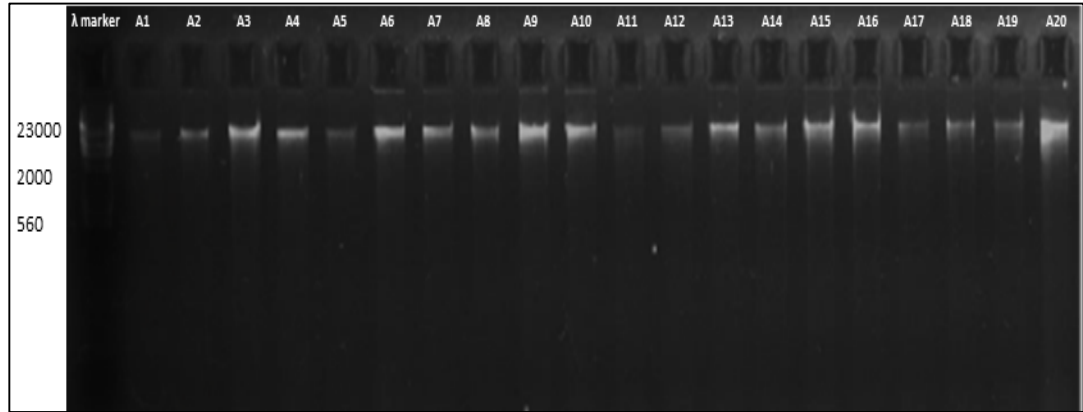


Figure 4.1: A representative photograph of extracted DNAs of 20 different individuals of Abant population. M: λ DNA-Hind III, A1-20: Abant populations.

Extracted DNAs have been used as a template for amplification, of specific barcode genes: *rbcL*, *matK*, *trnH-psbA* and *ITS* by PCR. PCR products of *trnH-psbA*, *matK* and *ITS* in different populations (Düzce, Mollafenari and Yedigöller) were visualized, and verified by their molecular weight on 1% agarose gel.

According to PCR results of 139 *P. avium* individuals, amplifications were obtained with *trnH-psbA* primers at a length of 270 bp (Figure 4.2), *matK* primers at a length of 643 bp (Figure 4.3), *ITS* primers at a length of 660 bp (Figure 4.4). However, no amplification was observed with *rbcL* primers (data not shown). Following the PCR reactions, the amplified gene regions were sequenced.

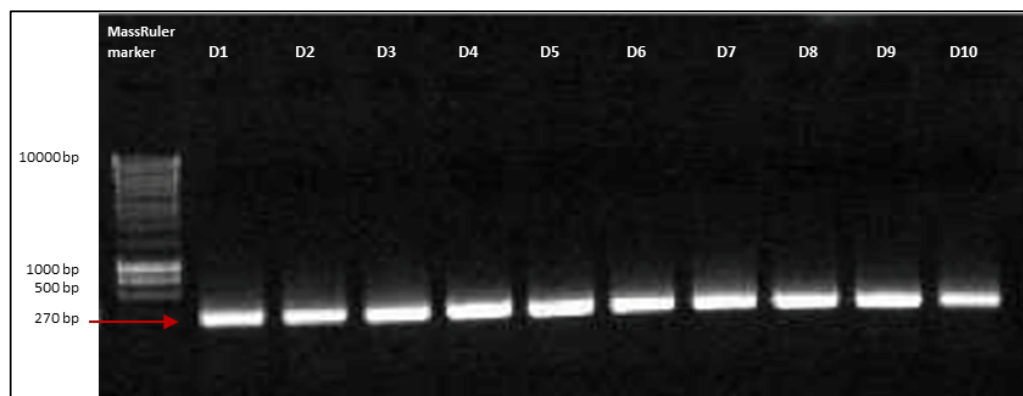


Figure 4.2: A representative figure of amplified sequences of *trnH-psbA* gene of Düzce population. M: MassRuler DNA ladder, D1-10: Düzce populations.

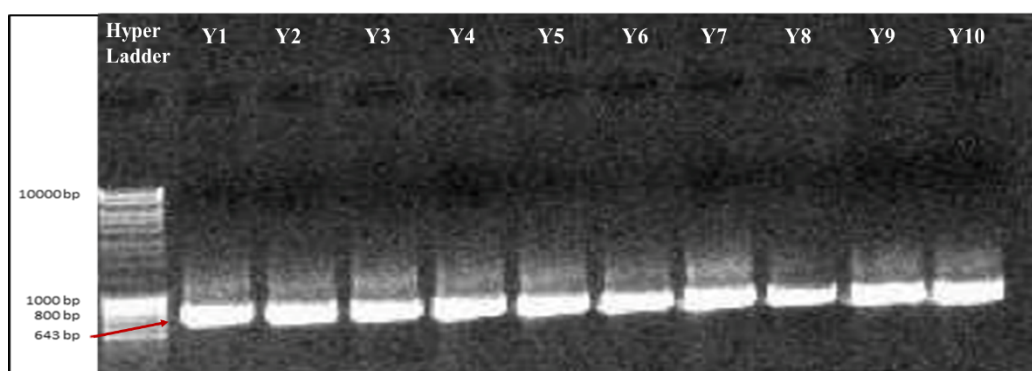


Figure 4.3: A representative figure of amplified sequences of *matK* gene of Yedigöller population. M: HyperLadder 1kb, Y1-10: Yedigöller populations.

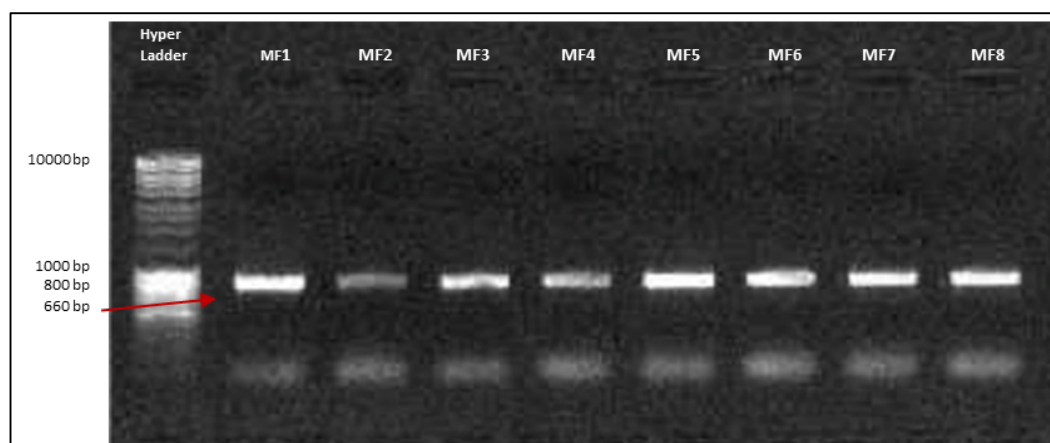


Figure 4.4: A representative figure of amplified sequences of *ITS* spacer of Mollafenari population. M: HyperLadder 1kb, MF1-8: Mollafenari populations.

4.2. DNA Sequencing (Barcoding) and Statistical Analysis

In this study, sequencing analysis of 139 *P. avium* individuals of the same species performed with *trnH-psbA*, *matK* and *ITS* genes. Results of sequence information for each sample and electropherograms of the sequences were obtained and representatively shown in the Figure 4.5. and Figure 4.6, respectively. According to the results of electropherograms, each base [adenine (A), guanine (G), cytosine (C) and thymine (T)] were detected by virtue of dideoxyribonucleoside triphosphates which were used as fluorescence dyed dNTPs. Adenine was detected as green color by virtue of dR6G dye, while guanine was detected as black color by virtue of dTAMRA dye, cytosine was detected as blue color by virtue of dR110 dye and thymine was detected as red color by virtue of dROX dye in the electropherograms (Figure 4.5). Then the sequence data were obtained based on electropherograms (Figure 4.6). The all DNA barcoding results of *trnH-psbA*, *matK* and *ITS* gene were given as FASTA format as supplementary data (S1, S2, S3).

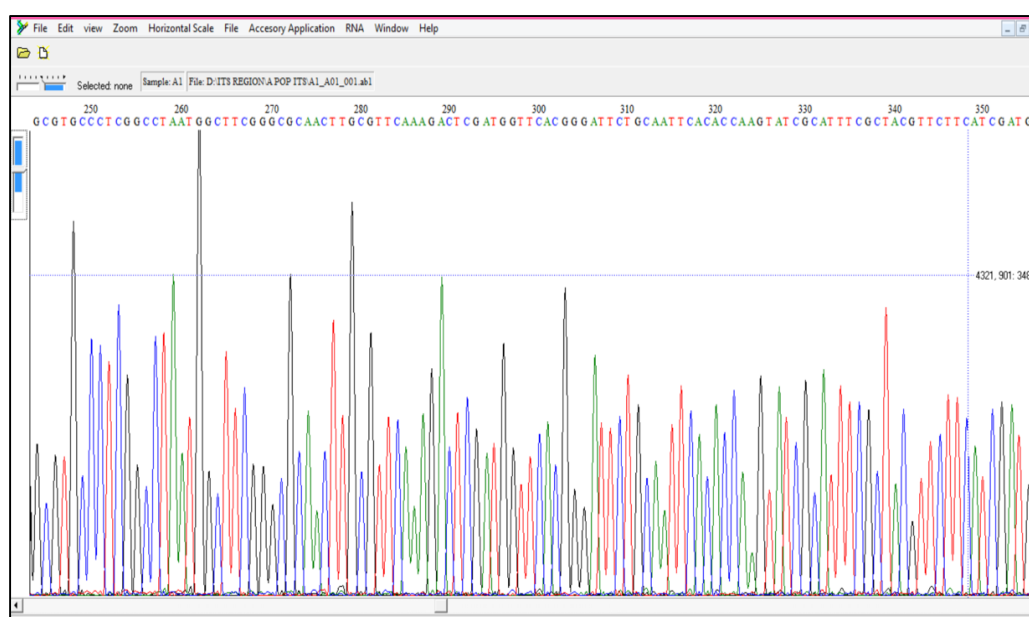


Figure 4.5: A representative screen display of electropherogram obtained as a result of DNA sequence analysis with BioEdit program.

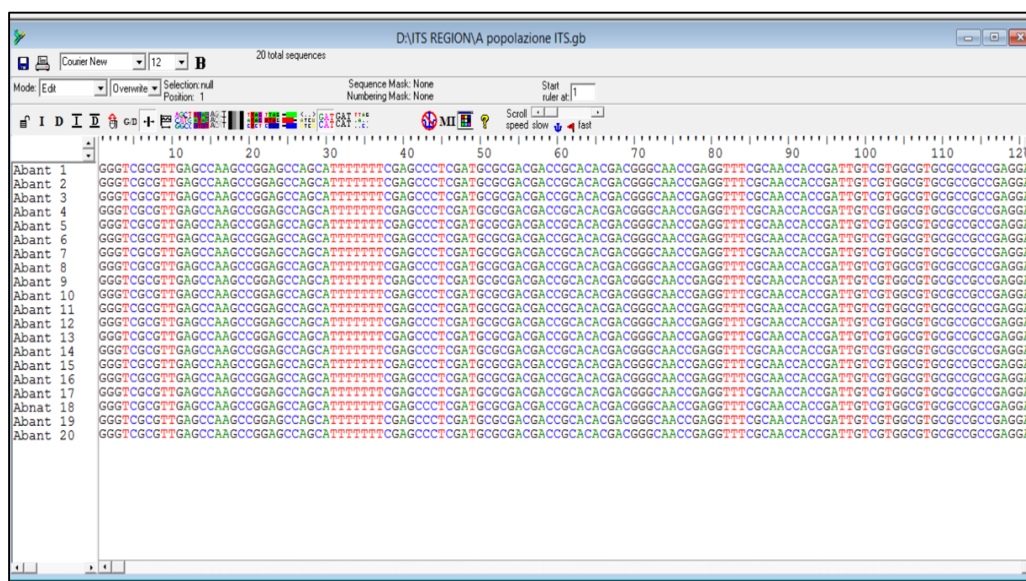


Figure 4.6: A representative screen display of the sequence data obtained as a result of DNA sequence analysis with BioEdit program.

As preliminary investigations, success rates of DNA barcodes (*trnH-psbA*, *matK* and *ITS*), (Table 4.1), molecular diversity parameters of each barcode genes (Table 4.2, 4.3 and 4.4, respectively) and percentage of variable sites obtained by aligning the sequenced regions (Table 4.5) were evaluated for each sequence obtained.

According to the results of PCR, while the success rates of *trnH-psbA* (270bp) and *ITS* (660bp) were obtained as 100%, the success rates of *rbcL* region and *matK* (643bp) were obtained as 0% and 89,93%, respectively. According to the results of sequencing, the success rates of *trnH-psbA* was 100%, whereas the success rates of *ITS* and *matK* were 95,2% and 97,13%, respectively (Table 4.1).

Table 4.1: The success rates of DNA barcodes within the investigated data-set.

	<i>rbcL</i>	<i>trnH-psbA</i>	<i>matK</i>	<i>ITS</i>
Investigated individual (N)	139	139	139	139
PCR success (%)	0 %	100 %	89,93 %	100 %
Sequencing success (%)	-	100 %	95,2 %	97,13 %
Sequenced length (bp)	-	270	643	660
Number of sequences of individuals obtained (S)	-	139	119	135

According to the obtained sequences of *trnH-psbA* gene, variable sites which are due to deletion, insertion, transition or transversion were determined for each population (Table 4.2). No variable site was observed for Abant, Melen and Kefken populations. In addition to that, while two variable sites were observed for Mollafenari, Düzce, Gölcük populations, three variable sites were observed for Yedigöller populations, respectively.

Table 4.2: Molecular diversity parameters for *trnH-psbA* gene.

TrnH-psbA	Abant	Mollafenari	Yedigöller	Düzce	Melen	Kefken	Gölcük
Total length (bp)	270	270	270	270	270	270	270
Conserved sites (bp)	270	269	268	269	270	270	269
Variable sites (bp)	0	1	2	1	0	0	1
Deletion sites (bp)	0	1	0	0	0	0	1
Insertion sites (bp)	0	0	0	0	0	0	0
Transitional sites (bp)	0	0	1	0	0	0	0
Transversional sites (bp)	0	0	1	1	0	0	0

According to the sequences of *matK* gene, no variable site was observed for Düzce and Melen populations, whereas more variable sites were observed for the rest of tested populations. While Abant and Mollafenari populations have the highest value of variable sites with the value of eight was obtained in both Abant and Mollafenari populations in comparison to other tested populations. On the contrary, Gölcük population has the lowest value of variable sites with the value of one (Table 4.3).

Table 4.3: Molecular diversity parameters for *matK* gene.

matK	Abant	Mollafenari	Yedigöller	Düzce	Melen	Kefken	Gölcük
Total length (bp)	643	643	643	643	643	643	643
Conserved sites (bp)	635	635	637	643	643	640	642
Variable sites (bp)	8	8	6	0	0	3	0
Deletion sites (bp)	7	11	11	0	0	9	1
Insertion sites (bp)	3	0	0	0	0	0	0
Transitional sites (bp)	1	0	3	0	0	0	0
Transversional sites (bp)	3	1	1	0	0	9	1

According to the obtained sequences of *ITS* gene, only one variable site was observed for Abant population, whereas more variable sites were observed for the rest of tested populations. Yedigöller population have the highest value of variable sites with the value of seven by comparison to the rest populations (Table 4.4).

Table 4.4: Molecular diversity parameters for *ITS* gene.

ITS	Abant	Mollafenari	Yedigöller	Düzce	Melen	Kefken	Gölcük
Total length (bp)	660	660	660	660	660	660	660
Conserved sites (bp)	659	654	653	656	657	654	654
Variable sites (bp)	1	6	7	4	3	6	6
Deletion sites (bp)	0	22	0	5	0	2	0
Insertion sites (bp)	0	0	7	0	0	0	0
Transitional sites (bp)	4	16	6	7	18	10	2
Transversional sites (bp)	0	1	3	0	1	3	2

When the observed variable sites were calculated on the basis of aligned sequence results, the total percentage value of 3.030 as the highest value and the total value of 1.661 as the lowest value were obtained for *ITS* and *trnH-psbA*, respectively.

Table 4.5: Percentage of variable sites of aligned sequences.

	Variable Sites %		
	<i>trnH-psbA</i> [270bp]	<i>matK</i> [643bp]	<i>ITS</i> [660bp]
Abant Population	0	1.088	0.151
Mollafenari Population	0.740	0.155	0.909
Yedigöller Population	0.307	0.933	1.060
Düzce Population	0.307	0	0.909
Melen Population	0	0	0.454
Kefken Population	0	0.466	0.909
Gölcük Population	0.307	0.155	0.909
Total	1.661	2.022	3.030

According to the sequence data, six haplotypes for *trnH-psbA* gene (Table 4.6), eighteen haplotypes for *matK* gene (Table 4.7) and twenty-nine haplotypes for *ITS* gene (Table 4.8) were found in the seven tested populations consisting of a total of 139 individuals.

As it can be evident from the Table 4.6 all individuals of Abant, Melen and Kefken populations were monomorphic and were represented by the haplotype H1 only. Nevertheless, Haplotype H1 which is peculiar to *trnH-psbA* gene is the most common haplotype, present in all seven populations.

Table 4.6: Haplotype composition of the seven wild populations of *P. avium* for *trnH-psbA* gene.

<i>trnH-psbA</i> Gene								
	Populations							
Haplotypes	Abant	Mollafenari	Yedigöller	Düzce	Melen	Kefken	Gölcük	Number of Individual
H1	20	18	18	19	20	20	19	134
H2	0	1	0	0	0	0	0	1
H3	0	1	0	0	0	0	0	1
H4	0	0	1	0	0	0	0	1
H5	0	0	0	0	0	0	1	1
H6	0	0	0	1	0	0	0	1
Total	20	20	19	20	20	20	20	139

Haplotype H1 which is peculiar to *matK* gene is the most common haplotype, present in all seven populations. Abant population has the maximum number of haplotypes (eight – H1, H2, H4, H5, H7, H9, H10 and H11), of which six (H4, H5, H7, H9, H10 and H11) are unique to it. In the case of Yedigöller population five haplotypes (H1, H14, H15, H16 and H17) of which H14, H15 and H16 which are unique were obtained. No unique haplotypes were found in Melen population (Table 4.7).

Table 4.7: Haplotype composition of the seven wild populations of *P. avium* for *matK* gene.

<i>matK</i> Gene								
	Populations							
Haplotypes	Abant	Mollafenari	Yedigöller	Düzce	Melen	Kefken	Gölcük	Number of Individual
H1	12	16	14	18	19	17	19	115
H2	2	1	0	0	0	1	0	4
H3	0	0	0	1	0	0	0	1
H4	1	0	0	0	0	0	0	1
H5	1	0	0	0	0	0	0	1
H6	0	0	0	0	0	0	1	1
H7	1	0	0	0	0	0	0	1

Table 4.7: continued.

