GENETIC CHARACTERIZATION OF PEAR CULTIVARS (*PYRUS COMMUNIS*) IN ÇORUH RIVER BASIN

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

BY

ABDULBAKİ ÇOBAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
MOLECULAR BIOLOGY AND GENETICS

APRIL 2019

Approval of the Thesis:

GENETIC CHARACTERIZATION OF PEAR CULTIVARS (*PYRUS COMMUNIS*) IN ÇORUH RIVER BASIN

Submitted by **ABDULBAKİ ÇOBAN** in partial fulfillment of the requirements for the degree of **Master of Science in Molecular Biology and Genetics**, **Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar _____

Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Orhan Adalı _____

Head of the Department, **Biological Sciences**

Prof. Dr. Zeki Kaya _____

Supervisor, **Biological Sci., METU**

Examining Committee Members:

Prof. Dr. İrfan Kandemir _____
Biology, Ankara Universit

Prof. Dr. Zeki Kaya _____
Biological Sci., METU

Prof. Dr. Musa Doğan _____
Biological Sci., METU

Prof. Dr. Sertaç Önde _____
Biological Sci., METU

Prof. Dr. Özgür Eminağaoğlu _____
Forest Eng. Artvin Çoruh University

Date: 12.04.2019

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

Name, Last name: Abdulbaki Çoban

Signature:

ABSTRACT

GENETIC CHARACTERIZATION OF PEAR CULTIVARS (*PYRUS COMMUNIS*) IN ÇORUH RIVER BASIN

Çoban, Abdulbaki

MSc., Department of Molecular Biology and Genetics

Supervisor: Prof. Dr. Zeki Kaya

April 2019, 93 Pages

The genus *Pyrus* contains more than 20 different pear species which are used as food source, horticultural, and ornamental purposes. Overall, pear is the second most consumed pome fruit after apple. *P. communis* is the most cultivated pear species in Europe and Asia Minor. However, more than 10 species of genus *Pyrus* can be found naturally in Europe and Asia Minor. In this study, 8 *P. communis* populations (one wild and 7 cultivated) and 1 *P. eleagnifolia* population was used to reveal genetic structure of those populations by using 11 SSR markers previously used in genus *Pyrus* and *Malus*. 84 genotypes of *P. communis* were sampled from Artvin Province, Turkey and 20 genotypes of *P. eleagnifolia* were sampled from METU campus in Ankara, Turkey for the current study.

Overall, there were no null alleles in the studied populations. There is no duplicated genotypes and no linkage between loci were found. Expected heterozygosity, allelic richness, and polymorphic information content were calculated to check the usefulness of the studied loci. All loci were found to be highly polymorphic for the further studies.

Structure analysis of the studied populations of *P. communis* revealed that there are significant gene flow between populations. Therefore, no clear population

differentiation was found. AMOVA results supported this finding as most of the differentiation was among genotypes within population. However, the wild population of *P. communis* were found to be distinct from other populations. Thus, it is summed as during domestication of *P. communis,* there were hybridization between genotypes within and between populations. As for the genetic structure analysis of *P. communis* and *P. eleagnifolia* populations, it was revealed that *P. communis* and *P. eleagnifolia* populations were significantly different from each other as expected. Since those are totally different species and due to geographical isolation, there were no gene flow between those populations.

The study was one of the first studies conducted on *P. communis* populations in Turkey. Thus, outcomes of the study are important for possible further studies which will be conducted on genus *Pyrus.* Besides, findings are important for further conservation studies of *P. communis* genotypes and breeding studies.

Keywords: *Pyrus communis, Pyrus eleagnifolia,* SSR, population structure, Genetic diversity

ÖZ

ÇORUH HAVZASINDA BULUNAN ARMUT KÜLTÜR ÇEŞİTLERİNİN (*PYRUS COMMUNIS*) GENETİK KARAKTERİZASYONU

Çoban, Abdulbaki

Yüksek lisans, Moleküler Biyoloji ve Genetik Bölümü

Tez Danışmanı: Prof. Dr. Zeki Kaya

Nisan 2019, 93 Sayfa

Pyrus cinsi süs bitkisi ve kişisel tüketim amaçlı kullanılan 20'den fazla armut türünü içeren bir cinstir. Tüm bu türlerle beraber, armut elmadan sonra en çok üretimi yapılan ikinci yumuşak çekirdekli meyve türünün genel adıdır. *P. communis* Avrupa ve Anadolu'da en çok üretimi yapılan armut türüdür. Bununla beraber, 10'dan fazla armut türünün doğal popülasyonuna bu coğrafyada rastlamak mümkündür. Bu çalışmada, genetik yapı ve çeşitliliği test etmek adına 8 *P. communis* popülasyonu (1 doğal ve 7 bahçe popülasyonu olmak üzere) ve 1 *P. eleagnifolia* popülasyonu kullanıldı. Daha önce *Pyrus* cinsi ve *Malus* cinsi için kullanılmış olan 11 tane SSR belirteci kullanıldı. *P. communis* örnekleri Artvin iline bağlı köylerden ve *P. eleagnifolia* örnekleri ODTÜ Ankara yerleşkesinden toplandı. Bu çalışmada toplam 104 olmak üzere, 84 *P. communis* ve 20 *P. eleagnifolia* örneği kullanıldı.

Genel olarak, popülasyonlarda null alellere rastlanmadı fakat yalnızca bazı popülasyonlardaki bazı lokuslarda null alel gözlendi. Bununla beraber, klonal çoğalma ve lokuslarda bağlantı dengesizliği gözlenmedi. Lokusların kullanılabilirliğini test etmek adına beklenen heterezigotluk, alel zenginliği ve polimorfizm bilgi içeriği hesaplandı. Bütün lokuslar devam edecek olan testler için yüksek ölçüde polimorfik bulundu.

Popülasyon yapı analiziyle beraber, popülasyonlar arasında yüksek oranda gen akışı olduğu bulundu. Bu yüzden popülasyonlar arasındaki çeşitlilik net bir şekilde bulunamadı. AMOVA sonuçlarıyla beraber çeşitliliğin popülasyonlar arasında olmasından ziyade bireyler arasında olduğu doğrulandı. Bununla beraber, *P. communis*'in doğal popülasyonunun diğer popülasyonlardan oldukça farklı olduğu bulundu. Bu sonuç, *P. communis'*in ehlileştirilmesi sürecinde bireyler ve popülasyonlar arasında hibridizasyon olduğunu özetledi. *P. communis* ve *P. eleagnifolia* popülasyonları arasındaki genetik yapı analizinde bu popülasyonların birbirinden tamamen farklı oldukları görüldü. Bu farklılık coğrafi izolasyon sebebiyle popülasyonlar arasında gen akışının olmamasıyla açıklandı.

Bu çalışma Türkiye'de *P. communis* üzerine yapılmış ilk çalışmalardan biri olması sebebiyle, gelecek çalışmalar için önemli sonuçlar içermektedir. Çalışma sürecinde *Malus* ve *Pyrus* cinslerinde kullanılmış SSR belirteçleri kullanıldığı için, gelecekte bu cinsler üzerine yapılma ihtimali olan çalışmalar için literatür bilgisi taşımaktadır. Bununla beraber elde edilen veriler, *P. communis* türünü genetik kaynaklarının korunmasında ve yabani türlerin ehlileştirme sürecinde faydalı olacaktır.

Anahtar Kelimeler: *Pyrus communis, Pyrus eleagnifolia,* Popülasyon yapısı, Genetik çeşitlilik

To My Mother

ACKNOWLEDGEMENTS

I would like to greatly thank to my supervisor Prof. Dr. Zeki Kaya for his guidance, supervision and endless patience throughout the study. I would like to also thank to Prof. Dr. Sertaç ÖNDE for his concern and kindness.

I would like to express my thanks to all jury members for their helpful comments and criticisms on the manuscript.

I specifically thank to TÜBİTAK (The Scientific and Technological Research Council of Turkey) because of the MS scholarship that I was honored.

I am also deeply thankful to my labmates Funda Özdemir Değirmenci, Asiye Çiftçi, Çiğdem Kansu, Bircan Taşkıran, Mevlüde Alev Ateş, Pelin Acar from Plant Genetics and Tissue Culture Laboratory for their endless support.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVATIONS

%P	Percentage of Polymorphic Loci
AMOVA	Analysis of Molecular Variance
Ar	Allelic Richness
CLUMPP	CLUster Matching and Permutation Program
CTAB	Cetyl Trimethyl Ammonium Bromide
DAPC	Discriminant Analysis of Principle Components
DNA	Deoxyribonucleic Acid
dNTP	Deoxy ribonucleotide triphosphate
EDTA	Ethylenediaminetetraaceticacid disodium salt
FAO	Food and Agriculture Organization of the United Nations
F	Inbreeding Coefficients
Fct	Difference Among Groups For The Total Population
Fis	Inbreeding Coefficient Within Individuals
Fit	Inbreeding Coefficient Within Total Population
Fsc	Differences Among Population Within Groups
Fst	Differences Among Subpopulation (AMOVA)
Fst	Inbreeding Coefficient Within Subpopulations
GDA	Genetic Data Analysis
GPS	Global Positioning System
G-W Index	Garza-Williamson Index
He	Expected Heterozygosity
Ho	Observed Heterozygosity
HWE	Hardy–Weinberg Equilibrium
I	Shannon’s Information Index
IAA	Isoamyl Alcohol
LD	Linkage Disequilibrium
MCMC	Markov Chain Monte Carlo
MgCl2	Magnesium Chloride

MLGs	Multilocus Genotypes
MSN	Minimum Spanning Network
Na	Number of Alleles
Ne	Number of Effective Alleles
Nm	Number of Migrants
OD	Optical Density
PCoA	Principal Coordinate Analysis
PCR	Polymerase Chain Reaction
PI	Probability of Identity
PIC	Polymorphic Information Content
rd	Standardized Index of Association
SSR	Simple Sequence Repeats
TBE	Tris-Borate-EDTA
TE	Tris EDTA
TÜBİTAK	The Scientific and Technological Research Council of Turkey
UPGMA	Unweighted Pair-Group Method with Arithmetic Mean

CHAPTER 1

INTRODUCTION

Pear is one of the most produced and consumed tree fruits around the world. It has been cultivated for more than two thousand years by human kind (Bell, 1990) and it is used as a general name for more than 20 species (Bell *et al.,* 1996). According to the Food and Agriculture Organization of the United Nations (FAO), it is the second most produced pome fruit after apple with around 27 million metric tons of production per year worldwide occupying an area of 1.6 million hectares (FAOSTAT, 2010). As for the production in Turkey, almost half a million metric tons of pear was produced in Turkey in 2016 occupying an area of 25.4 thousand hectares (FAOSTAT, 2010; TUIK, 2017).