H8	0	1	0	0	0	0	0	1
H9	1	0	0	0	0	0	0	1
H10	1	0	0	0	0	0	0	1
H11	1	0	0	0	0	0	0	1
H12	0	0	0	0	0	1	0	1
H13	0	0	0	0	0	1	0	1
H14	0	0	1	0	0	0	0	1
H15	0	0	1	0	0	0	0	1
H16	0	0	2	0	0	0	0	2
H17	0	1	1	1	1	0	0	4
H18	0	1	0	0	0	0	0	1
Total	20	20	19	20	20	20	20	139

Similarly, *trnH-psbA* and *matK* genes, the most common haplotype was haplotype H1 for *ITS* region that was present in all tested seven populations. Yedigöller population has the maximum number of haplotypes (nine – H1, H4, H8, H10, H11, H12, H13, H15 and H16), of which six (H10, H11, H12, H13, H15 and H16) are unique to it. Gölcük population contains five haplotypes (H1, H23, H28, H29 and H30) of which H23, H28, H29 and H30 are unique. No unique haplotypes were found in Abant population (Table 4.8).

Table 4.8: Haplotype composition of the seven wild populations of *P. avium* for *ITS* gene.

<i>ITS</i> Gene								
	Populations							
Haplotypes	Abant	Mollafenari	Yedigöller	Düzce	Melen	Kefken	Gölcük	Number of Individual
H1	19	13	11	14	16	15	16	104
H2	1	0	0	0	1	0	0	2
H3	0	1	1	0	0	0	0	2
H4	0	1	0	0	0	0	0	1
H5	0	1	0	0	0	0	0	1
H6	0	1	0	0	0	0	0	1
H7	0	2	1	0	0	0	0	3
H8	0	1	0	1	0	0	0	2
H9	0	0	1	0	0	0	0	1
H10	0	0	1	0	0	0	0	1
H11	0	0	1	0	0	0	0	1
H12	0	0	1	0	0	0	0	1
H13	0	0	0	0	0	1	0	1
H14	0	0	1	0	0	0	0	1
H15	0	0	1	0	0	0	0	1
H16	0	0	0	1	0	0	0	1
H17	0	0	0	1	0	0	0	1

Table 4.8: continued.

H18	0	0	0	1	0	0	0	1
H19	0	0	0	0	1	1	0	2
H20	0	0	0	1	0	0	0	1
H21	0	0	0	1	0	0	0	1
H22	0	0	0	0	0	0	1	1
H23	0	0	0	0	1	1	0	2
H24	0	0	0	0	1	0	0	1
H25	0	0	0	0	0	1	0	1
H26	0	0	0	0	0	1	0	1
H27	0	0	0	0	0	0	1	1
H28	0	0	0	0	0	0	1	1
H29	0	0	0	0	0	0	1	1
Total	20	20	19	20	20	20	20	139

The percentages of the frequency of the individual nitrogen bases (A, G, C and T) for amplified *trnH-psbA* gene sequence were presented in Table 4.9. The numbers following the population name represents the individual number in its population. Abant: represents all individuals of population; Mollafenari (a): represents (4), (6), (17) and (19) individuals of population; Mollafenari: represents all individuals of population except (9) and (a); Yedigöller: represents all individuals of population except (20); Düzce: represents all individuals of population except (2); Melen: represents all individuals of population; Kefken: represents all individuals of population; Gölcük: represents all individuals of population except (13). According to the reported values, it was observed that, *trnH-psbA* gene had A as the most represented base (43.33 – 43.7%) on a large scale, while C and G percentages were the lowest bases (12.22 – 12.97%).

Table 4.9: Percentages of the frequency of the individual nitrogen bases (A, G, C and T) for amplified *trnH-psbA* gene sequence.

<i>trnH-psbA</i>				
	A %	T %	G %	C %
Abant	43.70	31.11	12.22	12.97
Mollafenari (a)	43.33	31.48	12.22	12.97
Mollafenari (9)	43.50	31.20	12.30	13.00
Mollafenari	43.70	31.11	12.22	12.97
Yedigöller (20)	43.70	31.11	12.22	12.97
Yedigöller	43.70	31.11	12.22	12.97
Düzce (2)	43.33	31.11	12.22	13.34
Düzce	43.70	31.11	12.22	12.97
Melen	43.70	31.11	12.22	12.97
Kefken	43.70	31.11	12.22	12.97
Gölcük (13)	43.50	31.20	12.30	13.00
Gölcük	43.70	31.11	12.22	12.97

The percentages of the frequency of the individual nitrogen bases (A, G, C and T) for amplified *matK* gene sequence are shown in Table 4.10. The numbers following the population name represents the individual number in its population. Abant: represents all individuals of population except (2), (9), (10), (12), (13) and (16); Mollafenari: represents all individuals of population except (1), (14) and (b); Mollafenari (b): represents (2), (4), (8) and (9) individuals of population; Yedigöller: represents all individuals of population except (6), (11), (13) and (14); Düzce: represents all individuals of population; Melen: represents all individuals of population; Kefken: represents all individuals of population except (3), (a), and (b); Kefken (a): represents (14), (15) and (15) individuals of population; Kefken (b): represents (12) and (17) individuals of population; Gölcük: represents all individuals of population except (1). For *matK* gene, A (36.76 – 37.07%) and T (28.60 – 29.14%) were dominant basis when they were compared with the other purine (17.40 – 17.92%) and pyrimidine (16.44 – 16.72%).

Table 4.10: Percentages of the frequency of the individual nitrogen bases (A, G, C and T) for amplified *matK* gene sequence.

matK				
	A %	T %	G %	C %
Abant	36.90	29.14	17.45	16.51
Abant (2)	37.05	28.95	17.40	16.60
Abant (10)	36.77	28.95	17.84	16.44
Abant (9)	36.76	28.81	17.92	16.51
Abant (12)	37.00	28.78	17.57	16.65
Abant (13)	36.90	28.81	17.78	16.51
Abant (16)	36.83	28.86	17.78	16.53
Mollafenari (1)	36.76	28.81	17.78	16.65
Mollafenari	36.84	28.78	17.73	16.65
Mollafenari (b)	36.90	28.81	17.78	16.51
Mollafenari (14)	37.00	28.78	17.73	16.49
Yedigöller (6)	37.00	28.78	17.57	16.65
Yedigöller	36.84	28.78	17.73	16.65
Yedigöller (13)	37.07	28.64	17.78	16.51
Yedigöller (14)	36.84	28.61	17.90	16.65
Yedigöller (11)	36.90	28.81	17.78	16.51
Düzce	36.90	28.81	17.78	16.51
Melen	36.90	28.97	17.62	16.51
Kefken (3)	36.84	28.78	17.73	16.65
Kefken	37.00	28.78	17.57	16.65
Kefken (a)	36.87	28.60	17.81	16.72
Kefken (b)	36.97	28.86	17.63	16.54
Gölcük	37.07	28.64	17.78	16.51
Gölcük (1)	37.17	28.61	17.73	16.49

The percentages of the frequency of the individual nitrogen bases (A, G, C and T) for amplified *ITS* spacer sequence are shown in Table 4.10. The numbers following the population name represents the individual number in its population. Abant: represents all individuals of population except (a); Abant (a): represents (4), (5), (11) and (19) individuals of population; Mollafenari (c): represents (8), (9), (10) and (16) individuals of population; Mollafenari (d): represents (1), (2), (3), (4), (19) and (20) individuals of population; Mollafenari: represents all individuals of population except (c), (d), (11), (12), (15) and (17); Yedigöller (a): represents (1), (5), (6), (9), (10), (11), (12), (14), (15), (16), (18) and (20) individuals of population; Düzce: represents all individuals of population except (a), (4), (5), (8) and (11) individuals of population; Düzce (a): represents (9), (13) and (19) individuals of population; Melen (a): represents (1), (4), (5), (10), (16) and (17) individuals of population, Melen (b): represents (15) and (19) individuals of population, Melen (c): represents (3), (6), (11) and (13) individuals of population; Melen: represents all individuals of population except (a), (b), (c), (12) individuals of population; Kefken (c): represents (2) and (12) individuals of population; Kefken (d): represents (6) and (17) individuals of population; Kefken (e): represents (4), (5), (7), (9), (10), (11), (14), (19) individuals of population; Kefken: represents all individuals of population except (c), (d), (e), (8) and (20); Gölcük: represents all individuals of population except (3), (15) and (20) individuals of population. It was observed that C was the most represented base (30,30 - 30,92%), while the lowest represented base was A (19.24 – 19.69%) for *ITS* spacer.

Table 4.11: Percentages of the frequency of the individual nitrogen bases (A, G, C and T) for amplified *ITS* spacer sequence.

	<i>ITS</i>			
	A %	T %	G %	C %
Abant	19.55	21.21	28.79	30.45
Abant (a)	19.55	21.36	28.79	30.30
Mollafenari (d)	19.42	21.09	28.83	30.66
Mollafenari (17)	19.30	21.27	28.87	30.56
Mollafenari (11)	19.27	21.24	28.98	30.51
Mollafenari (c)	19.27	21.09	28.98	30.66
Mollafenari (12)	19.27	21.39	28.83	30.51
Mollafenari (15)	19.45	21.27	28.72	30.56
Mollafenari	19.42	21.24	28.83	30.51
Yedigöller (a)	19.30	21.27	28.87	30.56
Yedigöller (13)	19.27	21.54	28.68	30.51
Yedigöller (7)	19.45	21.27	28.72	30.56

Table 4.11: continued.

Yedigöller (2)	19.69	21.23	28.63	30.45
Yedigöller (3)	19.27	21.24	28.98	30.51
Yedigöller (4)	19.42	21.24	28.83	30.51
Yedigöller (17)	19.42	21.39	28.68	30.51
Yedigöller (19)	19.42	21.24	28.83	30.51
Düzce (4)	19.33	21.30	28.92	30.45
Düzce	19.45	21.27	28.72	30.56
Düzce (5)	19.63	21.15	28.62	30.60
Düzce (8)	19.48	21.30	28.77	30.45
Düzce (a)	19.30	21.27	28.87	30.56
Düzce (11)	19.33	21.30	28.77	30.60
Melen (a)	19.40	21.06	28.94	30.60
Melen (b)	19.55	21.06	28.79	30.60
Melen (12)	19.55	21.36	28.79	30.30
Melen	19.55	21.21	28.79	30.45
Melen (c)	19.40	21.21	28.94	30.45
Kefken	19.55	21.21	28.79	30.45
Kefken (e)	19.40	21.21	28.94	30.45
Kefken (c)	19.40	21.06	28.94	30.60
Kefken (d)	19.42	21.24	28.83	30.51
Kefken (20)	19.24	21.01	28.95	30.80
Kefken (8)	19.55	21.21	28.79	30.45
Gölcük	19.40	21.21	28.79	30.60
Gölcük (3)	19.40	21.21	28.47	30.92
Gölcük (15)	19.40	21.36	28.79	30.45
Gölcük (20)	19.24	21.21	28.95	30.60

Moreover, *TrnH-psbA*, *matK* and *ITS* regions were identified as highly conserved gene, as revealed by the percentage of identity (between the ranges from 98 to 100%) of the matches between samples and registered other accessions of *P. avium* in NCBI, nucleotide BLAST database, according to BLAST analysis.

As shown in Figure 4.7, 4.8, 4.9 and 4.10, nucleotide sequences were used for graphical DNA barcoding representation of all individuals of 139 *P. avium* for studied DNA barcode genes (*trnH-psbA*, *matK* and *ITS*) as well as multilocus which were generated by using apex library works on the basis of R statistical program. The variable regions can visually be determined in the all graphical DNA barcoding figures.

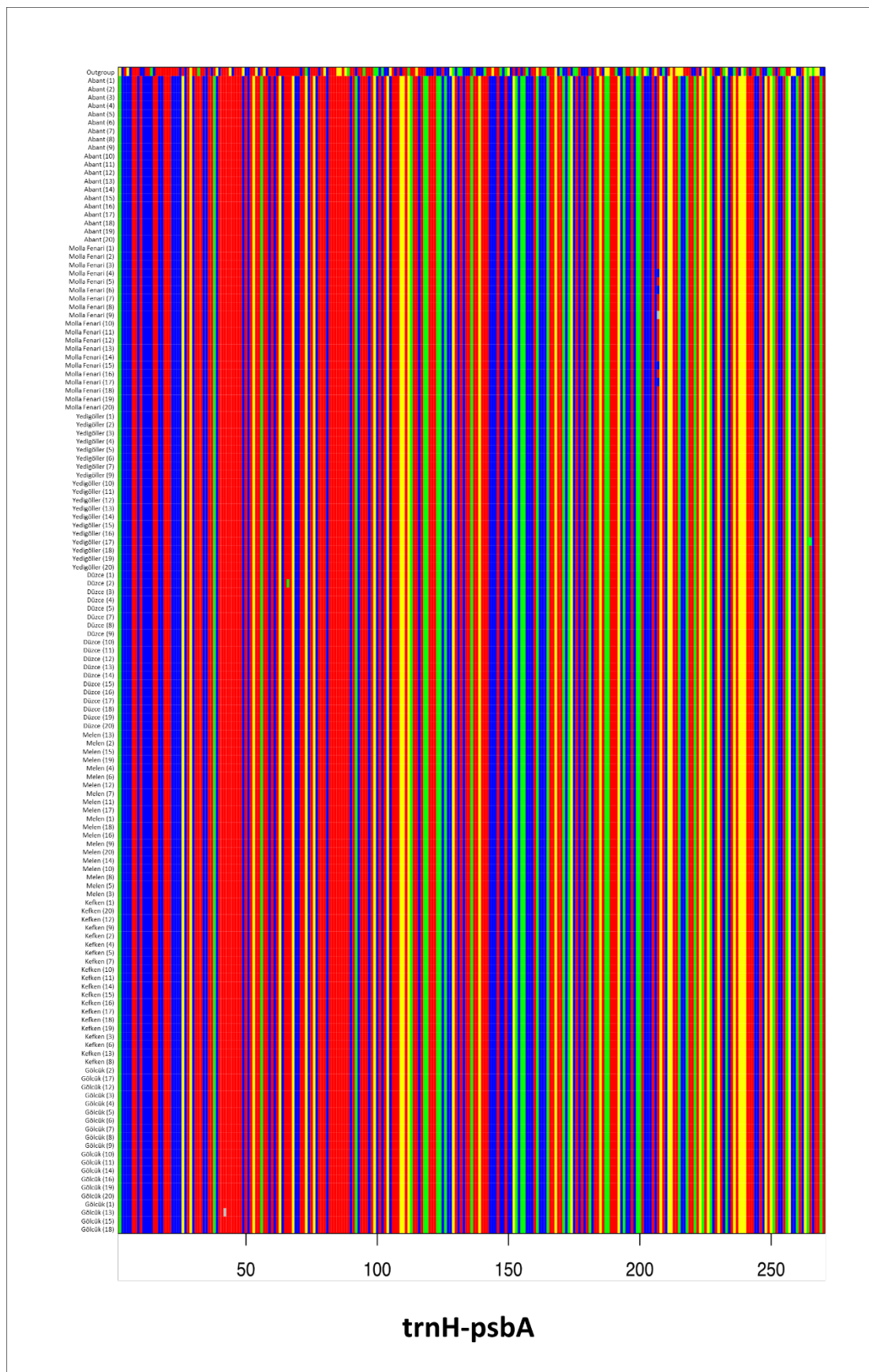


Figure 4.7: Graphical representation of 139 *P. avium* + *P. armeniaca* (outgroup) for DNA barcode of *trnH-psbA*.



Figure 4.8: Graphical representation of 139 *P. avium* + *P. armeniaca* (outgroup) for DNA barcode of *matK*.



Figure 4.9: Graphical representation of 139 *P. avium* + *P. armeniaca* (outgroup) for DNA barcode of ITS.



Figure 4.10: Graphical representation of 139 *P. avium* + *P. armeniaca* (outgroup) for DNA barcode of multilocus *trnH-psbA+matK+ITS*.

4.3. Phylogenetic Analysis

Both NJ (Figure 4.11) and ML methods (Figure 4.12), used for *trnH-psbA* sequences of our samples have shown similarities with each other. *P. armeniaca* was always considered as an out group so it separated primarily from the all *P. avium* tested populations with a bootstrap value of 100%. In addition to that, individual of *P. avium* Düzce population (2) was separated from the rest in both methods. NJ and ML methods fragmented the pair of *P. avium* Mollafenari (9) and *P. avium* Mollafenari (a) together with a bootstrap value of 53% and 56%, respectively. The other populations showed different placements which are not sharp on the phylogenetic trees in two different methods (Figure 4.11, 4.12).

When the two different phylogenetic methods were handled for *matK* sequences of our samples, it was obvious that *P. avium* Abant (13) was separated primarily following *P. armeniaca* with a bootstrap value of 100%. In addition to that, *P. avium* Kefken (b) moved away from the all other populations with bootstrap value of 33% in NJ method (Figure 4.13). *P. avium* Yedigöller (11) and *P. avium* Yedigöller (14) located together on the both phylogenetic trees with a bootstrap value of 63% and 72%, respectively. Phylogenetic methods NJ and ML fragmented the rest in a slightly different order (Figure 4.13, 4.14).

The most significant similarity was obtained for all subgroups of Gölcük population positioned in the same node in both phylogenetic methods, used for *ITS* sequences of tested samples. Whilst they positioned right after out group *P. armeniaca* in NJ method with a bootstrap value of 100%. In addition to that, individuals of Yedigöller population (17) and Yedigöller population (13) shared the same node with individuals of Melen population (12) and Abant population (a) in both two methods with bootstrap value of 40% and 8%, respectively. NJ and ML methods fragmented the set of Düzce population (8), Düzce population (5), Düzce population and Mollafenari population (15) together. It is also observed that Yedigöller population (4), Yedigöller population (2), Mollafenari population and Melen population were location closely (Figure 4.15, 4.16).

The NJ and ML studies according to the informational content of multilocus *trnH-psbA* + *matK* + *ITS* obtained by examining the set of all amplified nucleotide sequenced (for a total of 1573 bp) have outlined substantially different phylogenetic

profiles (Figure 4.17, 4.18). For multilocus trees, the numbers following the population name represents the individual number in its population. Abant: represents all individuals of population except (2), (9), (10), (12), (13), (16) and (x); Abant (x): represents (4), (5), (11), (19) and (20) individuals of population; Mollafenari: represents all individuals of population except (4), (6), (9), (12), (15), (17), (19), (x) and (y); Mollafenari (x): represents (8), (10) and (16) individuals of population; Mollafenari (y): represents (1), (2), (3) and (20) individuals of population; Mollafenari (y): represents all individuals of population except (c), (d), (11), (12), (15) and (17); Yedigöller: represents all individuals of population except (1), (2), (3), (4), (6), (7), (13), (19), (20), (x) and (y); Yedigöller (x) represents (11) and (14) individuals of population; Yedigöller (y) represents (1), (5), (9), (10), (12), (15), (16) and (18); Düzce: represents all individuals of population except (5), (8), (11) and (x) individuals of population; Düzce (x): represents (9), (13) and (19) individuals of population; Melen: represents all individuals of population except (12), (x), (y) and (z); Melen (x): represents (1), (5), (10), (16) and (17) individuals of population, Melen (y): represents (6) and (13) individuals of population, Melen (z): represents (15) and (19) individuals of population; Kefken: represents all individuals of population except (2), (3), (6), (12), (17), (x) and (y); Kefken (x): represents (15) and (16) individuals of population; Kefken (y): represents (4), (5), (7), (9), (10), (11) and (19) individuals of population; Gölcük: represents all individuals of population except (3) and (13) individuals of population.