1.1. The Genus Pyrus

The family Rosaceae is an angiosperm plant family containing approximately 3000 species and 90 genera including fruits like pears, apples, cherries and horticultural / ornamental trees and shrubs like roses (Potter *et. al.,* 2007, Christenhusz and Byng, 2016). It is a well-distributed family around the world, mainly in non-desert and non-tropical forest areas of the Northern hemisphere. In Rosaceae family, the genus *Malus* (apples) and the genus *Pyrus* (pears) are composed the most economically valued fruits (FAOSTAT, 2010). Beside the mostly consumed species of *Pyrus* which are consumed as food sources (*Pyrus communis, Pyrus pyrifolia, Pyrus ussuriensis*), some members of the genus *Pyrus* are used as ornamental purposes around the world such as; *Pyrus calleryana, Pyrus koehnii. Pyrus nivalis* cultivars are used to produce pear cider named perry (Hummer and Janick, 2009). In the Table 1.1, taxonomy of the *Pyrus communis* and *Pyrus eleagnifolia* were given.

The genus *Pyrus* is categorized under the subfamily Pomoideae in the Rosaceae family. It is broadly accepted that genus *Pyrus* has 22 species (Bell *et al.*, 1996). All of its species can be naturally found in the temperate regions of the old world. However, it is hard to define precise number of species of genus *Pyrus.* Terpo (1985) stated that there are 52 different species, Browicz (1993) defined 38 species and Kutzelnigg and Silbereisen (1995) suggested a list of 20 to 74 different species. Out of those species, *Pyrus communis* is mainly grown in Europe and Minor Asia while *Pyrus pyrifolia*, *Pyrus bretschneideri,* and *Pyrus ussuriensis* are the most cultivated ones in China and Japan.

Table 1.1: Taxonomy of *P. communis* and *P. eleagnifolia*

Kingdom	Plantae	
Phylum	Tracheophyta	
Class	Magnoliopsida	
Order	Rosales	
Family	Rosaceae	
Subfamily	Maloideae	
Genus	*Pyrus*	
Scientific Name	*Pyrus communis*	*Pyrus eleagnifolia*

1.2. The Origin and Distribution of *Pyrus*

The Maloideae subfamily has a basic chromosome number x = 17. The most accepted theories regarding the emergence of genus *Pyrus* are based on allopolyploid cross between Spiraeoideae with x = 9 and Prunoideae with x = 8 which are two primitive forms of Rosaceae family. Isozyme studies and univalent chromosomes during meiosis support those theories (Sax, 1931; Weeden and Lamb, 1987). Mainly, pear species are grouped into three: small fruits with three carpels (Asian pears), bigger

fruits with five carpels (European pears), and the hybrids of those with three-four carpels (Silva *et al., 2014*).

The genus *Pyrus* is thought to be originated from the highlands of southwestern and western China during the Tertiary period (around 60 million of years ago). Since those mountainous regions of China hosted large number of other species of the family Rosaceae, importantly species from subfamily of *Prunoideae* and *Pomoideae*, this thought is supported with the previous theories regarding the origin of the genus *Pyrus*. However, according to the fossil data, there have been some members of genus *Pyrus* in Caucasus and Western Europe since the Tertiary period (Rubstov, 1944). Besides, Vavilov (Vavilov, 1951) argued that it has three main diversity centers around the world: Asia Minor, Central Asia, and China. Today, it is easy to distinguish Eastern and Western pear morphologically. While Western Pear has an elongated body and full-bodied texture, Eastern Pear has a globular body with a sandy texture. In Figure 1.1, morphological differences between those two types of pear can be seen.

Figure 1.1: Western pear, namely *P. communis*, (on the left) and Eastern Pear, namely *Pyrus pyrifolia* (on the right).

Due to its importance as a fruit today, it is highly possible to see members of genus *Pyrus* all around the globe. However, wild populations of *P. communis* is mainly distributed in Europe and Caucasus. (Figure 1.2.) Besides, in overall pear production, China alone, holds the 70% of annual pear production or the world by almost 20 million metric tons (FAOSTAT 2010). In Figure 1.3, world top producers can be seen.

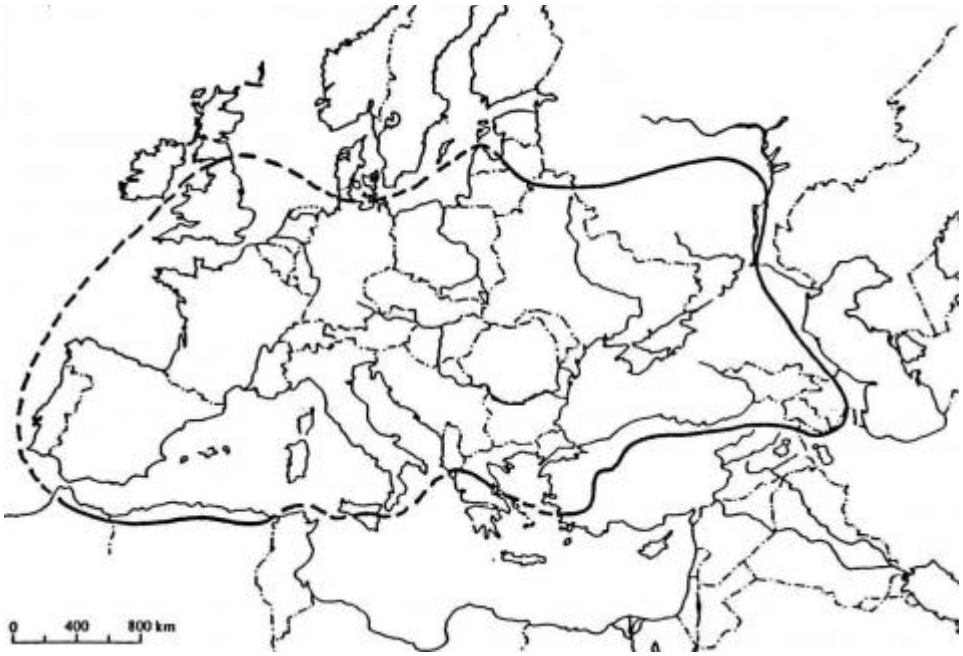


Figure 1.2: Distribution of wild *P. communis* (Zhoary, 1997)

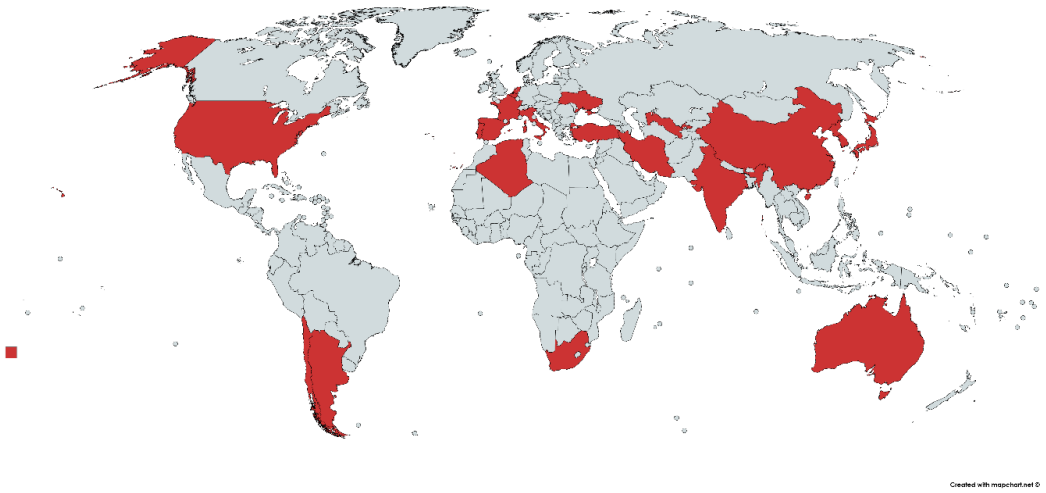


Figure 1.3: Top pear producer countries. (Map is created by Abdulbaki Çoban via mapchart.net by using FAOSTAT data. Top pear producers are colored red.)

1.3. Population Genetics and Molecular Markers

Population genetics is a branch of genetics which deals with the genetic diversity and change in the genetic diversity between and within populations. To do so, population genetics uses gene/allele data to calculate differences and similarities between individuals in a population and between populations. (Okasha, 2016) Therefore, it uses mathematics, namely statistics, to analyze data under the most important principle of population genetics: Hardy – Weinberg law. Simply, Hardy – Weinberg law states that allele frequencies in the gene pool of a population remain constant

over generations. This stability of allele frequencies is secured under certain assumptions:

- Mating should be random.
- There is no gene flow or migration (New alleles cannot be introduced into the gene pool).
- There is no mutation (New alleles cannot be produced by mutation).
- Population size is infinite.
- All individuals have equal chance to reproduce and survive (The gene/locus cannot affect selection) (Hartl and Clark, 1997).

If one or more of those assumptions are not met in a population, in other words, if there is inbreeding and/or gene flow and/or migration and/or mutations in a population, or population size is small, and/or the gene/locus of interest influences survivability or reproducibility, allele frequencies may change over generations.

As for molecular markers, a molecular marker can be defined as a certain DNA segment which reflects the genomic differences (Agarwal, 2008). To comprehend the evolutionary relationship between individuals or populations, it should be highly polymorphic (Cavalli-Sforza, 1998). Today, such molecular markers are used in many areas: Paternity tests, population genetics, gene mapping, forensics, and so on

(Schlötterer, 2004). An ideal molecular marker procedure and/or a molecular marker should have the followings:

- It should be distributed equally throughout the genome.
- It should be highly polymorphic.
- It should show the difference between heterozygous and homozygous alleles (It should be codominant).
- It should provide genomic / allelic differences easily.
- It shouldn't require prior information about the organism.
- It should be easy and cheap to use.
- It shouldn't be affected by environmental factors (Agarwal, 2008).

Taken into consideration those criteria, repetitive regions of the genome are useful tools as molecular markers. Tandem repeats (TR) are the repetitive DNA sequences that are dispersed throughout the genome (Thomas, 2005). According to their motif length, TRs are divided into two categories: microsatellites (or simple sequence repeats (SSRs) with unit size 1-10 bps) and minisatellites (unit size within 10-100 bps) (Mayer *et al.*, 2010). TRs are very unstable when compared to other parts of the genome. The mutation rate is between 10^{-3} to 10^{-6} per cell cycle. This high rate of mutations makes them suitable as molecular markers (Verstrepen *et al.*, 2005; Gemayel *et al.* 2010, 2012). As a subgroup of tandem repeats, SSR markers have the ideal marker criteria stated earlier. Besides, SSRs are very abundant in plant genomes which make them even suitable for plant conservation, ecology and/or population genetics studies.