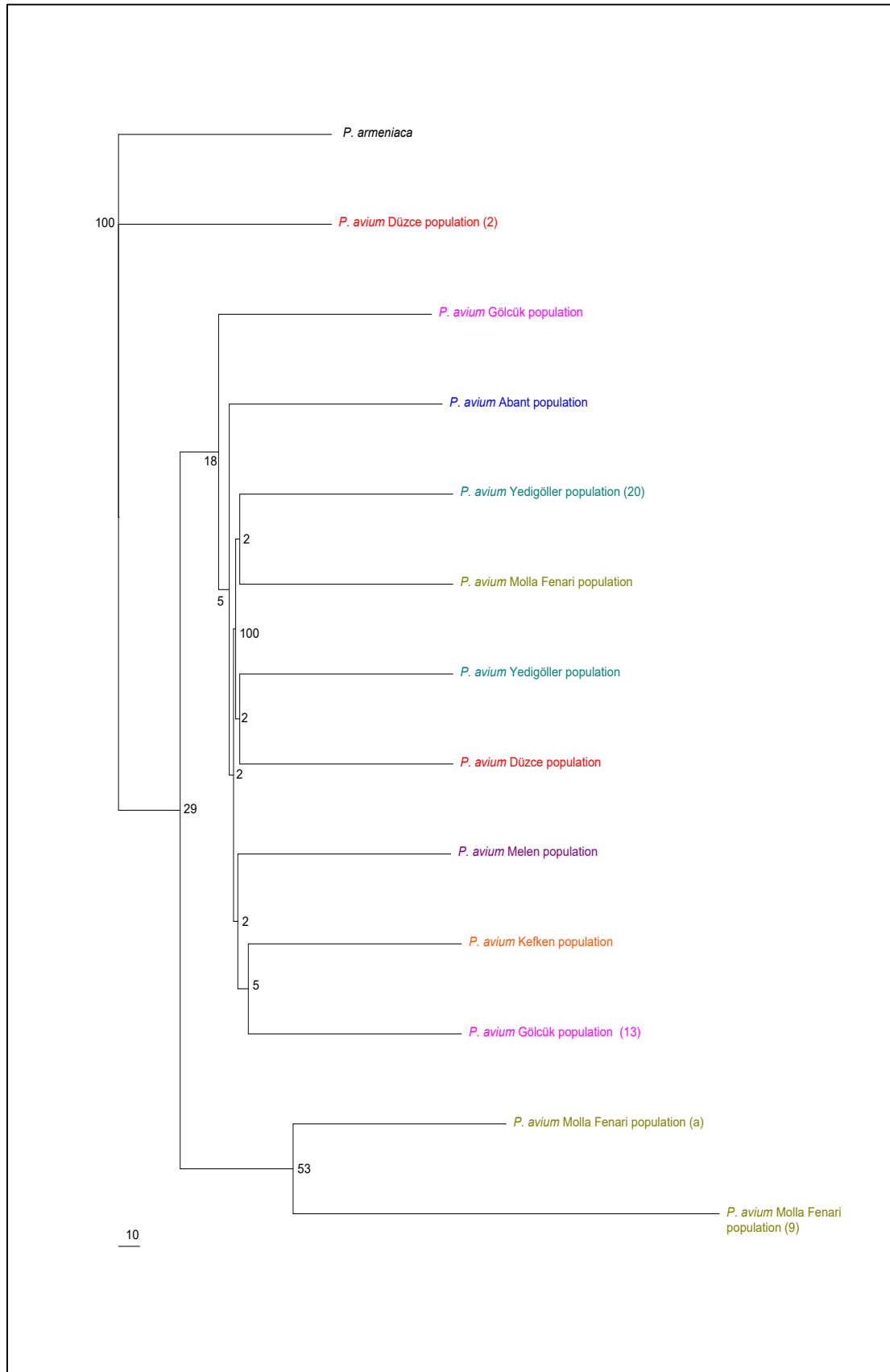


Figure 4.11: Bootstrap consensus tree generated by NJ method for the *trnH-psbA* sequences obtained for tested *P. avium* populations.

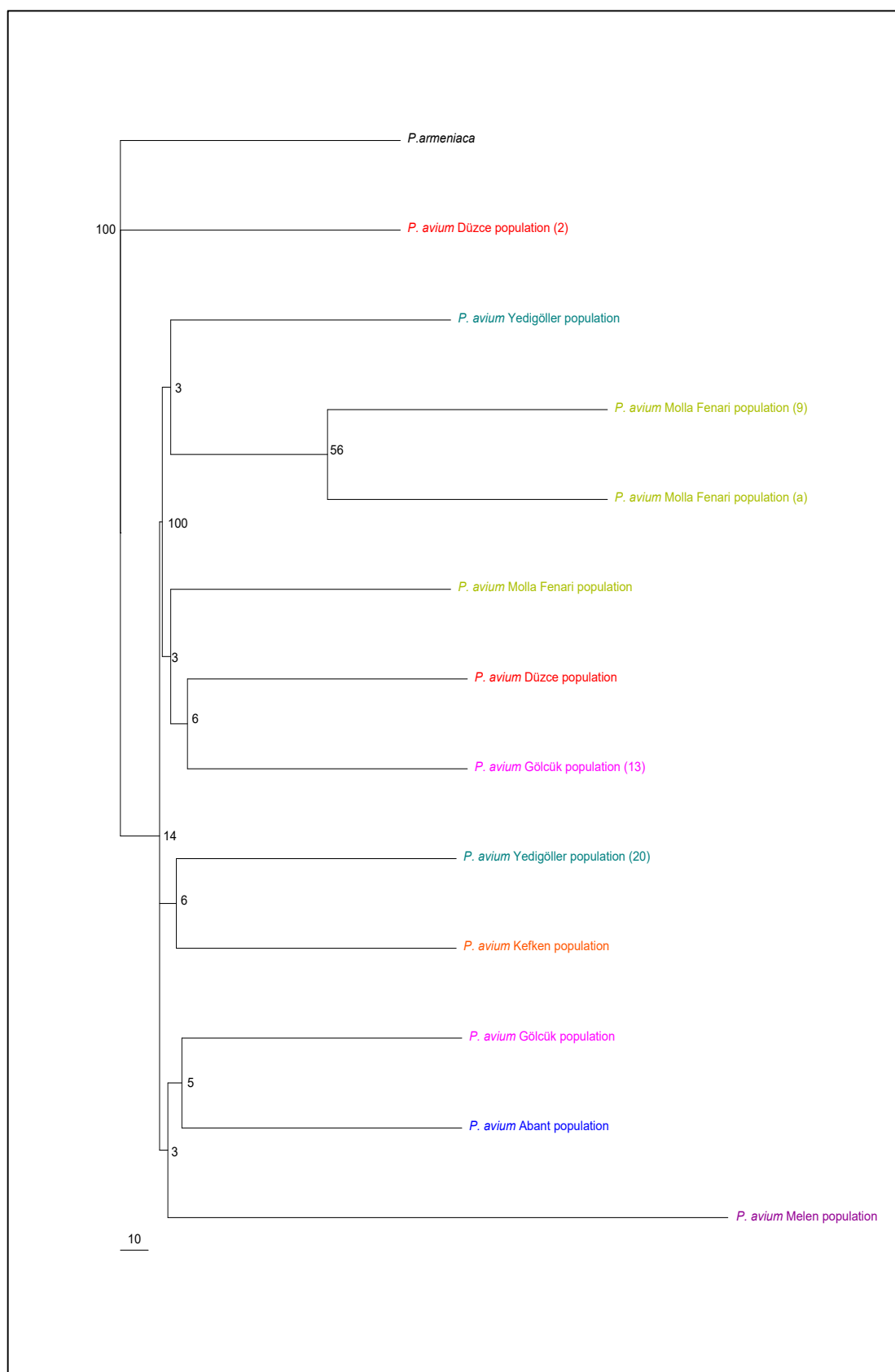
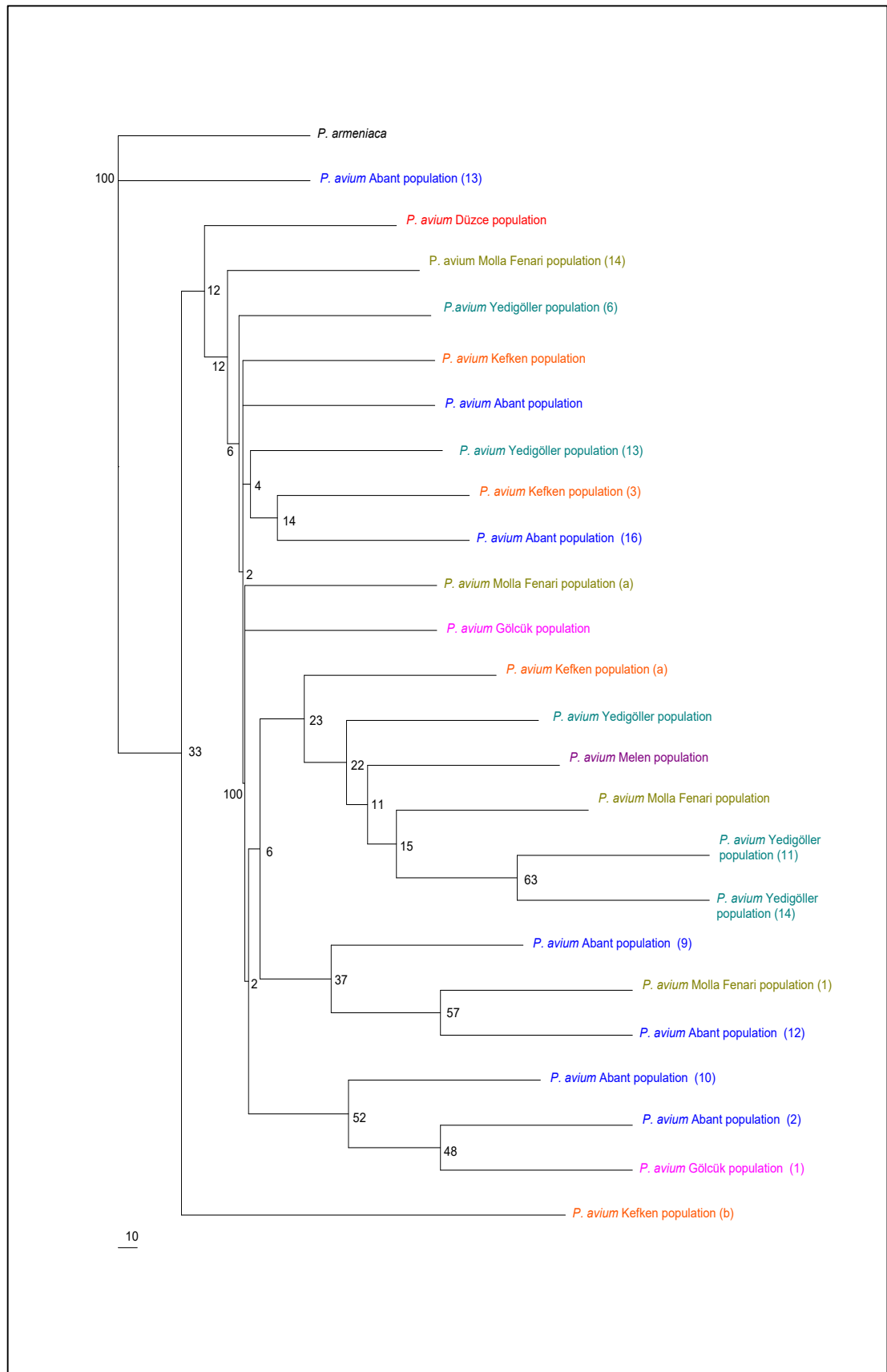


Figure 4.12: Bootstrap consensus tree generated by ML method for the *trnH-psbA* sequences obtained for tested *P. avium* populations.



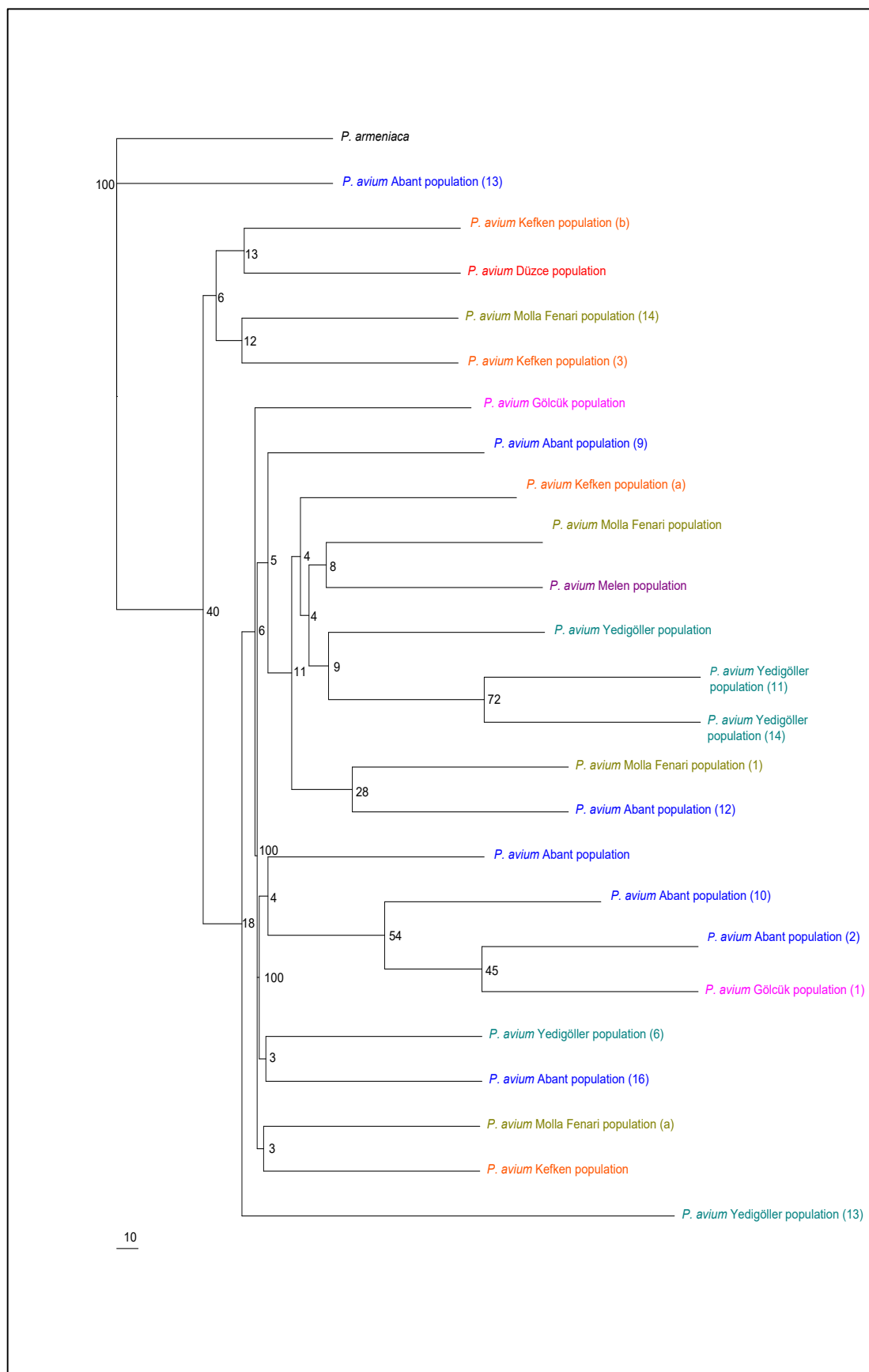


Figure 4.14: Bootstrap consensus tree generated by ML method for the *matK* sequences obtained for tested *P. avium* populations.

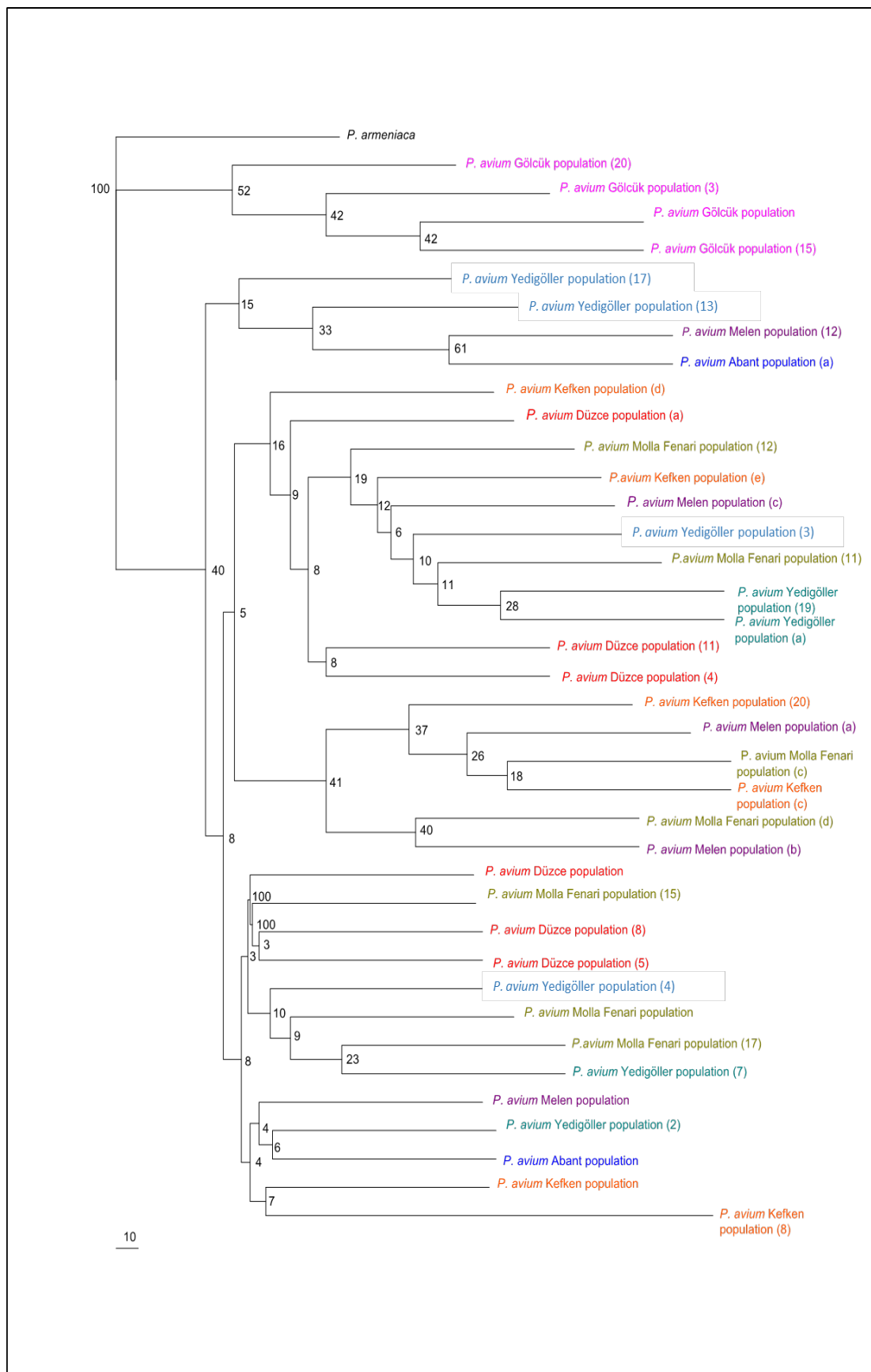


Figure 4.15: Bootstrap consensus tree generated by NJ method for the *ITS* sequences obtained for tested *P. avium* populations.