1.4. Genetic diversity parameters

Genetic diversity analysis is the key point of the population genetic studies. To assess genetic diversity, some parameters should be calculated: Number of alleles (Na), number of effective alleles (Ne), expected heterozygosity (He), observed

heterozygosity (Ho), Shannon's information index (I), F-statistics (Fis, Fst, Fit), fixation index (F), proportion of polymorphic loci (P%), allelic richness (Ar), polymorphic information content (PIC), the probability of identity (PI), Garza-Williamson index (G-W), number of migrants (Nm), pairwise Fst values, and Hardy – Weinberg chi-square statistics (Li *et al.*, 2008; Nei, 1973, 1978; Petit *et al.*, 1998).

Number of alleles is the observed number of alleles while the number of effective alleles stands for the alleles which are equally frequent and give the same expected heterozygosity as the study. As for observed and expected heterozygosity, observed heterozygosity is the total number of individuals which are heterozygous for the locus and expected heterozygosity is the calculated probability of the heterozygosity.

One of the assumptions of Hardy-Weinberg rule was stated as "mating should be random" in the previous part. Inbreeding is one of the forms of nonrandom mating which can be defined as mating between relative individuals instead of nonrelative ones. This, eventually, causes to increase in homozygosity, therefore decline in fitness (Keller and Waller, 2002). The inbreeding coefficient (F) is the probability that a random allele is identical with its ancestor / descent. As for F-statistics, Fis defines the increase in the homozygosity in subpopulations due to inbreeding where Fit is the increase in the homozygosity in individuals due to non-random mating. Fst is also another F-statistics parameter that estimates the decline in the heterozygosity caused by genetic drift (Wright, 1965; Nei and Chesser, 1983).

Allelic richness (Ar) is another genetic diversity parameter which defines the average number of alleles per locus. It is an important measure for the future of the population since it shows the persistence and adaptability of the population (Greenbaum *et al.*, 2014). Polymorphic information content (PIC) is used to check the genotypic variation. The value is between zero and one. If there is not any variation, the value is close to zero and if the allelic variation is high, it is close to one (Guo and Elston,

1999). As for probability of identity (PI), it checks randomly selected two individuals whether they are multilocus genotypes or not. Thus, it gives an important information about the usefulness of the selected primer group (Paetkau *et al.*, 1995).

Garza-Williamson index (M ratio), is an indicator of population bottleneck. When it's applied to the studied loci, if the index is lower than 0.68 which is the critical value, therefore, it can be said that population was subjected to bottleneck(s) through past generations (Garza and Williamson, 2001). Lastly, pairwise Fst values and number of migrants (Nm) are used to detect the differentiation between populations. They are important parameters to check whether there are gene flow between studied populations or not.

1.5. Literature review of Genus *Pyrus* and *P. communis*

Due to its economic value as a commercial fruit, genus *Pyrus* is a well-studied genus in terms of genetics, morphology, and population around the world. However, studies about pear in Turkey are limited. As for molecular markers, SSR markers which were developed for the genus *Pyrus* and the genus *Malus* were used in characterization of the genus *Pyrus* (Gianfranceschi *et al.*, 1998; Liebhard *et al.*, 2002; Bassil *et al.*, 2004; Fernandez-Fernandez, 2006; Nishitani *et al.*, 2009; Inoue *et al.*, 2009; Yue *et al.*, 2014).

Developed SSR markers were used by several scientists around the world to determine genetic structure and diversity within and between populations of *P. communis, P. pyrifolia, P. ussurensis* and their varieties. (Yamamoto *et al.* 2001, 2002a, 2002c, 2007; Bao *et al.*, 2007; Katayama *et al.*, 2007; Brini *et al.*, 2008; Bassil and Postman, 2010; Cao *et al.*, 2012, Sehic *et al.*, 2012; Zhang *et al.*, 2013; Rana *et al.*, 2015, Liu *et al.*, 2015) Besides, genetic diversity analysis and genetic linkage map

construction between species of genus *Malus* and genus *Pyrus* were performed by using SSRs derived from *Pyrus* and *Malus* (Yao *et al.,* 2010).

In Turkey, however, only two studies were conducted on the genus *Pyrus* in terms of molecular markers. (Altınbay, 2012; Akçay *et al.*, 2014). However, both studies used materials from clone banks located in Turkey. In the first study (Akçay *et al., 2014*), population differentiation of *P. communis* according to their geographical locations (7 eco-graphical regions of Turkey) were tested. Even though population differentiation was shown in the study, they summed as high gene flow between populations. On the second study conducted on *Pyrus,* 46 genotypes of *P. communis* from Erzincan Horticultural Research Institute were used to reveal differentiation between individuals. Same as previous study, even though there was no or limited multilocus genotypes; high rate of hybridization and gene flow between individuals were found in this study.

1.6. Justification of the Study

Pear is an important pome fruit in Turkey due to its economic value. According to FAO (FAOSTAT, 2016), Turkey produce half a million metric tons of pear in an area harvested 25.4 thousand hectares. Besides, according to YMS (Turkish Sector of Fresh Fruits and Vegetables) statistics (YMS, 2017), Turkey exported 36 thousand metric tons of pear in 2017 with a trade volume of 20 million US dollars. By this export value, pear is the 10^{th} most exported fruit in Turkey. Beside the economic value of pears for Turkey, pear has more than 600 cultivars in Turkey. Pear has been cultivated in Turkey for longer than two thousand years. Along with the agricultural cultivation, they are found naturally in steppes and planted in cities for ornamental purposes. Wild varieties of *P. communis* and *P. eleagnifolia* are important trees for

steppes; since they are highly adaptive to those environments, they can be used as reforestation.

Even though there are large number of cultivars, few genetic structure and diversity analysis studies were conducted on it. Until today, only one study was conducted on genus *Pyrus* in terms of population structure and differentiation analysis (Akçay *et al.*, 2014). Therefore, this study is important for further studies conducted on genus *Pyrus.*

1.7. Aim of the Study

The main objective of the study was to reveal the genetic diversity structure of the *P. communis* cultivars from Artvin Province and to determine the differences / similarities between wild and domesticated *P. communis* and to compare with wild *P. eleagnifolia* population from METU campus in Ankara by using Simple Sequence Repeats markers. Besides, characterizing cultivars in terms of genetic differences, therefore, detecting important genetic resources was also aimed in the study. Understanding the diversity of the populations, revealing of the possible sources of the diversity between populations were also important goals of the study. Finally, since there are less number of studies conducting on the genus *Pyrus* in terms of SSR analysis in Turkey, this study was aimed to provide literature information for further studies.

CHAPTER 2

MATERIALS AND METHODS

2.1. Plant Material – Sample Collection

Plant materials for the study were collected from METU campus in Ankara and North – East part of Turkey (Artvin Province). 84 green, fresh leaf samples of *Pyrus communis* were collected from seven different villages, and the samples from same collection site were grouped and named as populations of where they were collected. Wild population of *P. communis* which were collected in Meydancık, however, named as Panta. Furthermore, 15 sample were collected in "Örnek Bahçe" formed by Artvin Directorate of Food, Agriculture and Livestock in Dalkırmaz village, Şavşat, Artvin. This population was named as Special Clone Collection (SCC). Beside the naming according to the villages in which the samples were collected, a secondary grouping was done according to the labels which are given by the locals considering the phenotypic variations between the fruit samples. 34 different *P. communis* phenotype were identified according to the local farmers (Appendix A). As for the samples in METU campus in Ankara, green, fresh leaf samples of *Pyrus elaeagnifolia* genotypes were collected. All the leaf samples were stored immediately in silica gel – filled bags to protect leaf samples until the DNA extraction. The information about the populations, their GPS (global positioning system) coordinates, altitudes, and village name were tabulated in Table 2.1. Sample tree of a studied *Pyrus communis* and its fruit were photographed in their natural habitats (Figure 2.1.). Besides, fruit samples of *Pyrus eleagnifolia* were photographed in Figure 2.2.

Table 2.1: Detailed information of studied populations

Pop No	**Species**	**N**	**Location**	**Latitude (N)**	**Longitude (E)**	**Altitude (m)**
1	*P. communis*	12	A.Koyunlu	41.294	42.493	1598
2	*P. communis*	14	Camili	41.480	41.900	508.4
3	*P. communis*	11	Kirazlı	41.265	42.493	1487
4	*P. communis*	6	Meşeli	41.315	42.470	1684
5	*P. communis*	10	Meydancık	41.456	42.228	1726
6	*P. elaeagnifolia*	20	METU	39.891	32.778	908
7	*P. communis*	15	SCC	41.253	42.355	1107
8	*P. communis*	6	Panta	41.456	42.228	1726
9	*P. communis*	10	Veliköy	41.315	42.432	1398

Figure 2.1: A sample of *P. communis* tree (on the left) and its fruits (on the right) (Photographed by M. Alev Ateş).

Figure 2.2: Samples of *P. eleagnifolia* fruits (Photographed by Dr. Zeki Kaya).

2.2. DNA Extraction and Quantification

Collected fresh leaves were dried by using silica gel-filled bags. Dried leaves were crushed in mortar by the help of liquid nitrogen. Until the isolation of the DNA, samples were put and stored in -80°C. For isolation of the DNA, an altered version

of CTAB (cetyltrimethylammonium bromide) protocol were used (Doyle and Doyle, 1987). Detailed and the whole information about the altered version of this protocol, buffers, and all the other solutions were given in the Appendix B and Appendix C.

Quality and quantity of the isolated DNA samples were measured by using NanoDrop Spectrophotometer (NanoDrop 2000, Thermo Scientific, USA). To test the quality of the DNAs, optical density (OD) of the samples were measured in 230, 260, and 280 nm. After the measurements, OD ratios between 1.8 and 2.0 for A260/A280 and above 1.5 for A260/A230 were selected for further experiments. For other samples whose ODs were not within the satisfactory range, isolation step was repeated until their OD values were satisfactory. Until PCR (Polymerase Chain Reaction), samples were stored in -80°C.

2.3. Methods

2.3.1. SSR Markers

In the beginning of the study, 19 microsatellite loci that were previously developed for the genus *Pyrus* (Yamamoto *et al.*, 2002a, Yamamoto *et al.*, 2002b, Yamamoto *et al.*, 2002c, Nishitani *et al.*, 2009) and the genus *Malus* (Gianfranceschi *et al.*, 1998, Liebhard *et al.*, 2002) species were tested. 11 of those 19 microsatellite primers which had highest polymorphism and good amplification, were selected for further studies.

To carry on the experiments for further stages, 19 microsatellite primers were tested in Polymerase Chain Reaction (PCR). During conducting PCR experiments, each primer was tested in different reaction components and different annealing temperatures by using a thermocycler (Eppendorf-Master cycler, Eppendorf, Canada) to get optimal conditions. As a PCR reaction mix, 5x HOT FIREPol® Blend Master Mix Ready to Load (Solis BioDyne, Tartu, Estonia) was used. After the PCR conditions were optimized, the most polymorphic 11 loci were selected for further studies. The detailed information about the selected ones were tabulated in Table 2.2.

Detailed information about the PCR cycles and reaction mixtures were given in the Table 2.3 and 2.4, respectively.

After PCR is comleted, 5µl of amplified products were loaded on 3% agarose gel which was prepared by using TBE (Tris-Borate-EDTA) buffer. Then, they were run on agarose gel in 120 mA electric current for approximately 30 minutes. Visualization were done under UV light. (Vilber Lourmat, France).

Table 2.2: Detailed information about the selected SSR primers

Primer name	**Forward Primer**	**Reverse Primer**	**Genus**	**Reference**
KU10	AGTATGTGACCACCCCGATGTT	AGAGTCGGTTGGGAAATGATTG	*Pyrus*	Yamamoto *et al.*, 2002a
BGT23b	CACATTCAAAGATTAAGAT	ACTCAGCCTTTTTTTCCCAC	*Pyrus*	Yamamoto *et al.*, 2002a
NH013a	GGTTTGAAGAGGAATGAGGAG	CATTGACTTTAGGGCACATTTC	*Pyrus*	Yamamoto *et al.*, 2002b
NB113a	ATGAAATATGTCGTGTTGCCCTTA	CCCTTCCTCAGCATGTTTCCTAGAC	*Pyrus*	Yamamoto *et al.*, 2002c
TsuENH008	CTGAGGTCTCATTCGGTGATTCT	CCTTCTCTGCTTTCTTCTTCACG	*Pyrus*	Nishitani et al., 2009
NH007b	TACCTTGATGGGAACTGAAC	AATAGTAGATTGCAATTACTC	*Pyrus*	Yamamoto *et al.*, 2002b
CH03g06	ATCCCACAGCTTCTGTTTTTG	TCACAGAGAATCACAAGGTGGA	*Malus*	Liedhard et al., 2002
NH008b	GGAAAAGAGAAGGAAGAAGAGAAGG	TGATAGGGGCATTTCGGTAA	*Pyrus*	Yamamoto *et al.*, 2002b
CH02B10	CAAGGAAATCATCAAAGATTCAAG	CAAGTGGCTTCGGATAGTTG	*Malus*	Gianfranceschi et al. 1998
CH02F06	CCCTCTTCAGACCTGCATATG	ACTGTTTCCAAGCGATCAGG	*Malus*	Gianfranceschi et al. 1998
CH01F02	ACCACATTAGAGCAGTTGAGG	CTGGTTTGTTTTCCTCCAGC	*Malus*	Gianfranceschi et al. 1998

Table 2.3: PCR components of each SSR marker.

SSR loci	Master Mix (1X)	Primer pairs (10µM)	Water	DNA (20ng/µl)	Total (µl)
NB113a NH013a KU10 NH007b CH03G06 CH02F06	5 µl	0.5+0.5 µl	9 µl	5 µl	20 µl
CH01F02 CH02B10	3 µl	0.6+0.6 µl	10.8 µl		
NH008b TsuEnh008 Bgt23b	4 µl	0.5+0.5 µl	10 µl		

Table 2.4: PCR cycle conditions of each SSR marker

Step	Temperature	Time	Number of Cycles	Description
1	94ºC	4 min.	1	Denaturation
2	94ºC	40 sec.	30	Denaturation
	Ta	30 sec.		Annealing
	72ºC	2 min.		Extension
3	72ºC	10 min.	1	Final Extension

2.3.2. Data Collection

11 microsatellite markers were selected according to their optimization and polymorphism after the optimization of the PCR cycle temperatures and reaction mixtures. Then, by using fluorescent dyes (Tamra, Fam, Hex), forward primers of

those markers were labeled by SACEM company in Ankara. 104 DNA samples were amplified by using newly synthesized fluorescently labeled primers.

To analyze fluorescently labeled PCR products, products were grouped as three markers per group according to difference in their length and their dye. In the Table 2.5, groupings of the markers with their respective fluorescent dyes and their annealing temperatures (Ta) were given.

Table 2.5: Grouping of the SSR markers with their respective fluorescent dyes and their Ta's

SSR Loci		Fluorescent Dye			Annealing Temperatures Ta's (°C)
		HEX	FAM	TAMRA	
	TsuEnh008		x		56
	Bgt23b	x			53
Group 1	**CH02B10**			x	56.2
	KU10	x			56
	NH007b		x		54
Group 2	**CH03G06**			x	56.2
	NH008b			x	54.6
	NB113a		x		56.9
Group 3	**NH013a**	x			57
	CH02F06		x		56.2
Group 4	**CH01F02**	x			59

Analysis of the fluorescently labeled PCR products were made by BM Labosis Company (Çankaya, Ankara). Analysis assay was performed by using Applied Biosystems 3730 XL DNA Analyzer (Applied Biosystems, Foster City, CA, USA), using an internal standard size marker (The GeneScan ROX labeled 400HD) (APPENDIX D).

When the analysis was done by BM Labosis Company, electropherograms were checked manually and allele sizes were scored by using Peak Scanner Software 2.0 (Applied Biosystems Inc., Foster City, CA, USA). Electropherogram of a genotype which is heterozygous for NH013a locus was showed in Figure 2.3. In Figure 2.4,

electropherograms of a genotype homozygous for NH007b locus and heterozygous for CH03G06 were presented. Lastly, in Figure 2.5, an electropherograms of a genotype which is heterozygous for CH02B10, homozygous for Bgt23b and TsuEnh008 were given. Read allele sizes were tabulated in an Excel file for further analysis. A part of the excel file can be found in APPENDIX E.

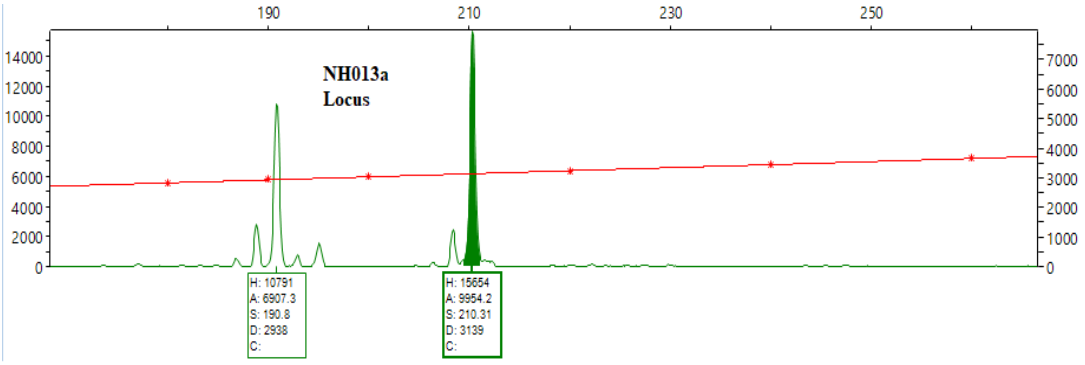


Figure 2.3: A phenogram showing a heterozygous genotype for NH013a SSR locus.

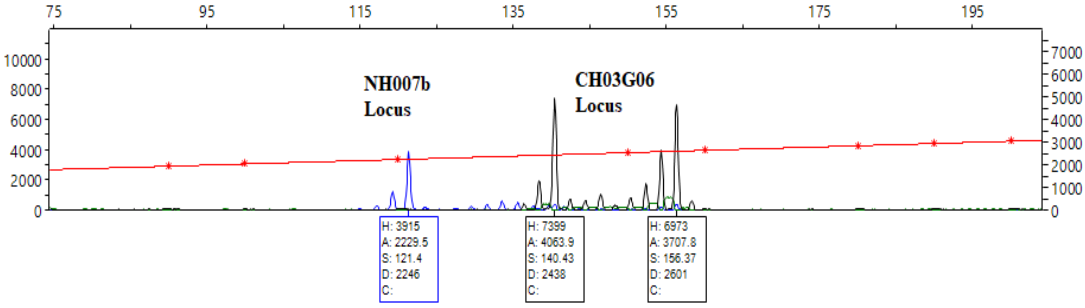


Figure 2.4: An electropherogram showing a genotype homozygous for NH007b locus and heterozygous for CH03G06 SSR locus.

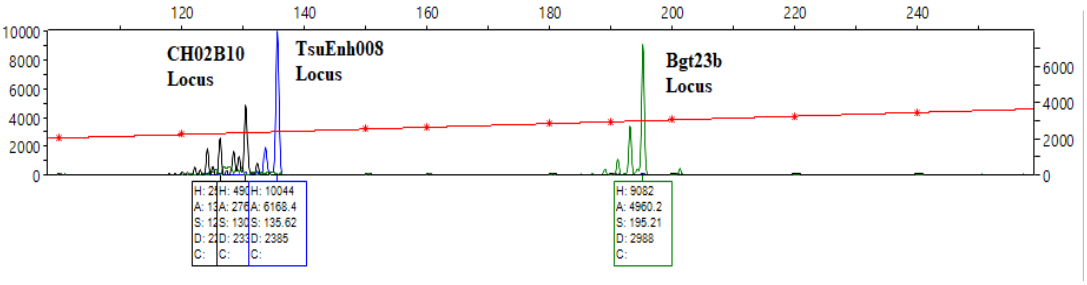


Figure 2.5: Electropherogram showing a genotype heterozygous for CH02B10 and homozygous for Bgt23b and TsuEnh008 SSR loci

2.4. Analysis of Data

Allelic data were converted into GDA (Genetic Data Analysis, Lewis and Zaykin, 2001), Genepop (Raymond and Rousset, 1995; Rousset, 2008), GenAlEx (Peakall and Smouse, 2012), MICRO-CHECKER (van Oosterhout et al., 2004) and STRUCTURE (Pritchard *et al.*, 2000) file formats by using a Python script coded by author (APPENDIX F and G, data types and Python scripts, respectively). Formatted data were used to import data into GDA software, STRUCTURE software, MICRO-CHECKER, GenAlEx and R environment for further analysis.