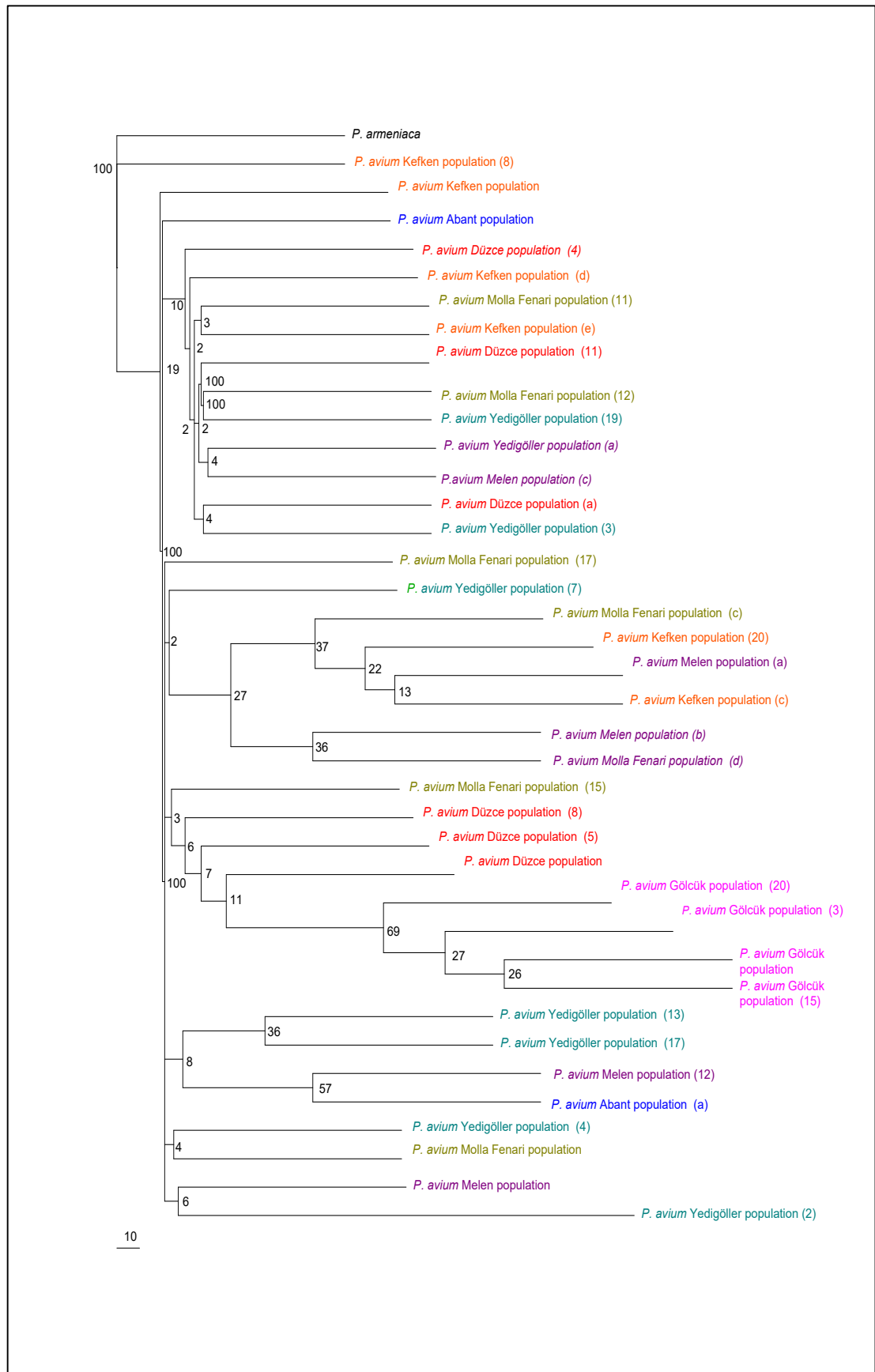


Figure 4.16: Bootstrap consensus tree generated by ML method for the ITS sequences obtained for tested *P. avium* populations.

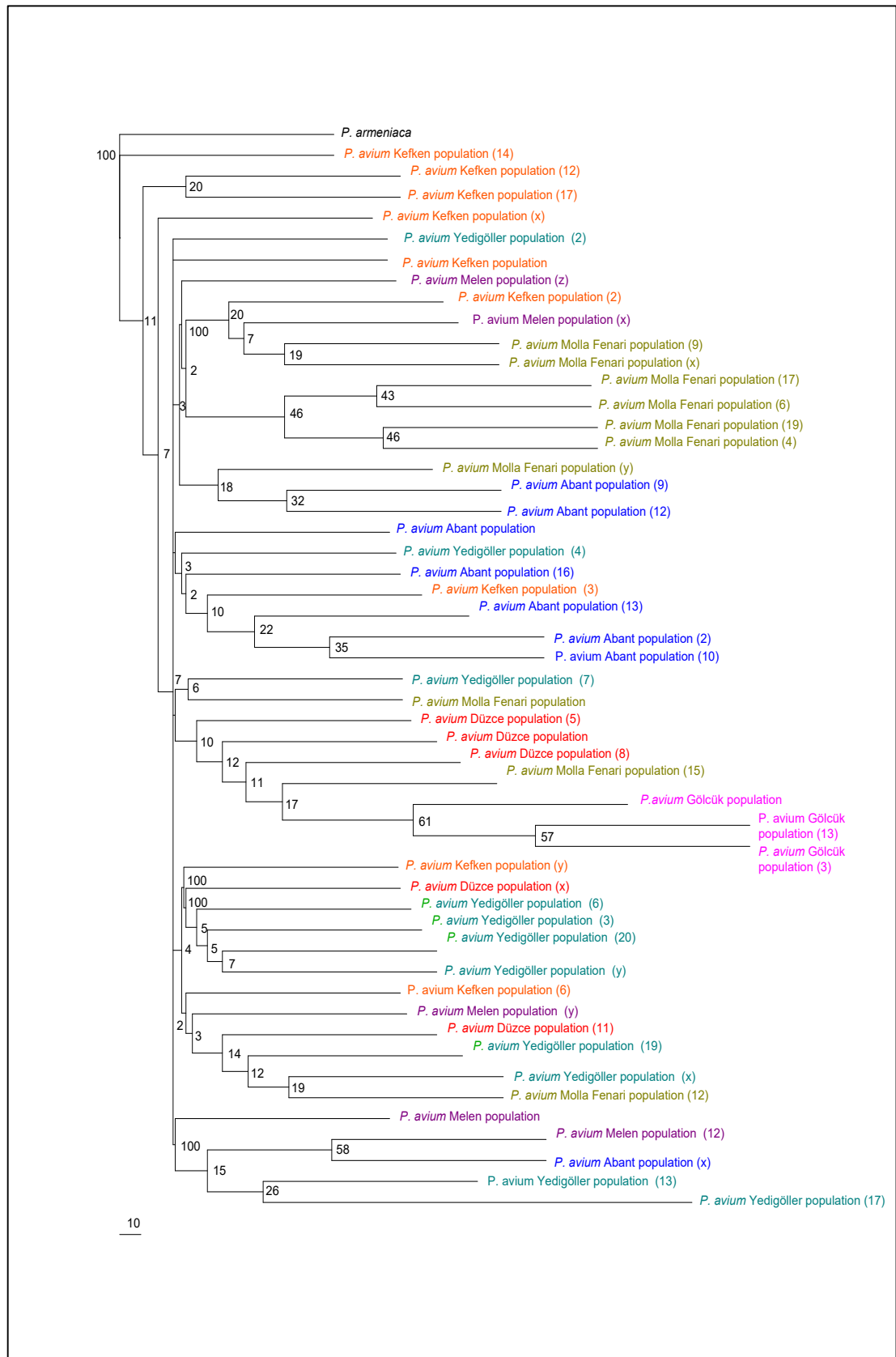


Figure 4.17: Bootstrap consensus tree generated by the NJ method for the *trnH-psbA+matK+ITS* combined sequences obtained for tested *P. avium* populations.

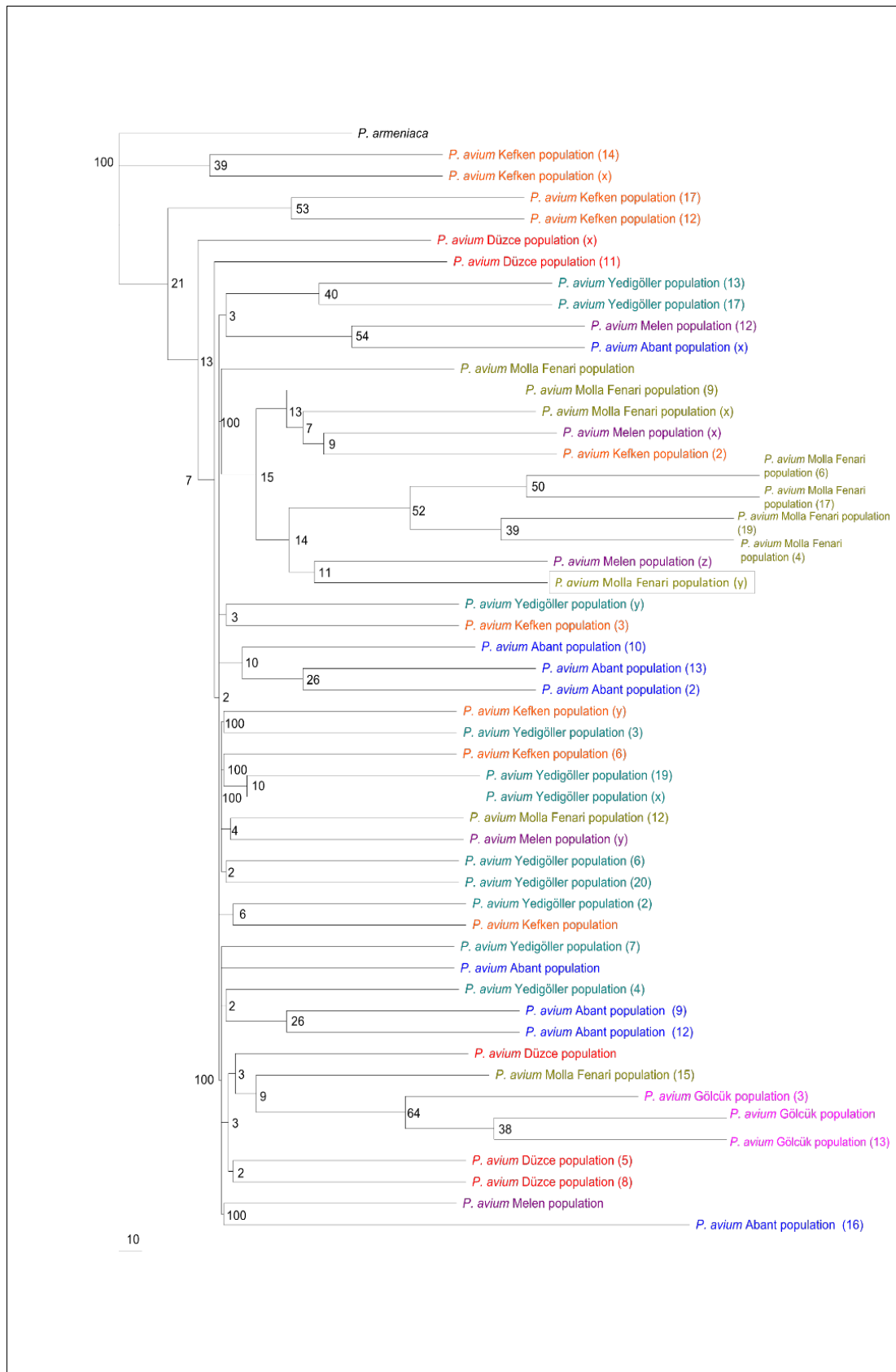


Figure 4.18: Bootstrap consensus tree generated by ML method for the *trnH-psbA*+*matK*+*ITS* combined sequences obtained for tested *P. avium* populations.

4.4. Principle Component Analysis (PCA)

By "GenAlEx 6.4" program, the distances between the haplotypes were plotted graphically on the spatial plane according to the "Kimura 2-parameter" model.

According to the results of PCA analysis for *trnH-psbA* gene which is represented in Figure 4.19., the haplotypes separated as three groups; i) twentieth individual of Yedigöller population, ii) second individual of Düzce population and iii) rest of the other individuals of all tested populations.

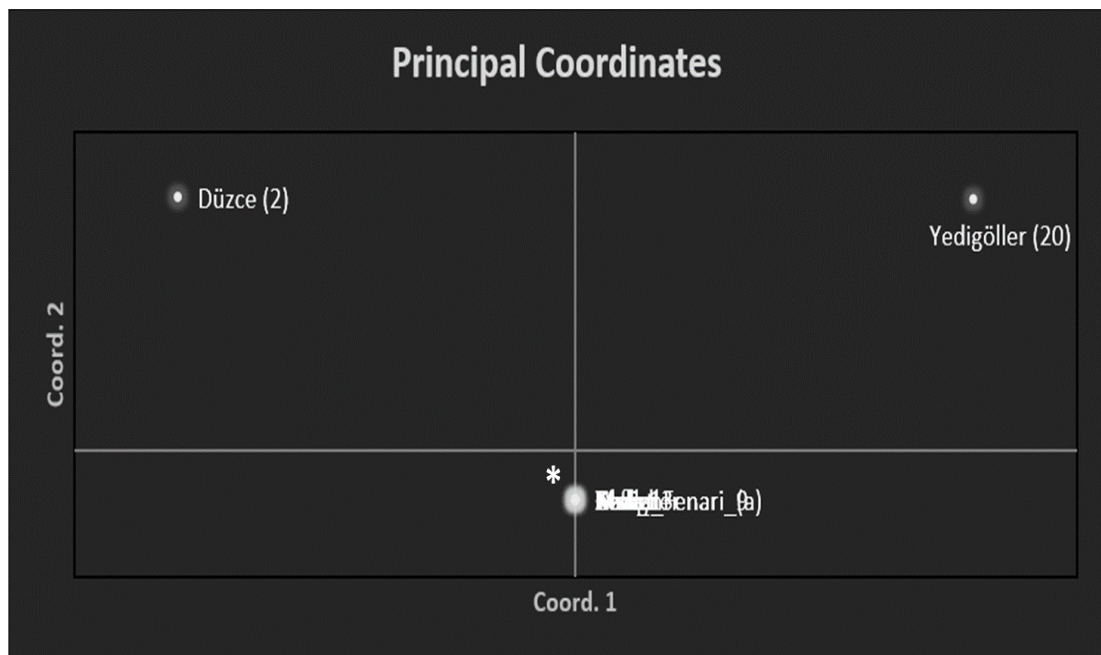


Figure 4.19: PCA analysis of *P. avium* L. haplotypes of the *trnH-psbA* gene obtained from "GenAlEx 6.4" program. The densities of the points representing the groups are directly proportional to the number of individuals. The symbol of * is represents all individuals except individual of Düzce (2) and individual of Yedigöller (20).

According to the results of PCA analysis for *matK* gene, which is represented in Figure 4.20, the haplotypes separated as three groups; i) eleventh and fourteenth individuals of Yedigöller population, ii) thirteenth individual of Yedigöller, iii) rest of the other individuals of all populations.

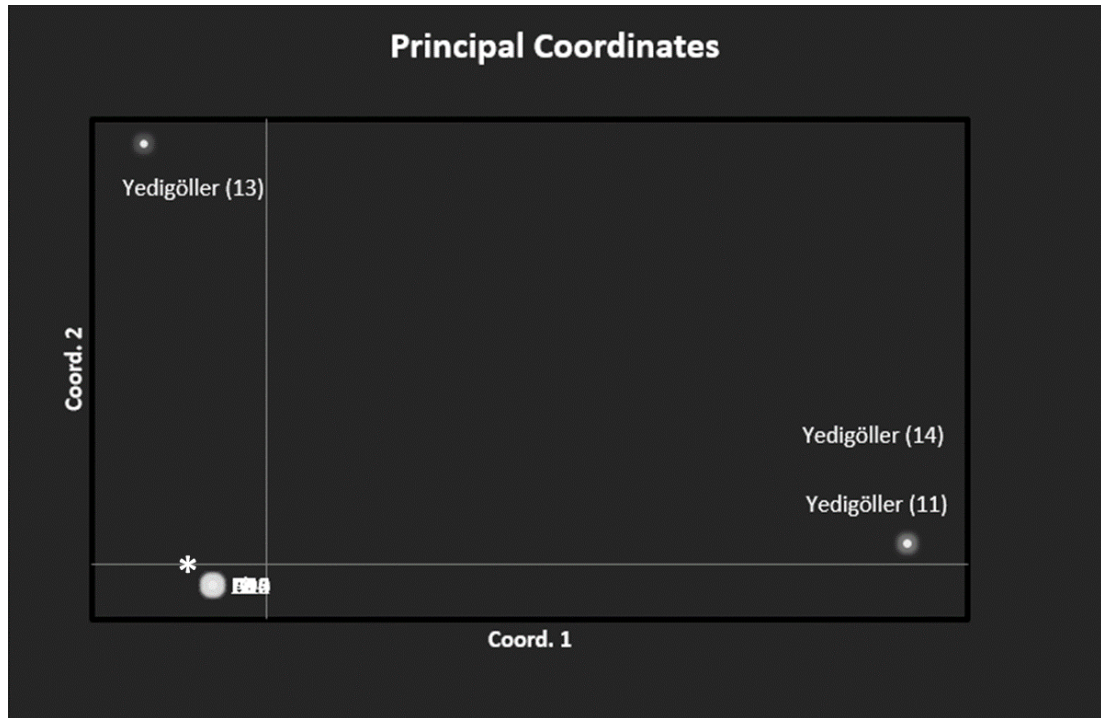


Figure 4.20: PCA analysis of *P. avium* L. haplotypes of the *matK* gene obtained from "GenAlEx 6.4" program. The densities of the points representing the groups are directly proportional to the number of individuals. The symbol of * is represents all individuals except individual of Yedigöller (13), individual of Yedigöller (11) and individual of Yedigöller (14).