2.4.1. Quality of the Markers and Detecting uninformative loci

Firstly, allele dropouts, null alleles, stuttering and typographic errors of the studied SSR loci due to possible DNA degradation or primer-site mutations were checked by MICRO-CHECKER software with Bonferroni-adjusted 95% confidence interval. Presence of null alleles was checked by using the Brookfield 1 equation (1996).

Before going through analysis of the data, genepop – formatted data file was imported into an R script to check missing data and by using *poppr* package (Kamvar, Brooks & Grünwald, 2015), locus-wise Simpson index, evenness, expected heterozygosity were calculated to check for determining uninformative loci. Clone – correction assay was also performed to check distinct multilocus genotypes (MLGs) in R by using the same package. MLGs were also checked by using GenClone 2.0 (Arnaud-Haond and Belkhir, 2007). By computing index of association, Linkage disequilibrium of the studied loci was tested (All R scripts can be found in APPENDIX H).

Lastly, Hardy-Weinberg equilibrium of the loci and the populations were tested with GENEPOP (Raymond and Rousset, 1995; Rousset, 2008) software. Markov Chain method was used to calculate P-values with parameters of 1000 dememorizations, 100 batches, 1000 iterations.

2.4.2. Locus and Population-wise genetic diversity analysis

Locus-wise genetic diversity parameters; mean number of individuals (N), number of alleles (Na), number of effective alleles (Ne), allelic richness (Ar), polymorphic information content (PIC), Shannon's information index (I), observed heterozygosity (Ho), expected heterozygosity (He), fixation index (F), number of migrants (Nm), and F statistics (Fis, Fit, Fst) were calculated by GenAlEx software. As for population-wise genetic diversity parameters, number of alleles (N), number of effective alleles (Ne), probability of identity (PI), percentage of polymorphic loci (%P), observed and expected heterozygosity (Ho and He), fixation index (F) and F statistics (Fis, Fit, Fst) were also calculated by GenAlEx software (Statistical formulas can be found in APPENDIX I).

2.4.3. Genetic Structure Analysis

To determine genetic structure and differentiation among the populations of *Pyrus communis,* five different methods were used. Firstly, pairwise Fst values and number of migrants were calculated and results were tabulated by using GenAlEx and *poppr* (Kamvar, Brooks & Grünwald, 2015) and *adegenet* (Jombart, 2008) packages of R. Based on the calculated Fst values, PCoA analysis were performed to see how populations are differentiated.

Then, STRUCTURE software was used to determine genetic structure of populations by individuals with and without pre-assigned population information. Based on allelic data of individuals, Bayesian clustering methods were applied. The software were run three times; one with the samples including *P. eleagnifolia* collected from METU and two with only *P. communis* samples with and without prior population information. The runs were done with 10 replicates, 250,000 Markov Chain Monte Carlo (MCMC) replications after 50,000 burning length for 1 to 8 clusters. The true number of clusters (Δ K statistics) were calculated via STRUCTURE HARVESTER (A web-based software, Earl and vonHodt, 2012). For the calculated K, CLUMPP (CLUster Matching and Permutation Program) software was used to find coefficient

matrices (Jakobsson and Rosenberg 2007). Finally, CLUMPP output was used as an input of POPHELPER software to visualize the results (Francis, 2016).

Later, a phenogram was constructed using GDA software based on UPGMA (Unweighted pair group method with arithmetic mean) cluster analysis.

Since STRUCTURE assumes populations are panmictic and markers are not linked and uses Bayesian clustering algorithm, another clustering algorithm (such as K-means clustering) was needed to use to check STRUCTURE results. Thus, DAPC (Discriminant analysis of Principle component) analysis was performed to group the populations by using R package of adegenet (Jombart, 2008).

Lastly, Minimum Spanning Network (MSN) analysis was performed with the data to visualize the relationships between individuals rather than populations by using poppr and magrittr packages of R (Kamwar, Tabima, Grünwald, 2015). Even though MSN analysis gives more informative results for clonal populations, it is an important visualization method for relationships between individuals and populations. Besides, the studied population includes local names based on phenotypic variations assigned by locals (Appendix A). To check the relationship between phenotypic classification and genotypic variations, MSN is an important tool.

2.4.4. Analysis of Molecular Variance (AMOVA)

Two AMOVA assay were carried out by using ARLEQUIN (Excoffier *et al.*, 2010) to portion the genetic variance between individuals, between populations, and within populations. For the first assay, all individuals of both species (*P. communis* and *P. eleagnifolia*) were used. In the second one, *P. communis* individuals were grouped according to their geographical distiribution.

CHAPTER 3

RESULTS

3.1. SSR marker selection

The selection process of suitable SSR loci was started with 19 SSR markers. Those 19 SSR markers were screened to acquire suitable ones for detecting polymorphisms among 84 genotypes of *P. communis* (104 when *P. eleagnifolia* genotypes included). Out of those 19 SSR markers, 11 were selected for further analysis with high allelic polymorphism. Due to the allele size differences between *P. communis* and *P. eleagnifolia*, allele data (alleles, number of alleles, and allele size ranges) were given in two different tables (Table 3.1 and Table 3.2). In both tables, the loci which are marked with asterisks (*) (KU10, CH03G06, Bgt23b) have similar allele size between both species, the others are significantly distinct. Missing data were found as 0.96% (Figure 3.1).

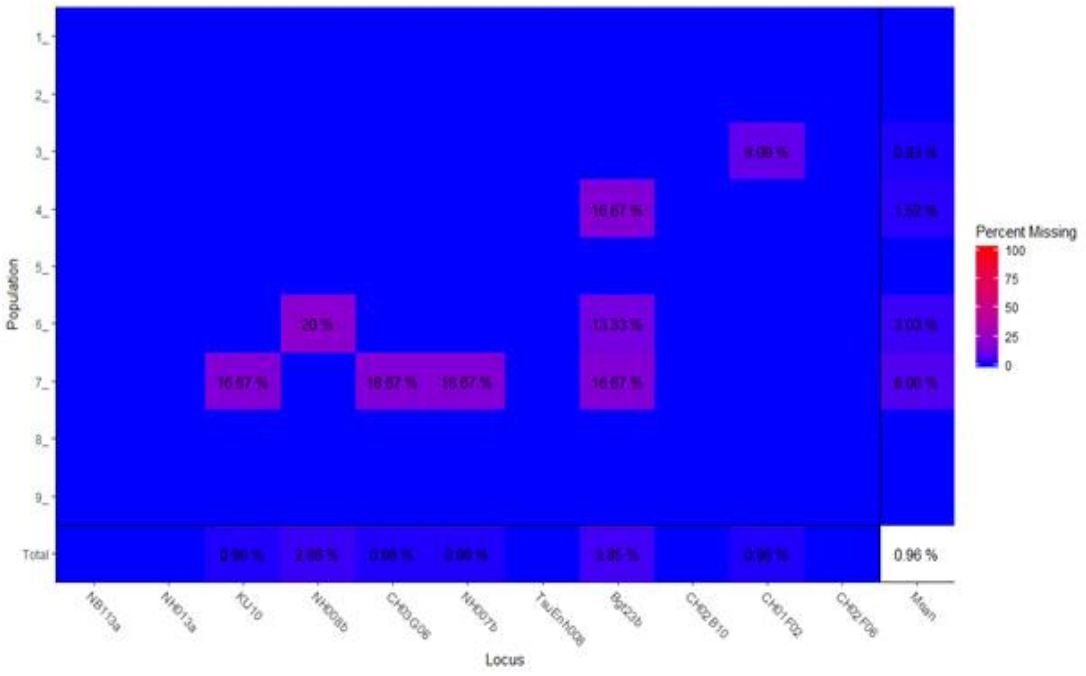


Figure 3.1: Percentage of missing data in each locus and population

Table 3.1: SSR loci allele size range in *P. communis*

SSR Loci	Allele Size Range	Number of Alleles	Alleles
NB113a	136-158	12	136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158
NH013a	185-225	12	185, 191, 193, 195, 197, 199, 205, 209, 211, 215, 219, 225
NH008b	185- 211	9	185, 191, 193, 195, 197, 201, 205, 207, 209, 211
KU10*	228-280	18	228, 230, 232, 234, 236, 240, 248, 250, 252, 258, 260, 262, 266, 268, 270, 274, 276, 280
CH03G06*	136-172	13	136, 140, 144, 146, 150, 152, 154, 156, 160, 162, 168, 170, 172
NH007b*	120-142	11	120, 122, 126, 128, 130, 132, 134, 136, 138, 140, 142
TsuEnh008	134-144	6	134, 136, 138, 140, 142, 144
Bgt23b*	187-229	18	187, 189, 191, 193, 195, 197, 199, 203, 205, 207, 209, 211, 213, 215, 221, 223, 225, 229
CH02B10	120-132	7	120, 122, 124, 126, 128, 130, 132
CH01F02	161-183	11	161, 163, 165, 167, 171, 173, 175, 177, 179, 181, 183
CH02F06	150-196	15	150, 154, 156, 164, 168, 170, 172, 174, 176, 178, 186, 190, 192, 194, 196

Table 3.2: SSR loci allele size range in *P. eleagnifolia*

SSR Loci	Allele Size Range	Number of Alleles	Alleles
NB113a	140-180	14	140, 144, 148, 154, 156, 158, 160, 162, 164, 170, 172, 174, 176, 180
NH013a	161-199	12	161, 165, 167, 169, 171, 173, 175, 177, 181, 185, 187, 199
NH008b	155- 187	12	155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 187
KU10*	224-286	13	224, 230, 232, 234, 240, 242, 244, 250, 252, 256, 260, 262, 286
CH03G06*	140-172	9	140, 144, 146, 150, 152, 154, 160, 164, 172
NH007b	120-152	10	122, 126, 128, 130, 132, 134, 136, 148, 150, 152
TsuEnh008	134-158	10	134, 140, 142, 144, 146, 150, 152, 154, 156, 158
Bgt23b*	175-233	16	175, 183, 195, 197, 199, 201, 203, 205, 209, 211, 213, 215, 217, 221, 223, 233
CH02B10	146-196	8	146, 162, 186, 188, 190, 192, 194, 196
CH01F02	181-211	9	181, 193, 195, 197, 201, 203, 205, 209, 211
CH02F06	132-158	7	132, 136, 138, 142, 146, 148, 158

Table 3.3: Estimated null allele frequencies of SSR loci for the studied *P. communis* populations

	A.Koyunlu	**Camili**	**Kirazlı**	**Meşeli**	**Meydancık**	**SCC**	**Panta**	**Veliköy**
NB113a	-0.0802	-0.0515	-0.0828	0.0323	0.0164	-0.0062	0.1756*	-0.0411
NH013a	-0.0038	0.108*	0.0793	0.2727*	0.1713*	0.0842	0.2821*	0.0761
KU10	0.0579	0.0502	0.0279	-0.0827	0.0423	-0.006	0.0323	0.0426
NH008b	-0.112	0.0482	-0.0121	-0.0313	0.0831	0.05	0.129	0.0341
CH03G06	-0.0611	-0.0479	-0.0023	-0.0909	0.0476	-0.0419	0.0698	-0.0286
NH007b	-0.0134	0.064	0.0649	-0.0313	0.0805	0.1187*	0.0123	0.0296
TsuEnh008	0.0637	0.2666*	0.1436	-0.0313	0.1011	0.1045	0.0609	0.1111
Bgt23b	0.2411*	0.0535	0.0876	0.1566	-0.0141	-0.0297	-0.0753	0.1957*
CH02B10	-0.1604	-0.1556	-0.0746	-0.1379	-0.1429	-0.0438	-0.1892	-0.1834
CH01F02	-0.0951	-0.12	-0.1173	-0.0909	-0.084	-0.0432	-0.0732	-0.0899
CH02F06	0.1076	-0.0687	0.0211	0.0625	-0.0795	0.0076	0.0476	0.0368

*Null alleles may be present according to Brookfield 1 equation.

Identifications of null alleles were performed according to Brookfield1 equation (Brookfield, 1996) by using MICRO-CHECKER software. Even though some loci are appeared to have higher null allele frequencies, it was found that they did not affect the further statistical analyses significantly. Therefore, it was assumed that there was no null alleles present and the estimations of genetic parameters were carried out without excluding the null alleles in loci.

As for linkage disequilibrium (LD) assessment, pairwise indices of association were calculated and heatmap was created by using poppr package in R (Figure 3.2). Index of association value between CH03G06 and NH007b was calculated slightly higher than others as 0.29. However, the pairwise index of association values of the samples vary between 0.29 and -0.113. Thus, no significant linkage LD was found between studied loci. To check whether there are clones in the studied populations, clone – correction assay was performed. Two genotypes in the "Meydancık" population were found to be clones. Therefore, one of the clones was excluded for further studies.

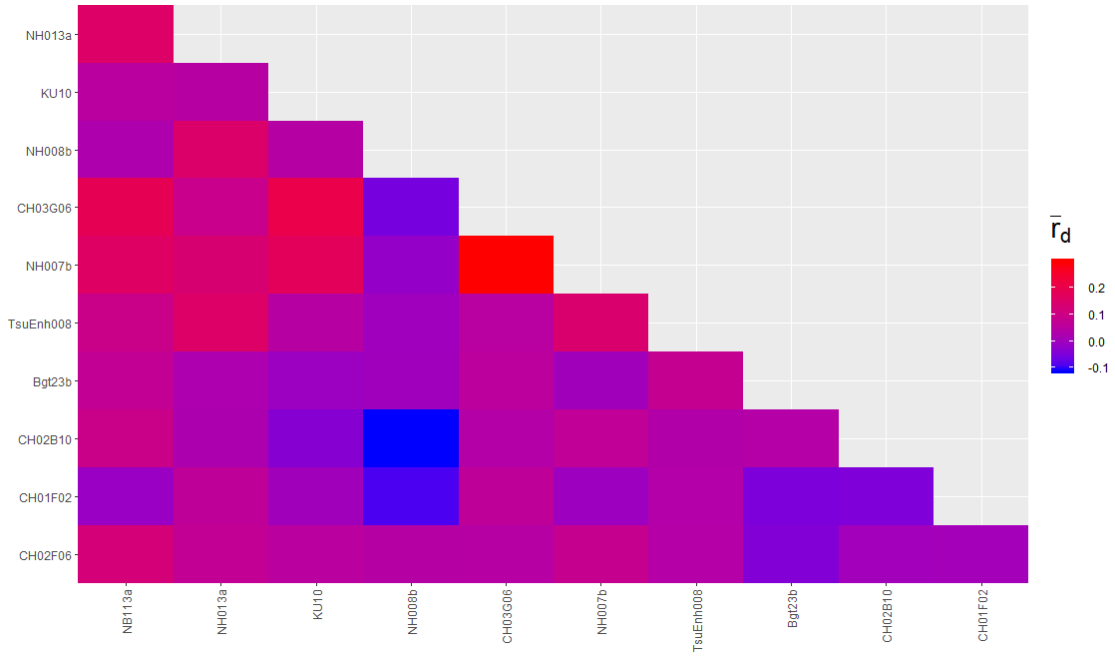


Figure 3.2: Pairwise index of association of 11 loci

3.2. Genetic Diversity of Loci and Populations

3.2.1. Genetic Diversity of Loci

As for locus-wise descriptive statistics, mean number of individual (N), mean number of different alleles (Na), mean number of effective alleles (Ne), allelic richness (Ar), private alleles (Pa), Polymorphic information content (PIC), Shannon index (I), observed heterozygosity (Ho), expected heterozygosity (He), fixation index (F), and Hardy-Weinberg equilibrium (HWE) were estimated and given in Table 3.4. The average mean of different allele was calculated as 7.06 ranging from 5.0 to 9.38. Besides, number of effective alleles was found as 4.74 ranging from 2.68 to 6.65.

Informativeness of each locus is related to Allelic richness, Polymorphic Information Content, and Shannon Index which varied from 3.63 (CH02B10) to 5.99 (KU10), 0.62 (CH02F06) to 0.83 (NH013a), and 1.22 (CH02B10) to 2.028 (KU10), respectively. Since markers with PIC values higher than 0.5 are considered as informative, all the studied markers are found to be informative.

Observed heterozygosity of the loci varied from 0.55 (TsuEnh008) to 0.97 (CH01F02) with a mean of 0.74 while the expected heterozygosity range between 062 (CH02B10) and 0.84 (KU10) with a mean of 0.76. With a negative Fixation index, excess of heterozygosity was observed in NB113a, CH03G06, CH02B10, and CH01F02. The rest of 11 loci were with positive fixation indices.

Table 3.4: Locus-wise descriptive statistics

SSR Locus	N	Na	Ne	Ar	PIC	Ho	He	F	HWE
NB113a	10.50±1.16	7.50±0.57	5.31±0.34	5.23	0.78	0.81±0.06	0.80±0.01	-0.01±0.66	ND
NH013a	10.50±1.16	7.62±0.68	5.74±0.53	5.32	0.83	0.57±0.08	0.81±0.03	0.31±0.09	ND
KU10	10.38±1.24	9.38±0.78	6.65±0.53	5.99	0.82	0.80±0.03	0.84±0.02	0.04±0.04	ND
NH008b	10.12±1.01	6.62±0.38	4.07±0.26	4.63	0.70	0.71±0.06	0.75±0.02	0.06±0.06	NS
CH03G06	10.38±1.24	6.88±0.64	4.24±0.48	4.67	0.63	0.78±0.04	0.74±0.02	-0.05±0.04	NS
NH007b	10.38±1.24	6.88±0.72	4.90±0.50	4.9	0.80	0.70±0.04	0.78±0.03	0.09±0.04	ND
TsuEnh008	10.50±1.65	5.38±0.32	3.83±0.34	4.23	0.67	0.55±0.06	0.72±0.03	0.25±0.07	***
Bgt23b	10.00±1.20	7.75±0.70	5.13±0.53	5.25	0.88	0.65±0.08	0.79±0.02	0.18±0.09	ND
CH02B10	10.50±1.16	5.00±0.46	2.68±0.17	3.63	0.76	0.84±0.03	0.62±0.02	-0.36±0.05	***
CH01F02	10.38±1.16	7.50±0.33	5.49±0.40	5.39	0.77	0.97±0.02	0.81±0.02	-0.20±0.02	ND
CH02F06	10.50±1.16	7.12±0.30	4.07±0.16	4.74	0.62	0.72±0.03	0.75±0.01	0.04±0.05	NS
Mean	10.38±0.33	7.06±0.20	4.74±0.16		0.75	0.74±0.02	0.76±0.01	0.03±0.03	

N=mean number of individuals with amplification, Na=mean number of different alleles, Ne=mean number of effective alleles, Ar=allelic richness, PIC=polymorphic information content, Ho=observed heterozygosity, He=expected heterozygosity, F=fixation index, HWE=Hardy Weinberg equilibrium (ND: non-deviating, NS: nonsignificant,***:p<0,001,**:p<0,01,*:p<0,05)

Table 3.5: Locus-wise F-statistics

	Fis	**Fit**	**Fst**	**Nm**
NB113a	-0.01	0.05	0.06	3.95
NH013a	0.29	0.35	0.08	2.94
KU10	0.04	0.09	0.04	5.56
NH008b	0.05	0.13	0.08	2.99
CH03G06	-0.04	0.02	0.06	3.75
NH007b	0.09	0.16	0.07	3.27
TsuEnh008	0.24	0.29	0.07	3.33
Bgt23b	0.17	0.25	0.09	2.47
CH02B10	-0.35	-0.32	0.02	9.62
CH01F02	-0.20	-0.14	0.05	4.80
CH02F06	0.04	0.11	0.07	3.15
Mean	0.03	0.09	0.06	4.17
SE	0.06	0.06	0.01	0.61

In Table 3.5, locus-wise F-statistics (Fis, Fit, Fst) were given. The mean value of Fis (inbreeding coefficient within individuals) was calculated as 0.03 and ranged from -0.35 (CH02B10) to 0.29 (NH013a). As for Fit (inbreeding coefficient within total population) and Fst (inbreeding coefficient within subpopulations), the average values are 0.09 and 0.06 respectively. The mean number of migrants was estimated as 4.17 and with a value of 9.62 in CH02B10 which had the highest value of number of migrants while Bgt23b had the lowest value of 2.47.

3.2.2. Genetic Diversity of the Populations

Regarding population – wise descriptive statistics, number of individuals (N), mean number of different alleles (Na), mean number of effective alleles (Ne), private alleles (PI), percentage of polymorphic loci (%P), Garza-Williamson index (G-W index (M)), observed heterozygosity (Ho), expected heterozygosity (He), and fixation index (F) were calculated and given in Table 3.6. All populations have high polymorphism rate with 100%.