According to the results of PCA analysis for *ITS* gene, which is represented in Figure 4.21, the haplotypes separated as eleven groups. While only eighth individual of Kefken population which is located very far from the all individuals, third individual of Gölcük population and seventeenth individual of Yedigöller population constitute different particular groups by themselves, individuals of d group of Mollafenari population and individuals of b group of Melen population constitute a separate group together. Some individuals of Yedigöller, Düzce, Mollafenari and Melen populations constitute five different groups which are closely located to each other and far located from all the other groups. All the remaining haplotypes which is difficult to particularly identify individuals in the graph formed another two groups which are very close the singular groups of Gölcük and Yedigöller.

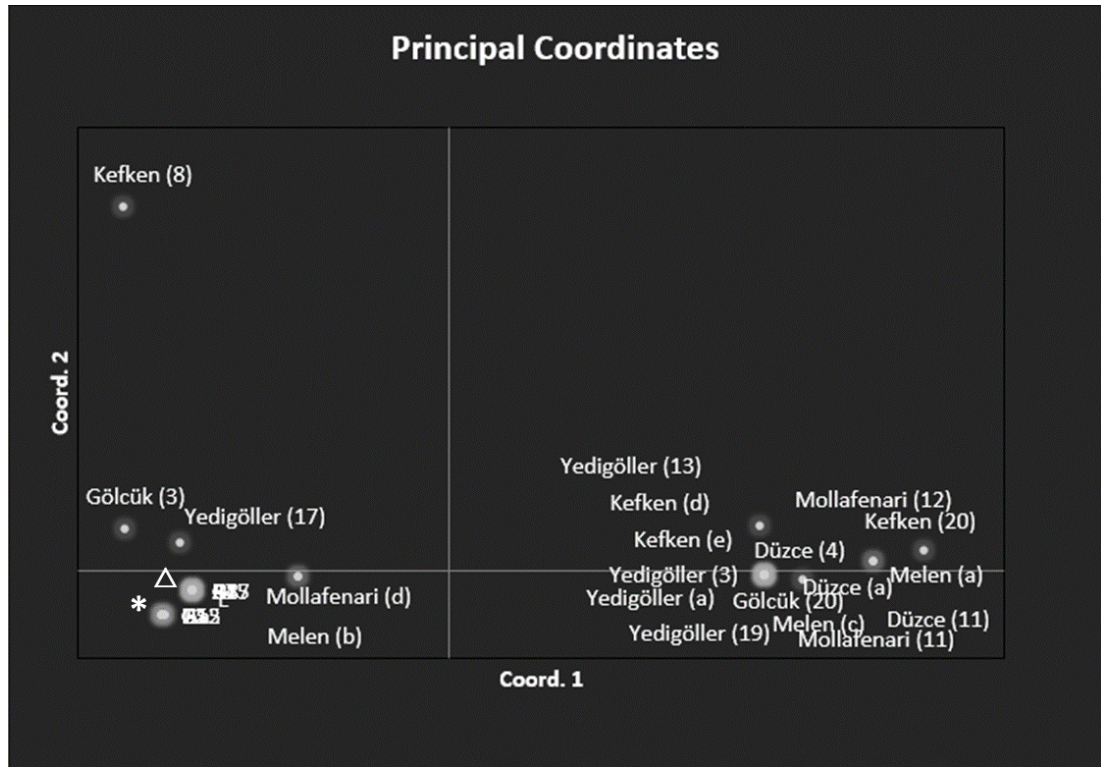


Figure 4.21: PCA analysis of *P. avium* L. haplotypes of the *ITS* gene obtained from "GenAlEx 6.4" program. The densities of the points representing the groups are directly proportional to the number of individuals. The symbol of Δ is represents the group of Melen, the group of Gölcük, the group of Kefken, individual of Yedigöller (4), the group of Düzce, the group of Mollafenari, individual of Düzce (8), individual of Mollafenari (15), individual of Mollafenari (17), the group of Abant. The symbol of * is represents: individual of Yedigöller (7), the group of Mollafenari (c), the group of Kefken (c), individual of Düzce (5), individual of Gölcük (15), individual of Melen (12), the group of Abant (a), individual of Yedigöller (2).

According to the results of PCA analysis for multilocus (*trnH-psbA+matK+ITS*) gene which is represented in Figure 4.22, the haplotypes separated as eleven groups. Fifth individual of Düzce population, third individual of Gölcük population, seventeenth individual of Yedigöller population and thirteenth individual of Yedigöller population constitute different particular groups by themselves. Third individual of Kefken population and second individual of Yedigöller population constitute a separate group together and located between seventeenth individual of Yedigöller population and fifth individual of Düzce population. Correlatively, twelfth individual of Melen population and x group of Abant population constitute a separate group together and located very closely to third individual of Gölcük population. While (x) group of Melen population, second and twelfth individuals of Kefken population, ninth individual and (x) group of Mollafenari constitute a particular group

and purvey a dense group for some haplotypes which is difficult to particularly identify individuals in the graph, similarly, z group of Melen population, y group of Mollafenari population and fourth and nineteenth individuals of Mollafenari population constitute a particular groups and purvey a dense group for some haplotypes which is difficult to particularly identify individuals in the graph.

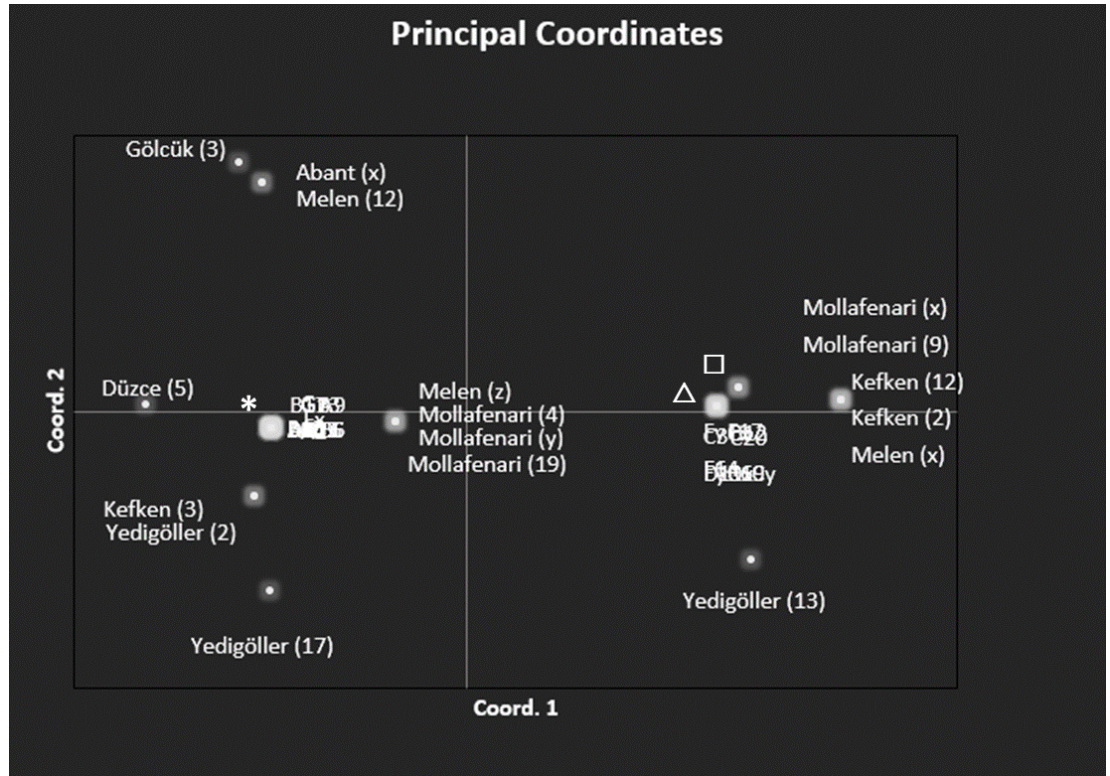


Figure 4.22: PCA analysis of *P. avium* L. haplotypes of multilocus (*trnH-psbA*+*matK*+*ITS*) obtained from "GenAlEx 6.4" program. The densities of the points representing the groups are directly proportional to the number of individuals. The symbol of * is represents: individual of Abant (10), the group of Yedigöller, individual of Mollafenari (15), the group of Mollafenari, individual of Gölcük (13), individual of Abant (13), the group of Kefken, the group of Gölcük, individual of Düzce (8), individual of Yedigöller (4), individual of Abant (2), the group of Düzce, the group of Kefken (x), individual of Mollafenari (17), individual of Mollafenari (6), the individual of Abant (16), the individual of Abant (12), the individual of Abant (9) and the group of Abant. The symbol of □ is represents: individual of Kefken (17), Kefken (y), individual of Yedigöller (3), individual of Yedigöller (20), Yedigöller (x) and Mollafenari (12). The symbol of Δ is represents: the group of Yedigöller (y), individual of Düzce (11), the group of Düzce (x), individual of Kefken (14), individual of Kefken (6), individual of Yedigöller (19), the group of Melen (y) and individual of Yedigöller (6).

4.5. Structure Analysis

Sequences of *trnH-psbA*, *matK*, and *ITS* DNA barcode genes from *P. avium* individuals for Structure analysis were aligned in the Mega 7 program (Figure 4.23) for identification of the *P. avium* haplotypes. 12, 24 and 37 haplotypes were identified, respectively.



Figure 4.23: SNP analysis of representative barcode gene (*ITS*) which used in the study by "Mega 7" program.

Using the SNPs obtained from sequence data of *P. avium* haplotypes, it was determined that the calculated number of *P. avium* haplotypes were divided into several population structures. As a result of 'Structure' analysis, it was determined that how seven *P. avium* populations are basically divided for the *trnH-psbA* gene (Figure 4.24), for the *matK* gene (Figure 4.25), for the *ITS* gene (Figure 4.26) and for the multilocus gene region (Figure 4.27). Genome of each accession is represented by a single vertical bar, which is partitioned into coloured segments in proportion to the estimated clusters according to K values.

The ΔK value which is the first step to calculate the value of K that separates the haplotypes, could not be calculated for *trnH-psbA* gene. The results of all K values from 1 to 10 were evaluated and the best result was obtained with K = 5 (Figure 4.24) which means that different haplotypes for *trnH-psbA* gene were figured out as five colored structure. However, it is not possible to distinguish the haplotypes from each other.

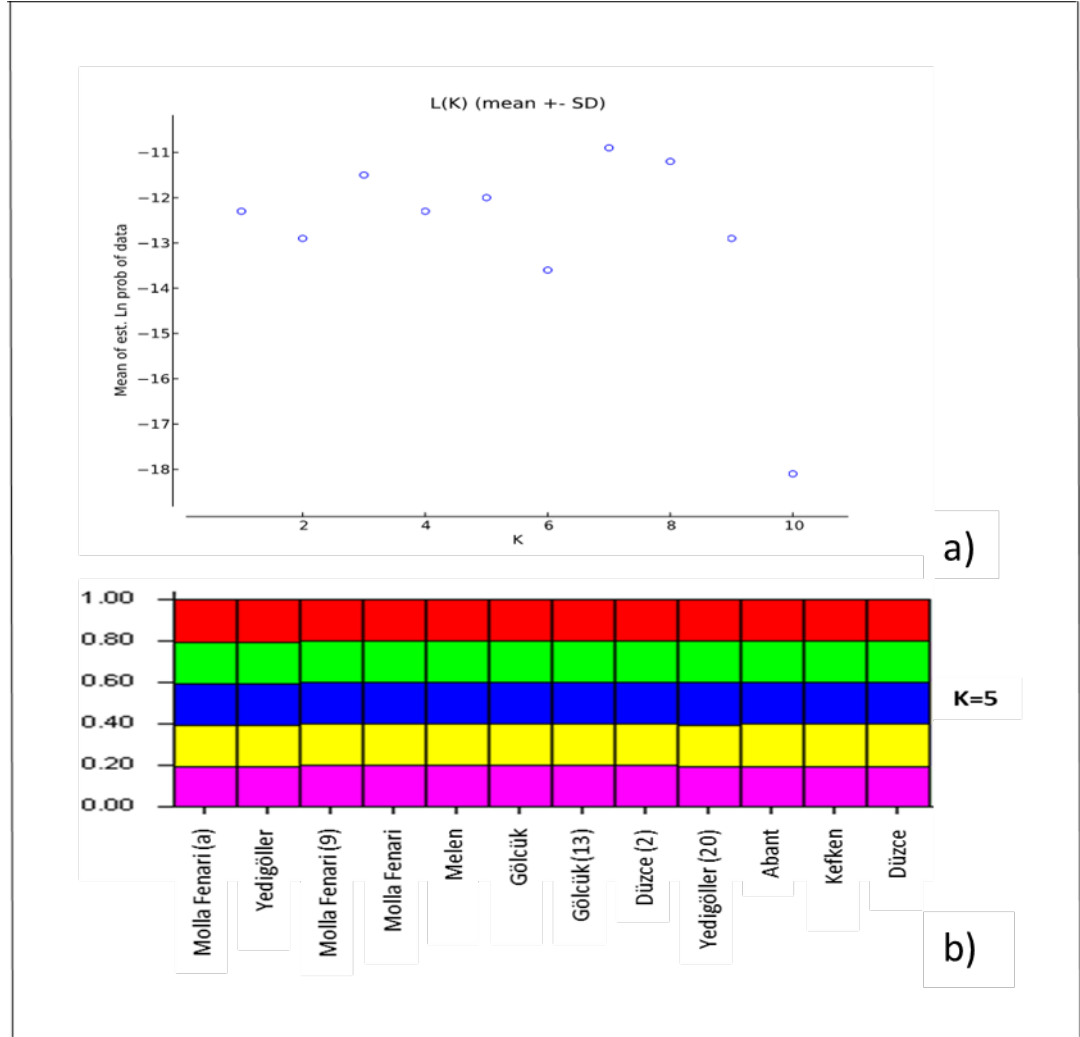


Figure 4.24: Estimation of L(K) for each K value and population structure for K=5 with samples of *P. avium* for *trnH-psbA* barcoding gene. a) Scatter diagram between mean of est. Ln prob. of data and hypothesized number of clusters, K. b) Estimated population structure with K = 5 ancestral gene pool ("cluster") using Structure program. The genome of each individual is represented by a vertical line that is divided into colored segments in proportion to the estimated membership of each of five clusters: Cluster I (red), Cluster II (green), Cluster III (blue), Cluster IV (yellow), Cluster V (purple).

With respect to ΔK graphics the value of 2 was chosen as K value for *matK* gene and haplotypes were figured out as two-colored structure (Figure 4.25). Likely with the structure analysis result of *trnH-psbA*, it is not possible to distinguish the haplotypes from each other as negligible differences were observed.

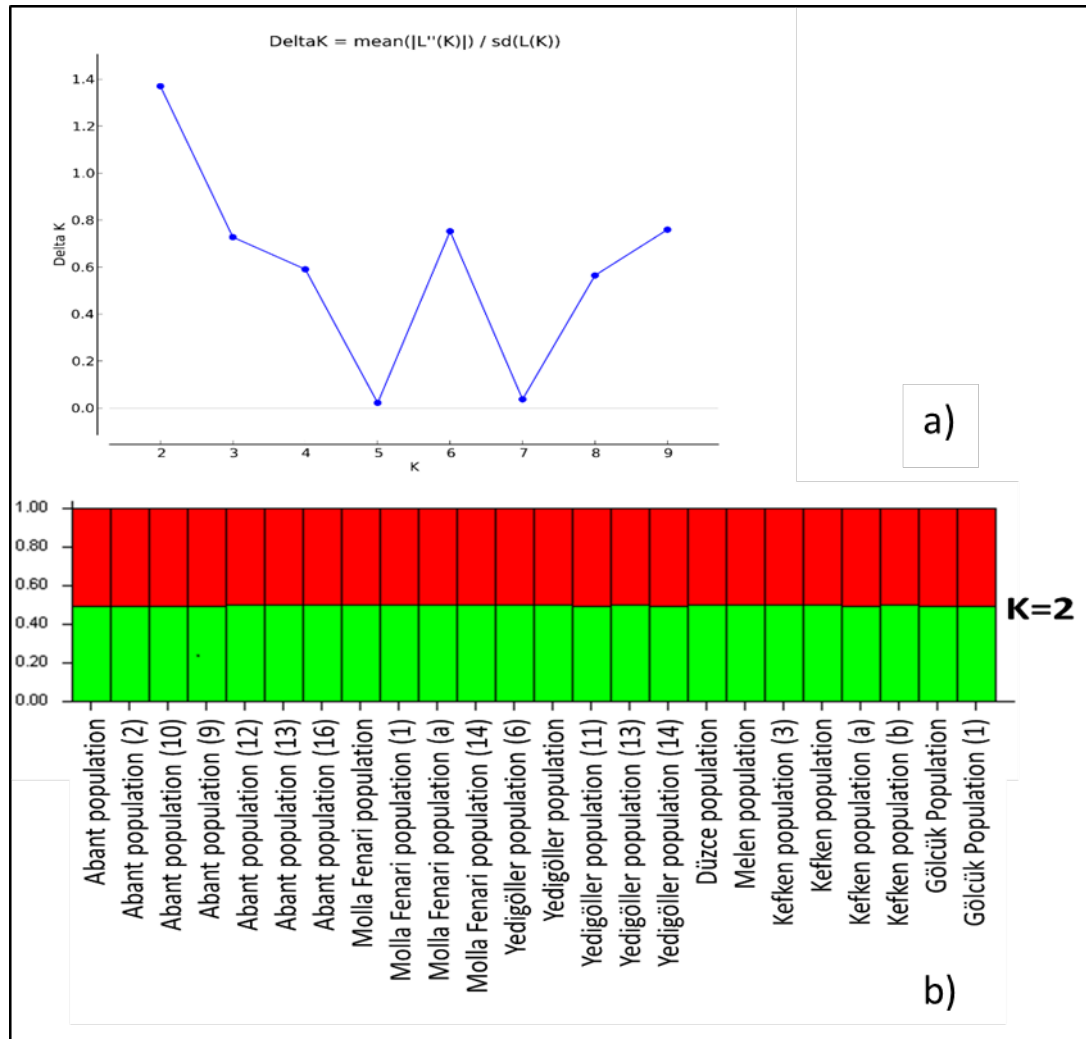


Figure 4.25: Estimation of L(K) for each K value and population structure for K=2 with samples of *P. avium* for *matK* barcoding gene. a) Evanno ΔK statistic as calculated for K-values of one to ten for 'Structure' analysis with samples of *P. avium* for *matK* barcoding gene. b) Estimated population structure with K = 2 ancestral gene pool ("cluster") using Structure program. The genome of each individual is represented by a vertical line that is divided into colored segments in proportion to the estimated membership of each of two clusters: Cluster I (red), Cluster II (green).