Table 3.6: Population-wise descriptive statistics of population genetic parameters.

Population	**N**	**Na**	**Ne**	**Pa**	**P (%)**	**Ho**	**He**	**F**
A.Koyunlu	12	7.27±0.54	4.56±0.42	2	100.00%	0.76±0.06	0.76±0.03	-0.02±0.08
Camili	14	7.64±0.62	4.69±0.53	4	100.00%	0.73±0.06	0.76±0.03	0.03±0.09
Kirazlı	11	8.00±0.38	4.92±0.49	2	100.00%	0.75±0.05	0.78±0.02	0.03±0.06
Meşeli	6	5.82±0.35	4.43±0.40	1	100.00%	0.75±0.06	0.75±0.02	0.01±0.09
Meydancık	10	6.36±0.31	4.54±0.33	1	100.00%	0.73±0.04	0.76±0.02	0.04±0.07
SCC	15	7.91±0.44	5.13±0.41	2	100.00%	0.76±0.04	0.79±0.02	0.04±0.04
Panta	6	5.36±0.58	4.03±0.54	3	100.00%	0.64±0.07	0.71±0.03	0.10±0.10
Veliköy	10	8.09±0.64	5.57±0.48	1	100.00%	0.77±0.05	0.81±0.02	0.03±0.07
Mean	10.38±0.33	7.06±0.20	4.74±0.16		100.00%	0.74±0.02	0.76±0.01	0.03±0.03

The mean number of different alleles was calculated as 7.56. The Na varied between 8.09 and 5.36 as Veliköy had the highest (8.09) and Panta (5.36) had the lowest. The number of effective alleles (Ne) varied between 4.03 and 5.57 and the average value of it was calculated as 4.74. Similar to the Na values, Veliköy population had the highest (5.57) Ne and Panta population had the lowest (4,04). Between populations, private allele count ranged between 1 to 4. All populations had at least one private allele. All private alleles were given in the APPENDIX J.

The average observed heterozygosity was calculated as 0.74 and the expected heterozygosity is 0.76. The observed heterozygosity ranged between 0.77 (Veliköy) and 0.64 (Panta) while expected heterozygosity varied from 0.81 (Veliköy) to 0.71 (Panta). Out of 8 populations, excess of heterozygosity is only seen in A.Koyunlu population with F value of -0,02.

3.3. Genetic Differentiation and Structure of the populations

3.3.1. Pairwise Fst and Number of Migrants

Pairwise Fst values range between 0,023 (between Camili and Kirazlı) and 0,068 (between Panta and Meşeli) between populations, thus, number of migrants vary from 3,421 to 10,646. In Table 3.7, pairwise Fst and Nm values were given. (below diagonal is pairwise Fst and above diagonal is Nm). Besides, in Figure 3.3, a heatmap of pairwise Fst values was given.

Table 3.7: Pairwise Fst (below diagonal) and number of migrants (above diagonal) between populations

	A.Koyunlu	**Camili**	**Kirazlı**	**Meşeli**	**Meydancık**	**SCC**	**Panta**	**Veliköy**
A.Koyunlu	-	8.737	7.246	6.363	4.426	7.793	4.126	10.272
Camili	0.028	-	10.644	8.464	5.096	9.681	4.792	10.312
Kirazlı	0.033	0.023	-	7.229	7.674	9.609	4.462	10.311
Meşeli	0.038	0.029	0.033	-	6.712	8.167	3.421	10.599
Meydancık	0.056	0.047	0.032	0.036	-	8.73	3.758	7.256
SCC	0.031	0.025	0.025	0030	0.028	-	3.693	9.26
Panta	0.057	0.050	0.053	0.068	0.062	0.063	-	4.855
Veliköy	0.024	0.024	0.024	0.023	0.033	0.026	0.049	-

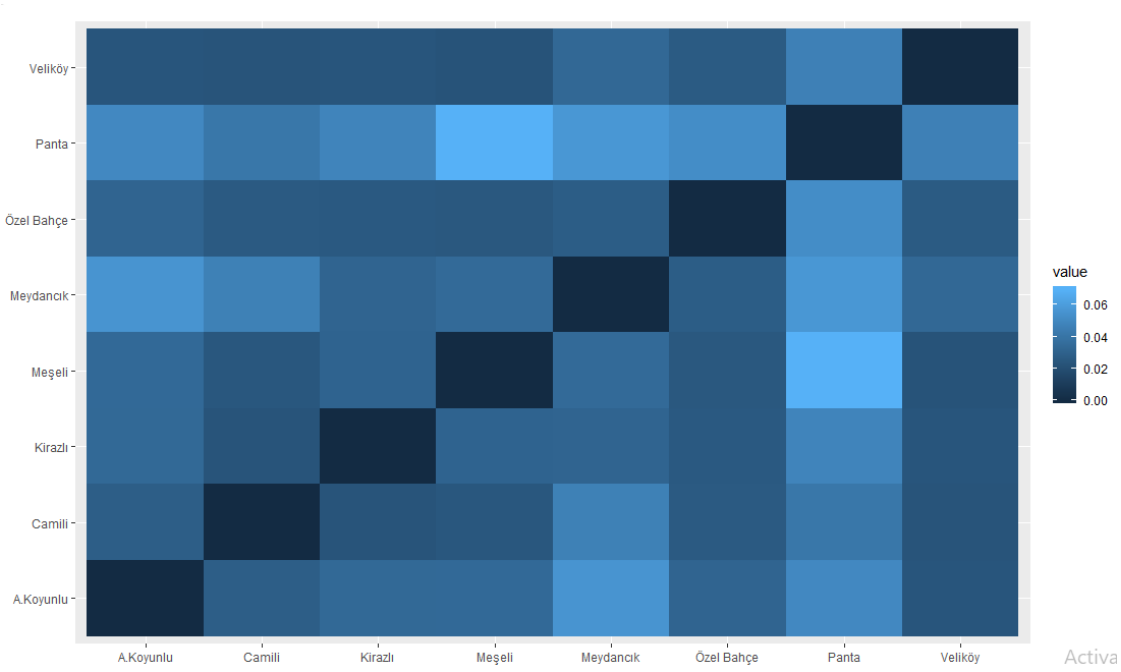


Figure 3.3: Heatmap of pairwise Fst values between populations.

3.3.2. UPGMA phenograms

Two UPGMA phenograms were constructed; one with only *P. communis* samples and the other with samples from both *P. communis* and *P. eleagnifolia.* (Figure 3.4 and Figure 3.5) The "panta" population is genetically most distant to other populations. It is followed by "Meydancık" population. The other populations, with respect to Panta and Meydancık, are genetically close to each other. (Figure 3.4). As for *P. eleagnifolia* samples, expectedly, the population named "METU" is distantly related to the *P. communis* populations. (Figure 3.5)

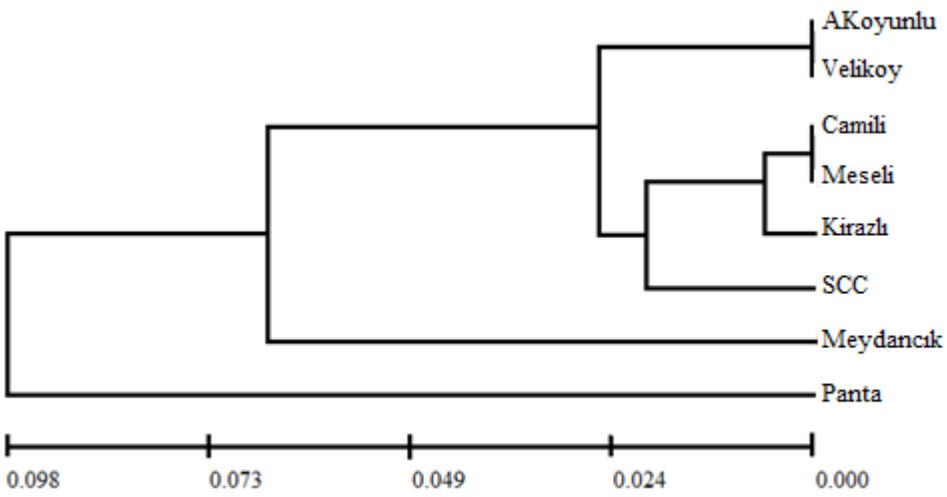


Figure 3.4: UPGMA phenogram based on coancestory identity of studied *P. communis* populations

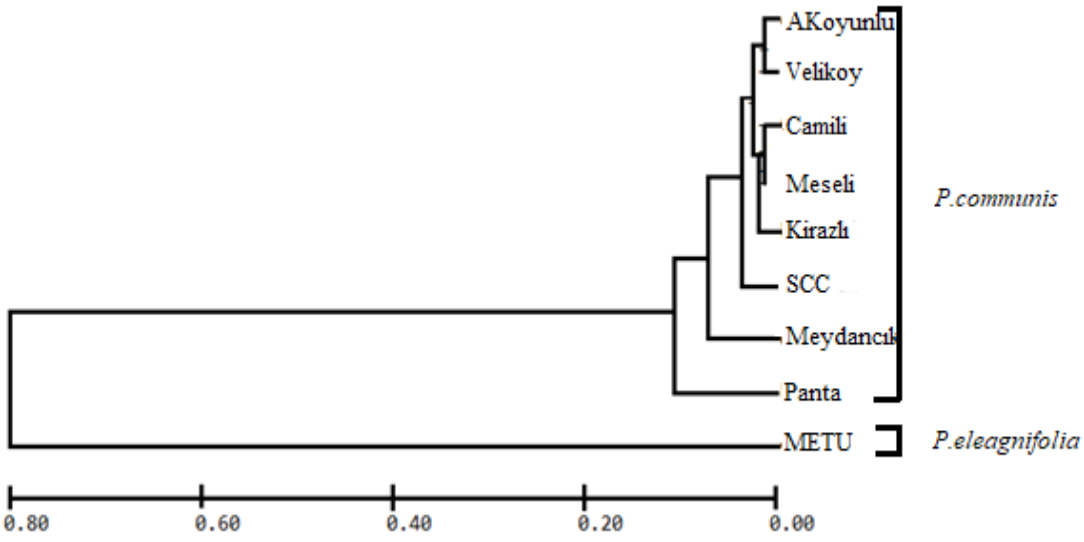


Figure 3.5: UPGMA phenogram based on coancestory identity of studied *P. communis* and *P. eleagnifolia* populations.