In the case of ITS region, with respect to ΔK graphics the value of 2 was chosen as K value as it can be seen in Figure 4.26 with red and green colors. Moreover, it was observed that haplotypes for *ITS* gene were separated as 3 different groups. The groups to which the ITS gene haplotypes belong are given below:

1st Group: Melen population (c), Kefken population (e), Kefken population (d), Mollafenari (11), Yedigöller population (19), Kefken population (c), Yedigöller population (3), Melen population (a), Düzce population (11), Düzce population (a), Yedigöller population (a), Düzce population (4).

2nd Group: Mollafenari population (d), Melen population (b), Yedigöller population (4), Abant population, Mollafenari population, Kefken population, Melen population, Mollafenari population (17), Mollafenari (15), Düzce population (8), Yedigöller population (7), Düzce population, Kefken population (20).

3rd Group: Gölcük population (3), Gölcük population (15), Yedigöller population (13), Kefken population (8), Gölcük population, Yedigöller population (17), Düzce population (5), Yedigöller population (2), Melen population (12), Abant population (a), Gölcük population (20), Mollafenari (12).

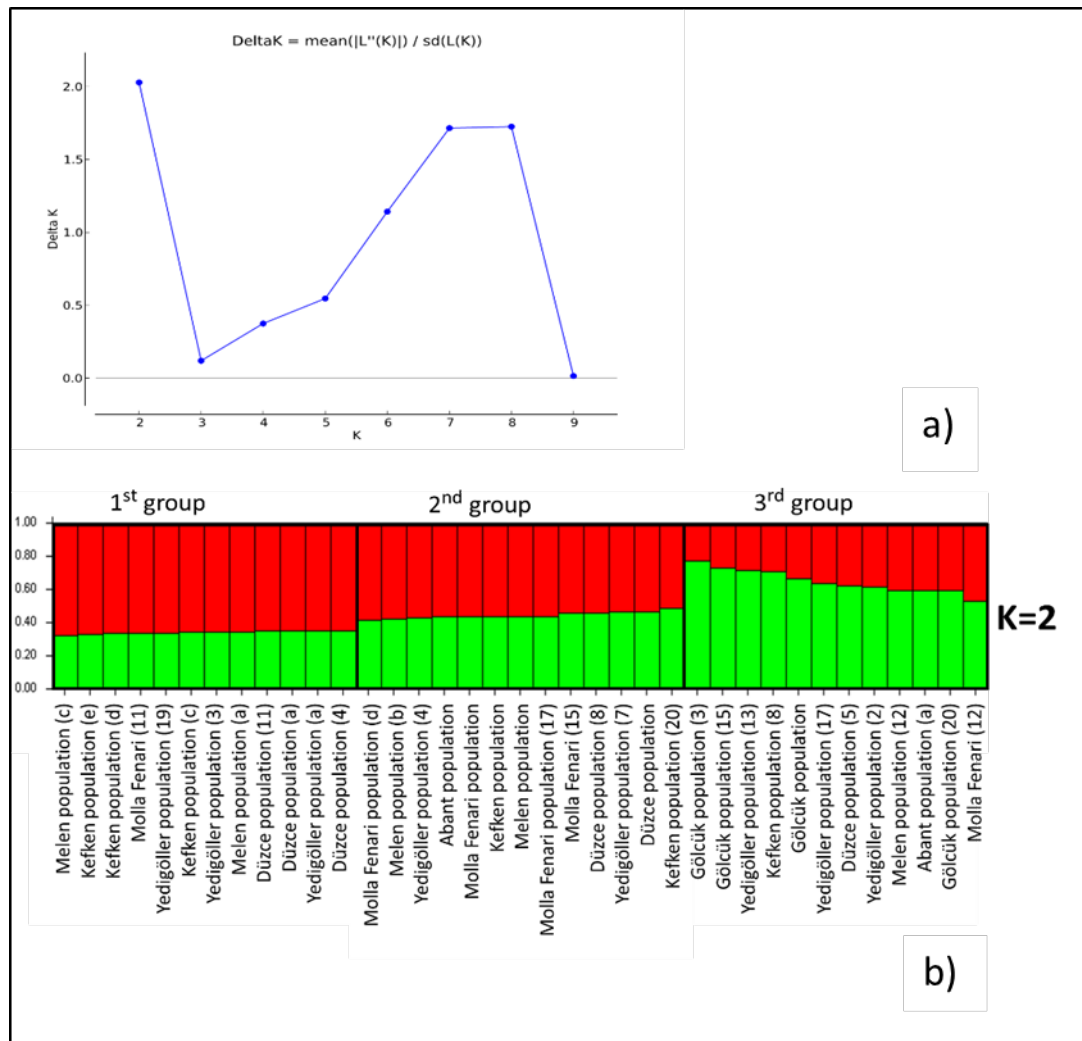


Figure 4.26: Estimation of $L(K)$ for each K value and population structure for $K=2$ with samples of *P. avium* for *ITS* barcoding gene. a) Evanno ΔK statistic as calculated for K -values of one to ten for 'Structure' analysis with samples of *P. avium* for *ITS* barcoding gene. b) Estimated population structure with $K = 2$ ancestral gene pool ("cluster") using Structure program. The genome of each individual is represented by a vertical line that is divided into colored segments in proportion to the estimated membership of each of two clusters: Cluster I (red), Cluster II (green).

For multilocus gene Estimated population structure with $K = 3$ which was the best result amongst the values from 1 to 10 ancestral gene-pool ("cluster") that was represented by red, green and blue colors (Figure 4.26). It was observed that haplotypes for multilocus (*trnH-psba+matK+ITS*) gene were separated as 4 different groups. The groups to which the multilocus gene haplotypes belong are given below.

Group 1: Abant population, Yedigöller population (y), Yedigöller population (4), Yedigöller population (19), Düzce population (x), Melen population, Kefken population, Abant population (9), Abant population (10), Yedigöller population (6),

Abant population (16), Yedigöller population (3), Düzce population (11), Melen population (y), Kefken population (y), Mollafenari population, Mollafenari population (15), Kefken population (6), Düzce population, Düzce population (8), Melen population (x), Kefken population (2), Mollafenari population (x), Mollafenari population (9).

Group 2: Kefken population (17), Kefken population (14), Kefken population (12), Melen population (12), Yedigöller population (x), Abant population (x), Yedigöller population (20), Melen population (z), Mollafenari population (y), Mollafenari population (17), Mollafenari population (6), Kefken population (x), Mollafenari population (4), Mollafenari population (19), Abant population (13), Yedigöller population (17), Yedigöller population (2), Abant population (12), Kefken population (3), Mollafenari population (12).

Group 3: Yedigöller population (7).

Group 4: Düzce population (5), Gölcük population (3), Yedigöller population (13), Gölcük population, Gölcük population (13), Abant population (2).

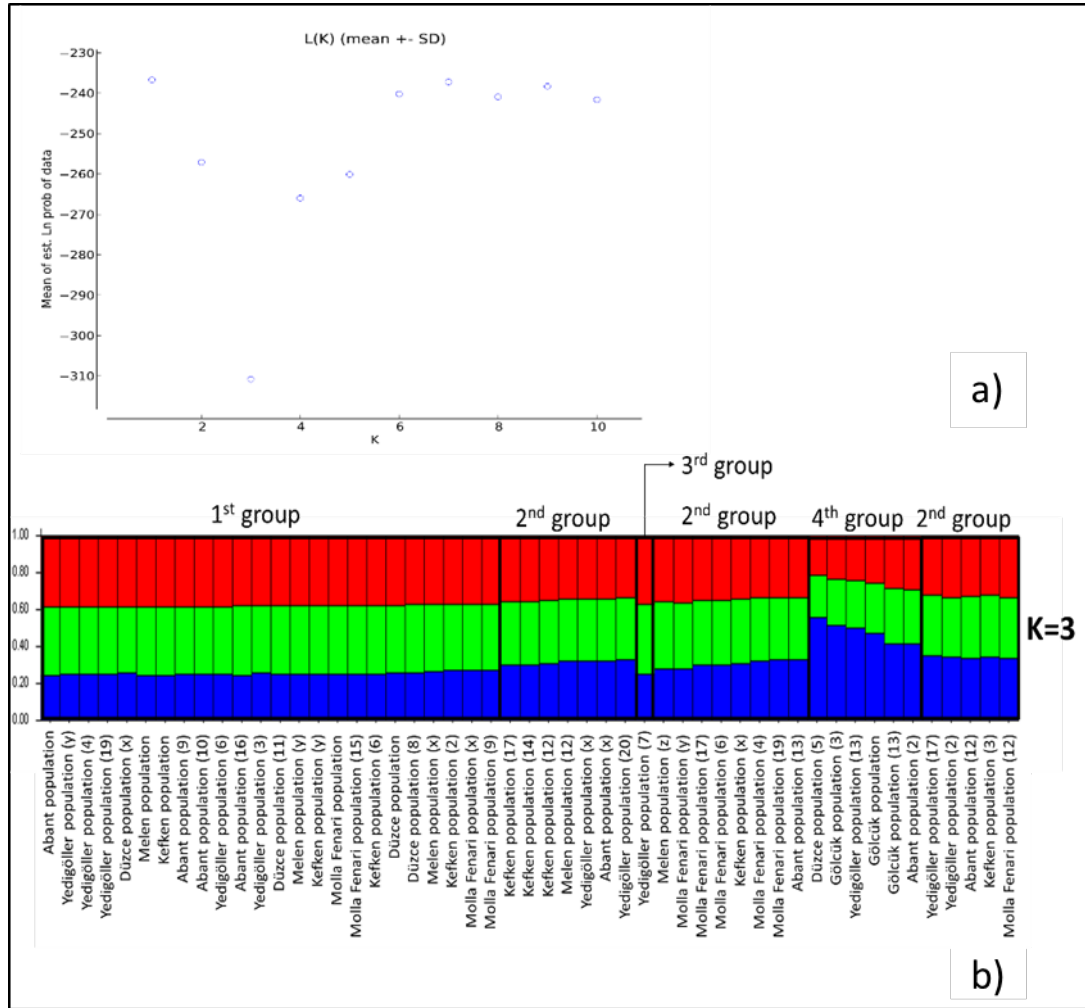


Figure 4.27: Estimation of $L(K)$ for each K value and population structure for $K=3$ with samples of *P. avium* for multilocus (*trnH-psbA+matK+ITS*) barcoding gene. a) Scatter diagram between mean of est. Ln prob of data and hypothesized number of clusters, K . b) Estimated population structure with $K = 5$ ancestral gene pool ("cluster") using Structure program. The genome of each individual is represented by a vertical line that is divided into colored segments in proportion to the estimated membership of each of three clusters: Cluster I (red), Cluster II (green), Cluster III (blue).

5. DISCUSSION

Plastidial (*rbcL*, *trnH-psbA*, *matK*) and nuclear (*ITS*) regions selected according to literature data and [Kress WJ et al., 2005], [Kress and Erickson, 2007]. These genes were compared to 139 *P. avium* individuals in 7 different subpopulations by using their discrimination power of single locus (*trnH-psbA*, *matK* and *ITS*) or multilocus region (*trnH-psbA* + *matK* + *ITS*). On the contrary to the literature [Prajapat and Chakraborty, 2011], [CBOL Plant Working group, 2009], [Hollingsworth, 2011] that considers *rbcL* as the best possible candidate for a plant barcode, no PCR amplification result was obtained with *rbcL* primers with all tested *P. avium* individuals (data not shown), although many different optimization (Tm degrees, DNA amount, BSA addition) were tested. Thus, it can be considered that current used *rbcL* primer has a limited utility as a plant barcode in wild *P. avium*. To this end, in future studies, it is suggested to use different primers that amplify different regions of *rbcL* in *P. avium* populations. It can also be evaluated that cpDNA can be isolated at the beginning of the study with the aim of to be able to obtain pure cpDNA for more amplification success with the current *rbcL* primer.

Overall PCR amplification results, *trnH-psbA* gene and *ITS* spacer showed 100% amplification success while *matK* was 89.93 % (Table 4.1). In accordance with the literature [Kress and Erickson, 2007], *matK* has a significant flaw like low PCR amplification success. *trnH-psbA* and *ITS* serve as good starting point for large scale testing of DNA barcoding in *P. avium*. Nevertheless, sequencing success of *ITS* were decreased by 2.87%.

Although only one species of *Prunus* was tested in this thesis, it was determined that the genetic diversity within the population is high according to the present number of haplotypes.

When the results of phylogenetic analysis of *P. avium* haplotypes belonging to 7 different subpopulations are examined by using two different parameters (ML and NJ), all phylogenetic trees had low bootstrap values which indicated that the haplotypes belonging to the species *P. avium* generally branch very close to each other. This result revealed that populations are genetically very close. Besides, it has been possible to observe that all the trees showed a correct distribution scheme according to the site of *P. armeniaca* as an outgroup with the high value of bootstrap (100%).

5.1. trnH-psbA

According to the phylogenetic construction results of *trnH-psbA* gene for both NJ and ML methods, individual of *P. avium* Düzce population (2) was separated from the rest in both methods. When the sequence of individual of *P. avium* Düzce population (2) was compared with all the other individuals, only one mutated base, which was a transversion was detected. It is obvious that this transversion and also its site on the sequence were very considerable and seemed to be relatively important to separate individual of *P. avium* Düzce population (2) from all the other individuals according to both applied methods. In addition to that, the pair of *P. avium* Mollafenari (9) and *P. avium* Mollafenari (a) were fragmented together. When the sequences of these individuals are compared with the rest of individuals of Mollafenari population, only one mutated base, which was a deletion for the individual of *P. avium* Mollafenari (9) and one transversion (A to T) for the group of *P. avium* Mollafenari (a) at the same site, was detected. It can be thought that the group of the rest individuals of Mollafenari population stand for the ancestral sequence as they located near to the other populations on the phylogenetic construction. To compare the differences between the individual of *P. avium* Mollafenari (9) and the group of *P. avium* Mollafenari (a) it is obvious that the amount of genetic change of the individual of *P. avium* Mollafenari (9) is more than the group of *P. avium* Mollafenari (a). Because the branches are related to the amount of genetic change in evolutionary lineages changing over time. It means that the individual of *P. avium* Mollafenari (9) has changed over the time. By NJ method, the type of the mutation, mutation sites and similarity to the other samples are discriminating features, but it is not possible to compare the differences among the other populations as there is not sharp differences observed on the phylogenetic trees constructed with both in two different methods. In the case of phylogenetic trees, the other mutations which are present in the other individuals are negligible. The mutations of them do not have enough power for branching on the phylogenetic tree in accordance with the low bootstrap values. It is possible to comment that the sequence regions used as barcodes are highly conserved among these populations.

According to the graph showing the results of PCA analysis of *P. avium* L. haplotypes of the *trnH-psbA* gene (Figure 4.19), three separate groups formed. Grouping of Düzce (2) haplotype is in agreement with the NJ and ML phylogenetic

contractions. Besides, PCA analysis and tree construction sufficiently corroborate each other with some slightly differences.

Estimated population structure was built for *trnH-psbA* gene by using Structure programme which is based on the differences of observed SNPs. The ΔK value which is the first step to calculate the value of K that separates the haplotypes, could not be calculated. According to results of the estimation of $L(K)$ for each K value, standard deviation of estimation $\ln \Pr(\text{Data})$ was less than 0.0000001 for each K value. The Evanno test requires division by the standard deviation of the estimated $\ln \Pr(\text{Data})$ values for all K between the first and last K value, and thus when the standard deviation is within Epsilon (0.000000) of zero “Structure Harvester” can not perform the test. As standard deviation $s[L(K)]$ was less than 0.0000001, ΔK was undefined for the sequence results of *trnH-psbA* gene. Another possibility in the case of concerning *trnH-psbA* gene, the inference of the number of gene pools (K) was not straightforward because log-likelihood values for the data conditional on K value, and $\ln P(x|K)$, increased progressively as K increased, as shown in Figure 4.24 (A). In such a case, it may not be possible to determine the true value of K . However, the nK values tried for all K classes indicated a strong signal at $K = 5$, as shown in Figure 4.24 (B). Estimated population structure with $K = 5$ which was the best result amongst the values from 1 to 10 ancestral gene-pool (“cluster”) was calculated by using Structure program for *trnH-psbA* gene. It was found that all the haplotypes were equally divided into five groups according to the Structure analysis. That’s why, no significant discrimination was observed among the haplotypes because negligible differences were observed. *trnH-psbA* barcode gene was not successful for discrimination of species-based haplotypes on the bases of structure analysis result.

5.2. matK

Even though the sequences were highly similar among members of each population, sufficient number of SNPs were observed among populations in the case of the *matK* gene. According to the phylogenetic construction results of *matK* gene for both NJ and ML methods, there are more considered haplotypes and mutations in comparison to *trnH-psbA* gene. It can be thought that usage of *matK* gene as a plant DNA barcode for *P. avium* is more efficient than usage of *trnH-psbA* gene.

Three separate groups formed according to the results of PCA analysis of *P. avium* haplotypes of *matK* gene (Figure 4.20). It has a tight bond with the phylogenetic tree of *matK* which constructed by ML. The phylogenetic tree of *matK* which constructed by NJ confirmed with the result of PCA in terms of the separation of the group of haplotypes of Yedigöller (11) and (14) together. PCA analysis and tree construction sufficiently corroborate each other with some differences that because of the usage of different statistical methods.

According to the structure analysis results of *matK* gene, because of the negligible differences among haplotypes as they equally divided into two groups, it can be thought that structure analysis is not convenient to discriminate the species-based haplotypes for *matK* gene.

matK proved to be the most informative barcode with *ITS* region as they had the highest numbers of variable sites.

5.3. ITS

As regards *ITS* gene for both NJ and ML methods, some individuals are divided into groups together. In ML method, individual of *P. avium* Yedigöller (2) showed a significant genetic alteration according to its length of horizontal branch. Similarly, individual of *P. avium* Kefken (8) showed a significant genetic change according to its length of horizontal branch in the phylogenetic tree of NJ. Among the two methods, the most significant similarity was detected for all subgroups of Gölcük population as it constructed a specific branch that separated them from all the other populations. All subgroups of Gölcük population isolated themselves but at the same time all of them are inserted inside the tree. In the light of these data, it can be stated that Gölcük population presents the highest genetic difference among all populations. All individuals of Kefken and Mollafenari populations are scattered inside the both phylogenetic trees. From this data, it can be assumed that Kefken and Mollafenari individuals were mostly pollinated with the other populations. As *ITS* spacer located in nuclear genome unlike to plastidial DNA barcode genes, it is influenced by gene transfusions sourced from both maternal and paternal.

The graph showing the results of PCA analysis of *P. avium* L. haplotypes of the *ITS* gene, and the phylogenetic constructions -especially ML- are in almost perfect

conformity as both of them considered the haplotype of Kefken (8) as an external group which is separated from all the other haplotypes. Moreover, haplotypes of Gölcük (3) and Yedigöller (17) closely located to each other on PCA that is very similar to the phylogenetic trees that belongs to *ITS* region. Similarly, haplotypes of Mollafenari (d) and Melen (b) were located in the same branch on the trees and they located on the same spot in the spatial plane of PCA analysis. The other specified haplotypes were also coherent with both of the phylogenetic trees. However, they were not as consistent as the groups given above. In addition to that, structure analysis results of *ITS* region are 100% compatible with PCA and phylogenetic trees with two exceptions which are Kefken population (20), Mollafenari (12). These differences were probably caused by different statistical methods which the analysis based on.

As it indicated before (p.81), CBOL has only regarded *ITS* as a supplementary locus [CBOL Plant Working Group, 2009] but, in this thesis study, *ITS* region can be recommended to be used for DNA barcode of wild *P. avium* although it currently presents some challenges for barcoding in plant species by the plant barcoding community.

5.4. Multilocus (*trnH-psbA*+*matK*+*ITS*)

According to the phylogenetic construction results of multilocus (*trnH-psbA* + *matK* + *ITS*) gene, the most remarkable result belongs to Gölcük population, similar to the phylogenetic tree construction of *ITS* spacer. Gölcük population is isolated from all the other populations. In addition to Gölcük population, a higher number of Mollafenari population was separated from the other populations. Phylogenetic trees, obtained from the analysis of the single *ITS* sequences and of all the other regions taken together (*trnH-psbA*+*matK*+*ITS*=1573 bp), initially separated different Kefken individuals from other populations which were subdivided by geographical origin. Kefken can be considered as one of an ancestor for wild cherry in Turkey. However, some individuals of them were influenced from the gene transfusions from the other populations.

According to the graph showing the results of PCA analysis of *P. avium* L. haplotypes of multilocus (*trnH-psbA*+*matK*+*ITS*) gene (Figure 4.22), eleven separate groups formed. As a result of the comparison between PCA and phylogenetic

constructions, there was not enough similarities in most groupings.

The ΔK value could not be defined for the sequence results of multilocus (*trnH-psbA+matK+ITS*) region. Because the standard deviation of estimation $\ln \Pr(\text{Data})$ was less than 0.0000001 for each K value. This low standard deviation could be due to the presence of *trnH-psbA* gene sequence in multilocus region in which ΔK value was not calculated properly. However, the nK values tried for all K classes indicated a strong signal at $K = 3$, as shown in Figure 4.27b. By virtue of the value of K, the population structure of the multilocus region haplotypes was divided into three different structures (Figure 4.27b). However, the result of structure is not compatible with PCA and phylogenetic tree results which possibly due to undefined ΔK . Thus, multilocus (*trnH-psbA+matK+ITS*) region is not convenient and suitable as a candidate plant DNA barcode in wild *P. avium*.

Overall, *P. avium* populations used in the study were evaluated in terms of regional distributions taking into account phylogenetic trees and geographical locations in Turkey. In phylogenetic analysis, Gölcük and Mollafenari populations are more isolated than others when the geographic location of Gölcük and Mollafenari populations was considered, it can be thought that they could be influenced by Europe by seaway. Moreover, it can also be assumed that, Kefken population can be carry ancestral tracks, due to its geographic location. As an interpretation, borderlines can give important data in ancestral sources studies of *P. avium* in Turkey as in the case of Kefken population. However, this assumption should be supported by more detailed molecular studies.

Abant, Yedigöller, Düzce and Melen populations are always mixed in every case of phylogenetic trees. A remarkable difference could not be obtained among them. This similarity could be due to pollination which occur by a range of insects, mainly wild bees, honeybees and bumblebees. Birds especially pigeons, starlings, thrushes and jays, and also larger vertebrates like foxes, badgers or wild boar play an important role in their dissemination. [Ducci, 2013], [Web 4].

Gismondi et al. (2013) also studied in *Crocus sativus* by using only one species. These two studies were important as both of them demonstrated the existence of barcode gene in intraspecific populations. This result can be due to the different habitat choices of the populations.

6. CONCLUSION

The primary objective of this study was to investigate the genetic variability of wild cherry populations collected from different geographical locations in Turkey. In the frame of this thesis, PCR and DNA sequencing were employed to test the discrimination power of each of four DNA barcode regions. All obtained results are listed below.

- PCR success rates of *trnH-psbA* (270bp) and *ITS* (660bp) were obtained as 100% while the rates of *rbcL* and *matK* (643bp) were obtained as 0% and 89,93%, respectively.
- Sequencing the success rates were 100%, 95,2% and 97,13% for of *trnH-psbA*, *ITS* and *matK*, respectively.
- The *ITS* sequences gave the highest variance among the populations while the lowest rates were observed in the *trnH-psbA* gene.
- The total number of haplotypes were observed in populations 6, 18 and 29 for *trnH-psbA* intergenic spacer, *matK* gene and *ITS* region, respectively.
- All individuals of Abant, Melen and Kefken populations are monomorphic and presented as H1 for to *trnH-psbA* gene. Abant population has the maximum number of haplotypes with the value of eight for *matK* gene while Yedigöller population has the maximum number of haplotypes with the value of nine in *ITS* gene.
- On the basis of aligned sequence results, the total percentage of observed variable sites was 3.030 for *ITS*, 2.022 for *matK* and 1.661 for *trnH-psbA*.
- According to the reported values, it was observed that, *trnH-psbA* gene had A as the most represented base (43.33 – 43.7%), while while C (36.76 – 37.07%) and T (28.60 – 29.14%) were found to be dominant basis for *matK* gene. In the case of *ITS* region, C was the most represented base.
- Overall results showed that *matK* was found to be one of the most variable noncoding cpDNA regions as plant DNA barcode in terms of percent variability while *ITS* was found to be one of the most variable noncoding nuclear DNA regions. In the light of these data, the usage of *ITS* region and *matK* gene as DNA barcode for *P. avium* species is recommended. However, studies have

indicated that the usage of multilocus combinations of candidate plant barcode genes can be a solution for plant barcoding studies. However, no clear consensus has yet emerged [Fazekas et al., 2008]. That's why, there is still a requirement for usage of various barcode gene regions in single or in combination for not only wild *P. avium* but also for other *Prunus* species in future studies.

REFERENCES

- Al-Qurainy F., Khan S., Tarroum M., Al-Hemaid F. M., Ali M. A., (2011), “Molecular authentication of the medicinal herb *Ruta graveolens* (Rutaceae) and an adulterant using nuclear and chloroplast DNA markers”, *Genetics and Molecular Research*, 10(4), 2806-2816.
- Álvarez I. and Wendel J.F., (2003), “Ribosomal ITS sequences and plant phylogenetic interference”, *Molecular Phylogenetics and Evolution*, 29, 417-434.
- Arnot D. E., Roper C., Bayoumi R. A., (1993), “Digital codes from hypervariable tandemly repeated DNA sequences in the *Plasmodium falciparum* circumsporozoite gene can genetically barcode isolates”, *Molecular and Biochemical Parasitology*, 61(1), 15-24.
- Asahina H., Shinozaki J., Masuda K., Morimitsu Y., Satake M., (2010), “Identification of medicinal *Dendrobium* species by phylogenetic analyses using matK and rbcL sequences”, *Journal of Natural Medicines*, 64(2), 133-138.
- Ausubel J. H., (2009), “A botanical macroscope”, *Proceedings of the National Academy of Science USA*, 106, 12569-12570.
- Avice J.C., Arnold J., Ball R.M., (1987), “Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics”, *Annual Review of Ecology, Evolution, and Systematics*, 18, 489-522.
- Barraclough T. G., Nee S., (2001), “Phylogenetics and speciation”, *Trends in Ecology and Evolution*, 16, 391–399.
- Bortiri E., Oh, S. H., Jiang J., Baggett S., Granger A., Weeks C., Parfitt D. E., (2001), “Phylogeny and systematics of *Prunus* (Rosaceae) as determined by sequence analysis of ITS and the chloroplast trnL-trnF spacer DNA”, *Systematic Botany*, 26(4), 797-807.
- CBOL Plant Working Group, (2009), “A DNA barcode for land plants”, *Proceedings of the National Academy of Sciences*, 106, 12794–12797.
- Chase M. W., Cowan R. S., Hollingsworth P. M., van den Berg C., Madriñán S., Petersen G., Pedersen N., (2007), “A proposal for a standardised protocol to barcode all land plants” *Taxon*, 56(2), 295-299.
- Chen S., Yao H., Han J., Liu C., Song J., Shi L., Luo K., (2010), “Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species”, *Plos One*, 5(1), 1-8.

Chiang C. H., Yu T. A., Lo S. F., Kuo C. L., Peng W. H., Tsay H. S., (2012), "Molecular authentication of *Dendrobium* species by multiplex polymerase chain reaction and amplification refractory mutation system analysis", *Journal of the American Society for Horticultural Science*, 137(6), 438-444.

Chiang Y. C., Chang W. T., Chen M. D., Lai G. H., Chen H. J., Chao J., Lee M. S., (2011), "Rapid identification of the medicinal plant *Taraxacum formosanum* and distinguishing of this plant from its adulterants by ribosomal DNA internal transcribed spacer (ITS) based DNA barcode", *African Journal of Biotechnology*, 10(24), 4838-4843.

Cohen I. M. (2011), "Testing the utility of the Consortium for the Barcoding of Life's two 'agreed upon' plant DNA barcodes, *matK* and *rbcL*", PhD Thesis, The University of Tennessee at Chattanooga.

Dahui L., Zaigui W., Xueshi L., Yi Y., (2012), "Identification of the medicinal plant *Dipsacus asperoides* from three other species in genus *Dipsacus* (Dipsaceae) by internal transcribed spacer of ribosomal deoxyribonucleic acid (rDNA ITS)", *Journal of Medicinal Plants Research*, 6(2), 289-295.

De Rogatis A., Ferrazzini D., Ducci F., Guerri S., Carnevale S., Belletti P., (2013), "Genetic variation in Italian wild cherry (*Prunus avium* L.) as characterized by nSSR markers", *Forestry*, 86(3), 391-400.

Dhar A., Minin V. N., (2016), "Maximum Likelihood Phylogenetic Inference", in *Encyclopedia of Evolutionary Biology*, Academic Press, 1st Edition, 499-506.

Dong W., Liu J., Yu J., Wang L., Zhou S., (2012), "Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding", *Plos One*, 7(4), 1-9.

Ducci F., De Cuyper B., de Rogatis A., Dufour J., Santi F., (2013), "Forest Tree Breeding in Europe" in *Managing Forest Ecosystems*, 463–511.

Dunning L. T., Savolainen V., (2010), "Broad-scale amplification of *matK* for DNA barcoding plants, a technical note", *Botanical Journal of the Linnean Society*, 164(1), 1-9.

Ercisli S., (2004), "A short review of the fruit germplasm resources of Turkey", *Genetic Resources and Crop Evolution*, 51(4), 419-435.

Fazekas A. J., Burgess K. S., Kesanakurti P. R., Graham S. W., Newmaster S. G., Husband B. C., Barrett S. C., (2008), "Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well", *Plos One*, 3(7), 1-12.

Felsenstein J., (1981), "Evolutionary trees from DNA sequences, a maximum likelihood approach", *Journal of Molecular Evolution*, 17, 368-376.

Felsenstein J., (1985), "Confidence limits on phylogenies: an approach using the bootstrap", *Evolution*, 39, 783–791.

Felsenstein J., (1989), "PHYLP-Phylogeny Inference Package Version 3.2", *Cladistics*, 5, 164–166.

Fernandez-Cruz J., Fernandez-Lopez J., Miranda-Fontaiña M. E., Diaz R., Toval G., (2014), "Molecular characterization of Spanish *Prunus avium* plus trees", *Forest Systems*, 23(1), 120-128.

Ganopoulos I., Aravanopoulos F. A., Argiriou A., Kalivas A., Tsaftaris A., (2011), "Is the genetic diversity of small scattered forest tree populations at the southern limits of their range more prone to stochastic events? A wild cherry case study by microsatellite-based markers", *Tree genetics and genomes*, 7(6), 1299-1313.

Gilani S. A., Qureshi R. A., Khan M., Ulla, F., Nawaz Z., Ahmad I., Potter D., (2011), "A molecular phylogeny of selected species of Genus *Prunus* L.(Rosaceae) from Pakistan using the TRN-L and TRN-F spacer DNA", *African Journal of Biotechnology*, 10(22), 4550-4554.

Gismondi A., Di Marco G., Delorenzo M., Canini A., (2015), "Upgrade of *Castanea sativa* (Mill.) genetic resources by sequencing of barcode markers", *Journal of genetics*, 94(3), 519-524.

Gismondi A., Fanali F., Labarga J. M. M., Caiola M. G., Canini A., (2013), "*Crocus sativus* L. genomics and different DNA barcode applications", *Plant Systematics and Evolution*, 299(10), 1859-1863.

Gismondi A., Leonardi D., Enei F., Canini A., (2013), "Identification of plant remains in underwater archaeological areas by morphological analysis and DNA barcoding", *Advances in Anthropology*, 3(04), 240-248.

Gismondi A., Rolfo M. F., Leonardi D., Rickards O., Canini A., (2012), "Identification of ancient *Olea europaea* L. and *Cornus mas* L. seeds by DNA barcoding", *Comptes rendus biologiques*, 335(7), 472-479.

Gonzalez M.A., Baraloto C., Engel J., Mori S.A., Pertronelli P., Riera B., Roger A., Thebaud C., Chave J., (2009), "Identification of Amazon trees with Barcodes", *Plos One*, 4(10), 1-7.

Goremykin V.V., Holland B., Hirsch-Ernst K.I., Hellwig F.H., (2005), "Analysis of *Acorus calamus* chloroplast genome and its phylogenetic implications", *Molecular Phylogenetics and Evolution*, 22, 1813-1822.

Gregory T.R., (2005), "DNA Barcoding does not compete with taxonomy", *Nature*, 434,1067.

Group C. P. W., Hollingsworth P. M., Forrest L. L., Spouge J. L., Hajibabaei M., Ratnasingham S., Fazekas A. J., (2009), "A DNA barcode for land plants", *Proceedings of the National Academy of Sciences*, 106(31), 12794-12797.

Guo H., Wang W., Yang N., Guo B., Zhang S., Yang R., Yu J., (2010), “DNA barcoding provides distinction between *Radix Astragali* and its adulterants”, *Science China Life Sciences*, 53(8), 992-999.

Guo X., Wang X., Su W., Zhang G., Zhou R., (2011), “DNA barcodes for discriminating the medicinal plant *Scutellaria baicalensis* (Lamiaceae) and its adulterants”, *Biological and Pharmaceutical Bulletin*, 34(8), 1198-1203.

Hajibabaei M., Singer G.A.C., Hebert P.D.N., Hickey D.A., (2007), “DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics”, *Trends in Genetics*, 23, 167-172.

Han J., Song J., Liu C., Chen J., Qian J., Zhu Y., Chen S., (2010), “Identification of *Cistanche* species (Orobanchaceae) based on sequences of the plastid *psbA-trnH* intergenic region”, *Yao xue xue bao= Acta pharmaceutica Sinica*, 45(1), 126-13.

He Y., Hou P., Fan G., Song Z., Arain S., Shu H., Zhang Y., (2012), “Authentication of *Angelica anomala* Avé-Lall cultivars through DNA barcodes”, *Mitochondrial DNA*, 23(2), 100-105.

Hebert P. D. N., Cywinska A., Ball S. L., (2003), “Biological identifications through DNA barcodes”, *Proceedings of the Royal Society of London B: Biological Sciences*, 270(1512), 313-321.

Hebert P. D. N., Stoeckle M.Y., Zemlak T.S., Francis C.M., (2004b), “Identification of birds through DNA barcodes”, *PLoS Biol.*, 2(10), 1657-1663.

Hedges, S. B., (2002), “The origin and evolution of model organisms”. *Nature Reviews Genetics*, 3(11), 838-849.

Hilu K. W., Black C., Diouf D., Burleigh J.G., (2008), “Phylogenetic signal in *matK* vs. *trnK*: A case study in early diverging eudicots [angiosperms]”, *Molecular Phylogenetics and Evolution*, 48, 1120–1130.

Hilu K.W., Liang H., (1997), “The *matK* gene: sequence variation and application in plant systematics”, *American Journal of Botany*, 84, 830-839.

Hilu K.W., Alice L.A., (1999), “Evolutionary implications of *matK* indels in Poaceae”, *American Journal of Botany*, 86, 1735-1741.

Hollingsworth P. M., Graham S. W., Little, D. P., (2011), “Choosing and using a plant DNA barcode” *Plos One*, 6(5), 1-13.

Hummer K. E., Janick J., (2009), “Rosaceae: taxonomy, economic importance, genomics”, in *Genetics and genomics of Rosaceae*, Springer New York, 1-17.

Hurst G.D.D., Jiggins F.M., (2005), “Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts”, *Proceedings of the Royal Society B: Biological Sciences*, 272, 1525.

Jarni K., De Cuyper B., Brus R., (2012), “Genetic variability of wild cherry (*Prunus avium* L.) seed stands in Slovenia as revealed by nuclear microsatellite loci”, *Plos One*, 7(7), 1-5.

Jinbo U., Kato T., Ito M., (2011), “Current progress in DNA barcoding and future implications for entomology”, *Entomological Science*, 14(2), 107-124.

Kazi T., Hussain N., Bremner P., Slater A., and Howard C., (2013), “The application of a DNA-based identification technique to over-the-counter herbal medicines”, *Fitoterapia*, 87, 27-30.

Khan S., Al-Qurainy F., Nadeem M., Tarroum M., (2012), “Development of genetic markers for *Ochradenus arabicus* (Resedaceae), an endemic medicinal plant of Saudi Arabia”, *Genetics and molecular research*, 11(2), 1300-1308.

Kim Y. S., Ryuk J. A., Ko B. S., (2012), “Discrimination of Korean *Rehmannia glutinosa* from Chinese *Rehmannia glutinosa* using sequence-characterized amplified region marker”, *Journal of the Korean Society for Applied Biological Chemistry*, 55(1), 1-6.

Kimura M., (1980), “A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences”, *Journal of Molecular Evolution*, 16(2), 111-120.

Kool A., de Boer H. J., Krüger Å., Rydberg A., Abbad A., Björk L., Martin G., (2012), “Molecular identification of commercialized medicinal plants in Southern Morocco”, *Plos One*, 7(6), 1-12.

Kress W. J., Erickson D.L., (2008), “DNA barcoding: genes, genomics, and bioinformatics”, *Proceedings of the National Academy of Sciences of the United States of America*, 105, 2761-2762.

Kress W. J., Wurdack K. J., Zimmer E. A., Weigt L. A., Janzen D. H., (2005), “Use of DNA barcodes to identify flowering plants”, *Proceedings of the National Academy of Sciences of the United States of America*, 102(23), 8369-8374.

Kritpetcharat O., Khemtonglang N., Kritpetcharat P., Daduang J., Daduang S., Suwanrungruang K., Chaveerach A., (2011), “Using DNA markers and barcoding to solve the common problem of identifying dried medicinal plants with the examples of *Smilax* and *Cissus* in Thailand”, *Journal of Medicinal Plants Research*, 5(15), 3480-3487.

Lahaye R., (2008), “DNA barcoding the floras of biodiversity hotspots”, *Proc Natl Acad Sci USA*, 105, 2923–2928.

Li J. H., (2008), “Phylogeny of *Catalpa* [Bignoniaceae] inferred from sequences of chloroplast *ndhF* and nuclear ribosomal DNA”, *Journal of Systematics and Evolution*, 46, 341–348.

Li X., Yang Y., Henry R.J., Rossetto M., Wang Y., Chen S., (2015), “Plant DNA barcoding: from gene to genome” *Biol. Rev. Camb. Philos. Soc.*, 90, 157–166.

Li M., Au K. Y., Lam H., Cheng L., Jiang R. W., But P. P. H., Shaw P. C., (2012), “Identification of Baiying (*Herba Solani Lyrati*) commodity and its toxic substitute Xungufeng (*Herba Aristolochiae Mollissimae*) using DNA barcoding and chemical profiling techniques”, *Food chemistry*, 135(3), 1653-1658.

Li R., Dao Z., (2011), “Identification of *Meconopsis* species by a DNA barcode sequence: The nuclear internal transcribed spacer (ITS) region of ribosomal deoxyribonucleic acid (DNA)”, *African Journal of Biotechnology*, 10(70), 15805-15807.

Liu T., Ji Y., (2012) “Molecular authentication of the medicinal plant *Paris polyphylla* Smith var. *yunnanensis* (Melanthiaceae) and its related species by polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP)”, *Journal of Medicinal Plants Research*, 6(7), 1181-1186.

Liu Y., Zhang L., Liu Z., Luo K., Chen S., Chen K., (2012), “Species identification of *Rhododendron* (Ericaceae) using the chloroplast deoxyribonucleic acid *psbA-trnH* genetic marker”, *Pharmacognosy magazine*, 8(29), 29.

Liu Z., Zeng X., Yang D., Ren G., Chu G., Yuan Z., Chen S., (2012), “Identification of medicinal vines by ITS2 using complementary discrimination methods”, *Journal of ethnopharmacology*, 141(1), 242-249.

Ma X. Y., Xie C. X., Liu C., Song J. Y., Yao H., Luo K., Chen S. L., (2010), “Species identification of medicinal pteridophytes by a DNA barcode marker, the chloroplast *psbA-trnH* intergenic region”, *Biological and Pharmaceutical Bulletin*, 33(11), 1919-1924.

Mathew B., (1982), “The Crocus: a revision of the genus *Crocus* (Iridaceae)”, London, BT Batsford Ltd, 127, 96.

McVean G., (2009), “A genealogical interpretation of principal components analysis”, *PLoS genetics*, 5(10).

Meizi L., Kun L., Hui Y., Ping L., (2011), “Authentication of *Sedum sarmentosum* and its adulterants by application of ITS2 sequences”, *Modern Chinese Medicine*, 13(12), 29-31.

Mendel J., Marešová E., Papoušek I., Halačka K., Vetešník L., Šanda R., Urbánková S., (2012), “Molecular Biodiversity Inventory of the Ichthyofauna of the Czech Republic”, In *Analysis of Genetic Variation in Animals*. InTech, 15, 287-314.

Mohanty A., Martin J. P., Aguinagalde I., (2001), “Chloroplast DNA study in wild populations and some cultivars of *Prunus avium* L.” *TAG Theoretical and Applied Genetics*, 103(1), 112-117.

Monkheang P., Sudmoon R., Tanee T., Noikotr K., Bletter N., Chaveerach A., (2011), "Species diversity, usages, molecular markers and barcode of medicinal Senna species (Fabaceae, Caesalpinioideae) in Thailand", *Journal of Medicinal Plants Research*, 5(26), 6173-6181.

Morgan G. J., (1998), "Emile Zuckerkandl, Linus Pauling, and the molecular evolutionary clock, 1959–1965", *Journal of the History of Biology*, 31(2), 155-178.

Mount D. W., (2001), "Phylogenetic Prediction", in *Bioinformatics: sequence and genome analysis*, Cold Spring Harbor, New York. 2nd edition, 237-280.

Nei M., Kumar S., (2000), "Molecular evolution and phylogenetics", Oxford university press, 333.

Newmaster S. G., Fazekas A. J., Ragupathy S., (2006), "DNA barcoding in land plants: evaluation of rbcL in a multigene tiered approach", *Canadian Journal of Botany/Revue Canadienne De Botanique*, 84, 335–341.

Patwardhan A., Ray S., Roy A., (2014), "Molecular markers in phylogenetic studies-a review", *Journal of Phylogenetics and Evolutionary Biology*, 2(2), 1-9.

Peng Z., Ludwig A., Wang, D., Diogo, R., Wei, Q., He, S., (2007), "Age and biogeography of major clades in sturgeons and paddlefishes (Pisces: Acipenseriformes)" *Molecular Phylogenetics and Evolution*, 42, 854–862.

Penny D., Watson E. E., Hickson R. E., Lockhart P. J., (1993), "Some recent progress with methods for evolutionary trees", *New Zealand Journal of Botany*, 31(3), 275-288.

Petersen G., Seberg O., Thorsøe S., Jørgensen T., Mathew B., (2008), "A phylogeny of the genus *Crocus* (Iridaceae) based on sequence data from five plastid regions", *Taxon*, 57(2), 487-499.

Peterson P. M., Romaschenko K., Johnson G., (2010), "A classification of the Chloridoideae (Poaceae) based on multi-gene phylogenetic trees", *Molecular Phylogenetics and Evolution*, 55, 580–598.

Phoolcharoen W., Sukrong S., (2013), "Molecular analysis of *Vitex* species using candidate DNA barcoding and PCR-RFLP of the matK gene for authentication of *Vitex glabrata*", *Natural product communications*, 8(1), 125-128.

Porras-Hurtado L., Ruiz Y., Santos C., Phillips C., Carracedo Á., Lareu M. V., (2013), "An overview of STRUCTURE: applications, parameter settings, and supporting software", *Frontiers in genetics*, 1-13.

Praciak A., Pasiecznic N., Sheil D., van Heist M., Sassen M., Correia C.S., Dixon C., Fyson G., Rashford K., Teeling C., (2013) "The CABI encyclopedia of forest trees", in CABI, Oxfordshire, 1st edition.

Prajapat A., Chakraborty D., (2012) "DNA Barcoding: A Rapid and Foolproof Method for Identification of Plant Material", in *Current Trends in Biotechnology*, 17, 351-380.

Primack R. B., R. Rozzi P. Feinsinger R. Dirzo F. Massardo (2001), “Fundamentos De Conservacion Biologica: Perspectivas Latinoamericanas.”, Fondo de Cultura Economica, 797.

Purty R., Chatterjee S., (2016), “DNA Barcoding: An effective technique in molecular taxonomy”, Austin J Biotechnol Bioeng, 3(1), 1059.

Rakonjac V., Mratinić E., Jovković R., Fotirić Akšić M., (2014), “Analysis of morphological variability in wild cherry (*Prunus avium* L.) genetic resources from Central Serbia”, Journal of Agricultural Science and Technology, 16(1), 151-162.

Russell K., (2003), “Technical Guidelines for genetic conservation and use for wild cherry (*Prunus avium*)”, EUFORGEN, International Plant Genetic Resources Institute, 1-6.

Sahare P., (2013), “EcoRV digestion patterns in ITS region of medicinal plants.”, International Journal of Scientific and Engineering Research, 4(6), 558.

Saitou N., Nei, M., (1987), “The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees”, Molecular Phylogenetics and Evolution. 4, 406-425.

Sass C., Little D. P., Stevenson D. W., Specht C. D., (2007), “DNA barcoding in the cycadales: testing the potential of proposed barcoding markers for species identification of cycads” Plos One, 2(11), 1-9.

Savolainen V., Cowan R. S., Vogler A. P., Roderick G. K., Lane R., (2005), “Towards writing the encyclopaedia of life: an introduction to DNA barcoding”, Philosophical Transactions of the Royal Society of London B: Biological Sciences, 360(1462), 1805-1811.

Savolainen V., Spichiger R., Manen J. F., (1997), “Polyphyletism of Celastrales deduced from a chloroplast noncoding DNA region”, Molecular Phylogenetics and Evolution, 7(2), 145-157.

Scaltsoyiannes A., Tsoulpha P., Iliev I., Theriou K., Tsaktsira M., Mitras D., Zaragotas D., (2009), “Vegetative propagation of ornamental genotypes of *Prunus avium* L.”, Propag Ornam Plants, 9, 198-206.

Scholz H., Scholz I., (1995), “Prunoideae”, in Gustav Hegi’s Illustrierte Flora von Mitteleuropa, Blackwell Wissenschafts-Verlag, Berlin, 2nd edition, 446–510.

Seberg O., and Petersen G., (2009), “How many loci does it take to DNA Barcode a Crocus?”, Plos One, 4, 1-6.

Selvaraj D., Shanmughanandhan D., Sarma R. K., Joseph J. C., Srinivasan R. V., Ramalingam S., (2012), “DNA barcode ITS effectively distinguishes the medicinal plant *Boerhavia diffusa* from its adulterants” Genomics, proteomics and bioinformatics, 10(6), 364-367.

Sharma A., Folch J. L., Cardoso-Taketa A., Lorence A., Villarreal M. L., (2012), "DNA barcoding of the Mexican sedative and anxiolytic plant *Galphimia glauca*", *Journal of ethnopharmacology*, 144(2), 371-378.

Shaw J., Lickey E.B., Beck J.T., Farmer S.B., Liu W., Miller J., Siripun K.C., Winder C.T., Schilling E.E., Small R.L., (2005), "The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis", *Amer. J. Bot.*, 92, 142–166.

Shneyer V. S., (2009), "DNA barcoding is a new approach in comparative genomics of plants" *Russian journal of genetics*, 45(11), 1267-1278.

Srirama R., Senthilkumar U., Sreejayan N., Ravikanth G., Gurumurthy B. R., Shivanna M. B., Shaanker R. U., (2010), "Assessing species admixtures in raw drug trade of *Phyllanthus*, a hepato-protective plant using molecular tools", *Journal of ethnopharmacology*, 130(2), 208-215.

Srirama R., Senthilkumar U., Sreejayan N., Ravikanth G., Gurumurthy B. R., Shivanna M. B., Shaanker R. U., (2010), "Assessing species admixtures in raw drug trade of *Phyllanthus*, a hepato-protective plant using molecular tools", *Journal of ethnopharmacology*, 130(2), 208-215.

Strimmer K. S. (1997), "Maximum likelihood methods in molecular phylogenetics.", PhD Thesis, University of Munich.

Sui X. Y., Huang Y., Tan Y., Guo Y., Long C. L., (2011), "Molecular authentication of the ethnomedicinal plant *Sabia parviflora* and its adulterants by DNA barcoding technique", *Planta medica*, 77(05), 492-496.

Sun Z., Gao T., Yao H., Shi L., Zhu Y., Chen S., (2011), "Identification of *Lonicera japonica* and its related species using the DNA barcoding method", *Planta medica*, 77(03), 301-306.

Sun Z., Song J., Yao H., Han J., (2012), "Molecular identification of *Cistanches Herba* and its adulterants based on nrITS2 sequence", *Journal of Medicinal Plants Research*, 6(6), 1041-1045.

Takamiya T., Wongsawad P., Tajima N., Shioda N., Lu J. F., Wen C. L., Yukawa T., (2011), "Identification of *Dendrobium* species used for herbal medicines based on ribosomal DNA internal transcribed spacer sequence", *Biological and Pharmaceutical Bulletin*, 34(5), 779-782.

Tanceva Crmaric O., Stambuk S., Satovic Z., Kajba D., (2011), "Genotypic Diversity of Wild Cherry [*Prunus avium* L.] in the Part of Its Natural Distribution in Croatia", *Sumarski List*, 135, 543-555.

Tavaud M., Zanetto A., Santi F., Dirlewanger E., (2001), "Structuration of genetic diversity in cultivated and wild cherry trees using AFLP markers", *Acta horticulturae*, 263-269.

Tezcan M., Vlachonasios K., Aki C., (2010), "DNA barcoding study on *Sideritis trojana* bornm: an endemic medicinal plant of IDA mountain, Turkey" *Fresenius Environ Bull*, 19(7), 1352-1355.

Theodoridis S., Stefanaki A., Tezcan M., Aki C., Kokkini S., Vlachonasios K. E., (2012), "DNA barcoding in native plants of the Labiatae (Lamiaceae) family from Chios Island (Greece) and the adjacent Çeşme- Karaburun Peninsula (Turkey)", *Molecular ecology resources*, 12(4), 620-633.

Tsai L. C., Wang J. C., Hsieh H. M., Liu K. L., Linacre A., Lee J. C. I., (2008), "Bidens identification using the noncoding regions of chloroplast genome and nuclear ribosomal DNA", *Forensic Science International: Genetics*, 2(1), 35-40.

Turet-Sayar M., Turkeç A., Demir T., (2012), "Identification of Sweet Cherry Cultivars (*Prunus avium* L.) and Analysis of Their Genetic Relationship Using Microsatellite DNA Fingerprinting", *Journal of Agricultural Science*, 4(8), 134-140.

Turkoglu Z., Bilgener S., Ercisli S., Yildirim N., (2013), "Simple sequence repeat (SSR) analysis for assessment of genetic variability in wild cherry germplasm", *Journal of Applied Botany and Food Quality*, 85(2), 229.

Ullah Z., Ahmad M., Milne R. I., (2013), "Phylogenetic relationship of Pakistani species of *Carex* L. based on matK gene sequence variation", *Pakistan Journal of Botany*, 45, 185-190.

Vaughan S. P., Russell K., (2004), "Characterization of novel microsatellites and development of multiplex PCR for large- scale population studies in wild cherry, *Prunus avium*", *Molecular Ecology Resources*, 4(3), 429-431.

Vaughan, S. P., Bošković R. I., Gisbert-Climent A., Russell K., Tobutt K. R., (2008), "Characterisation of novel S-alleles from cherry (*Prunus avium* L.)", *Tree genetics and genomes*, 4(3), 531-541.

Vaughan S. P., Cottrell J. E., Moodley D. J., Connolly T., Russell K., (2007), "Distribution and fine-scale spatial-genetic structure in British wild cherry (*Prunus avium* L.)", *Heredity*, 98(5), 274-283.

Wang H., Kim M. K., Kim Y. J., Lee H. N., Jin H., Chen J., Yang D. C., (2012), "Molecular authentication of the oriental medicines *Pericarpium Citri Reticulatae* and *Citri Unshius Pericarpium* using SNP markers" *Gene*, 494(1), 92-95.

Web 1, (2017), <http://www.pfaf.org/USER/Plant.aspx?LatinName=Prunus+avium>, (Accession date: 10/11/2017).

Web 2, (2017), <http://www.barcodeoflife.org/content/about/what-dna-barcoding>, (Accession date: 17/12/2017).

Web 3, (2017), http://epidemic.bio.ed.ac.uk/how_to_read_a_phylogeny, (Accession date: 19/12/2017).

Web 4, (2017), <https://www.woodlandtrust.org.uk/visiting-woods/trees-woods-and-wildlife/british-trees/native-trees/wild-cherry/>, (Accession date: 08/12/2017).

Welk E, de Rigo D, Caudullo G., (2016), “*Prunus avium* in Europe: distribution, habitat, usage and threats”, European Atlas of Foresty Species, 140-141.

Whitlock B.A., Hale A.M., Groff P.A., (2010), “Intraspecific inversions pose a challenge for the trnH-psbA plant DNA barcode”, Plos One, 5(11), 1-7.

Wilson C. A., (2009), “Phylogenetic relationships among the recognized series in Iris Section Limniris”, Systematic Botany, 34, 277–284.

Wilson A. C., Carlson S. S., White T. J., (1977), “Biochemical evolution”, Annual review of biochemistry, 46(1), 573-639.

Wiriyakarun S., Yodpetch W., Komatsu K., Zhu S., Ruangrunsi N., Sukrong S., (2013), “Discrimination of the Thai rejuvenating herbs *Pueraria candollei* (White Kwao Khrua), *Butea superba* (Red Kwao Khrua), and *Mucuna collettii* (Black Kwao Khrua) using PCR-RFLP”, Journal of natural medicines, 67(3), 562-570.

Xiao W. L., Motley T. J., Unachukwu U. J., Lau C. B. S., Jiang B., Hong F., Kennelly E. J., (2011), Chemical and genetic assessment of variability in commercial *Radix Astragali* (*Astragalus* spp.) by ion trap LC–MS and nuclear ribosomal DNA barcoding sequence analyses. Journal of agricultural and food chemistry, 59(5), 1548-1556.

Xu G., Wang X., Liu C., Li W., Wei S., Liu Y., Liu J., (2013), “Authentication of official Da-huang by sequencing and multiplex allele-specific PCR of a short maturase K gene”, Genome, 56(2), 109-113.

Yang J. Y., Jang S. Y., Kim H. K., Park S. J., (2012), “Development of a molecular marker to discriminate Korean *Rubus* species medicinal plants based on the nuclear ribosomal DNA internal transcribed spacer and chloroplast trnL-F intergenic region sequences”, Journal of the Korean Society for Applied Biological Chemistry, 55(2), 281-289.

Yang Y., Zhai Y., Liu T., Zhang F., Ji Y., (2011), “Detection of *Valeriana jatamansi* as an adulterant of medicinal Paris by length variation of chloroplast psbA-trnH region”, Planta medica, 77(01), 87-91.

Yang Z., Rannala B., (2012), “Molecular phylogenetics: principles and practice”, Nature Reviews Genetics, 13(5), 303-314.

Yu J., Xue J. H., Zhou S. L., (2011), “New universal matK primers for DNA barcoding angiosperms”, Journal of Systematics and Evolution, 49(3), 176-181.

Zuo Y., Chen Z., Kondo K., Funamoto T., Wen J., Zhou S., (2011), “DNA barcoding of *Panax* species”, Planta medica, 77(02), 182-187.

Zurawski G., Bottomley W., Whitfeld P. R., (1982), "Structure of the genes for the beta and epsilon subunits of the spinach chloroplast ATPase indicates a dicistronic mRNA and an overlapping translation stop/start signal", *Proceedings of the National Academy of Sciences, USA*, 79, 6260-6264.

BIOGRAPHY

Selin Gül ÜNSAL was born in 1991 in Istanbul. She earned her B.Sc degree from Istanbul Kültür University on Molecular Biology and Genetics and earned her M.Sc degree from Acıbadem University on Management of Drug Industry, Institute of Health Science. She worked in Sativa Biotechnology Company, Düzen Laboratory Groups and Farmasina Medical and Chemical Products Company, respectively. She earned her Master of Science degree at Gebze Technical University Graduate School of Natural and Applied Sciences, Department of Molecular Biology and Genetics in 2018.