3.3.3. Population Genetic Structure

Three STRUCTURE analyses were performed: One including samples from METU and the other two with only populations of *P. communis.* The analyses with only *P. communis* populations were conducted with and without prior population information. However, in both cases, delta K were estimated as 2. Therefore, the result of the analysis of with population information was not included. The graph of delta K and Evanno method using each delta K were shown in Figure 3.6. and Table 3.8. Individuals in the all populations were clustered into two groups not relatedly with their original populations, except the wild population Panta. The members of the populations clustered into the cluster 1 with membership values ranging between 40 % and 85.7 %. However, all members of the Panta were found to be in cluster 1. Proportions of the population clusters and graph of cluster membership of 83 individuals can be found on Table 3.9. and Figure 3.7, respectively.

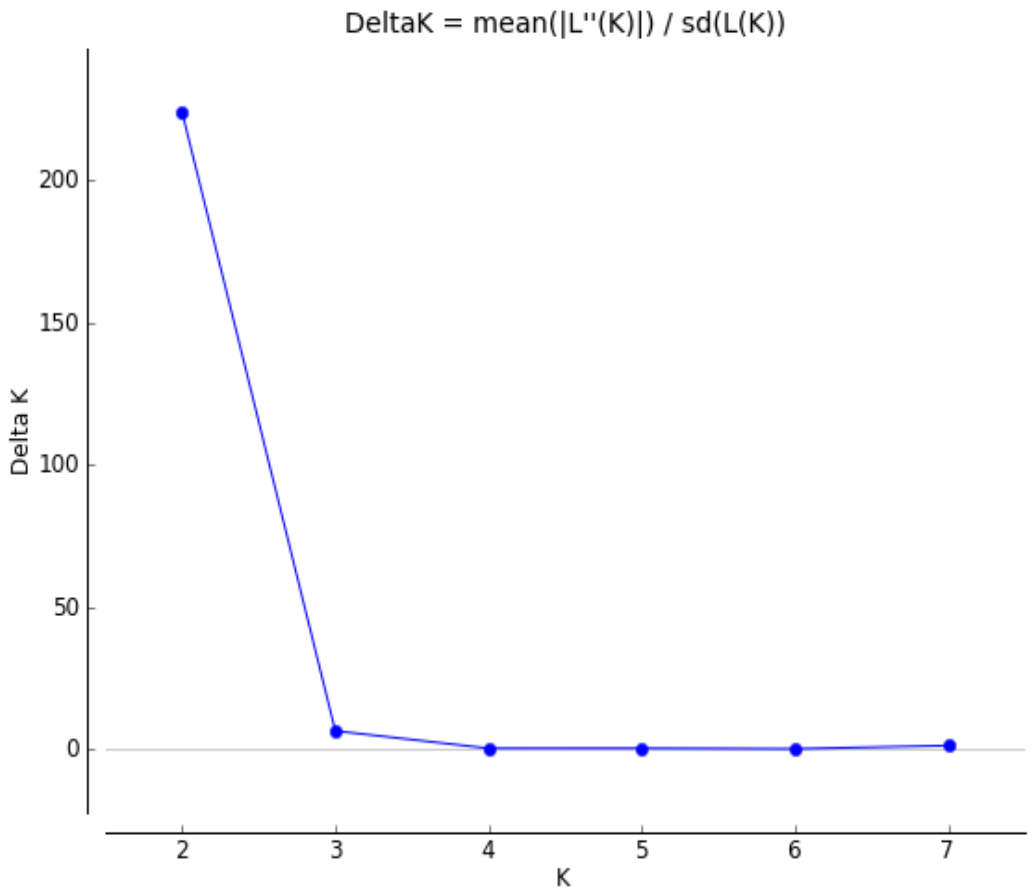


Figure 3.6: Delta K values graph of 8 population of *P.communis* without prior population information

Table 3.8: Evanno method using delta K

K	**Reps**	**Mean LnP(K)**	**Stdev LnP(K)**	**Ln'(K)**	**|Ln''(K)|**	**Delta K**
1	10	-3739.41	0.27	NA	NA	NA
2	**10**	**-3526.73**	**0.57**	**212.68**	**128.09**	**224.07**
3	10	-3442.14	5.81	84.59	37.31	6.42
4	10	-3394.86	8.95	47.28	2.74	0.31
5	10	-3350.32	16.18	44.54	5.12	0.32
6	10	-3310.90	55.94	39.42	7.58	0.14
7	10	-3263.90	28.94	47.00	36.65	1.27
8	10	-3253.55	59.28	10.35	NA	NA

Table 3.9: Assigned individuals of population into two clusters according to STRUCTURE analysis

Pop	N	Cluster 1	Cluster 2
A.Koyunlu	12	9 (75%)	3 (25%)
Camili	14	12 (85.7%)	2 (14.3%)
Kirazlı	11	7 (63.6%)	4 (36.4%)
Meşeli	6	3 (50%)	3 (50%)
Meydancık	9	4 (44.4%)	5 (55.6%)
SCC	15	6 (40%)	9 (60%)
Panta	6	6 (100%)	0 (0%)
Veliköy	10	6 (60%)	4 (40%)

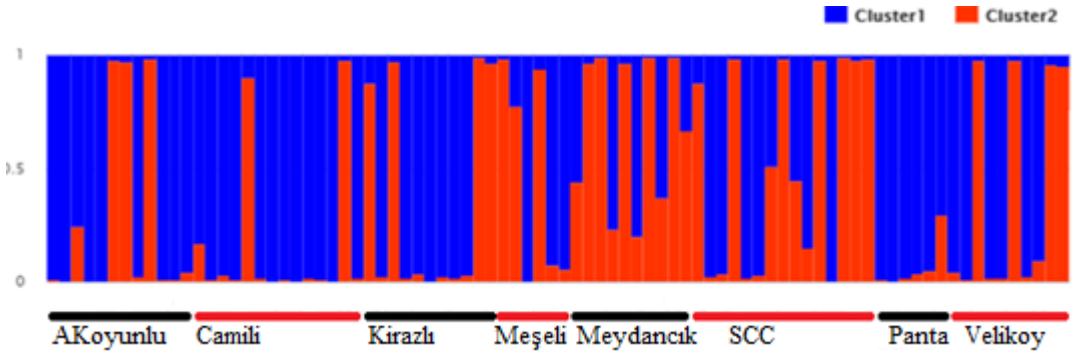


Figure 3.7: Graph of cluster membership of the studied 83 individuals

As for the population structure analysis of *P. communis* samples from Artvin and *P. eleagnifolia* samples from METU campus, delta K value was calculated as 2. The results of with and without prior population information did not affect the results. Graphical representation of delta K value and Evanno method based on delta K values were shown in Figure 3.8 and Table 3.10, respectively. Assignment of the individuals to each cluster and their membership values were given in Table 3.11. Finally, graphical representation of membership assignment of the individuals was provided in Figure 3.9. Differently from the previous STRUCTURE analysis, two clusters were separated two populations of *P. communis* and *P. eleagnifolia.* While all members of *P. communis* were clustered into the cluster 1, all members of the *P. eleagnifolia* population were clustered into cluster 2.

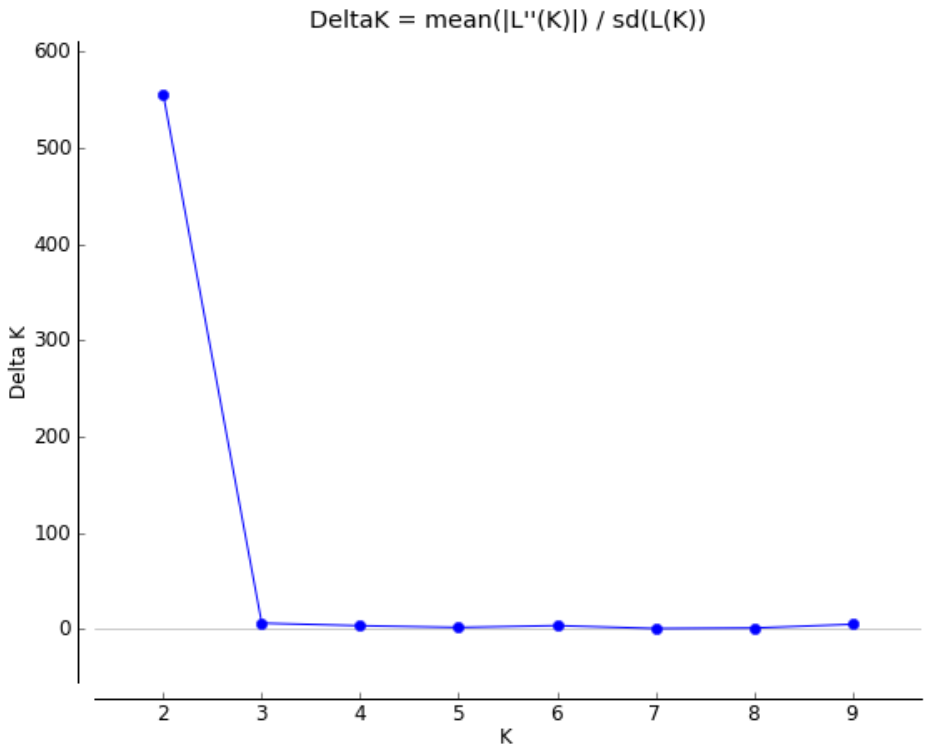


Figure 3.8: Delta K values of total analysis of *P. communis* and *P. eleagnifolia*

Table 3.10: Evanno method using K

K	**Reps**	**Mean LnP(K)**	**Stdev LnP(K)**	**Ln'(K)**	**|Ln''(K)|**	**Delta K**
1	10	-5469.87	0.9129	NA	NA	NA
2	**10**	**-4808.54**	**0.7863**	**661.33**	**436.93**	**555.69**
3	10	-4584.14	19.81	224.40	114.92	5.79
4	10	-4474.66	20.31	109.48	63.93	3.14
5	10	-4429.11	27.11	45.55	36.54	1.34
6	10	-4347.02	13.89	82.09	45.21	3.25
7	10	-4310.14	32.19	36.88	7.66	0.23
8	10	-4265.60	29.18	44.54	22.23	0.76
9	10	-4243.29	32.66	22.31	150.05	4.59
10	6	-4371.03	251.47	-127.74	NA	NA

Table 3.11: Number of individuals and their membership values in each cluster

Pop	N	Cluster 1	Cluster 2
A.Koyunlu	12	12 (100%)	-
Camili	14	14 (100%)	-
Kirazlı	11	11 (100%)	-
Meşeli	6	6 (100%)	-
Meydancık	9	9 (100%)	-
SCC	15	15 (100%)	-
Panta	6	6 (100%)	-
Veliköy	10	10 (100%)	-
METU	20	-	20 (100%)

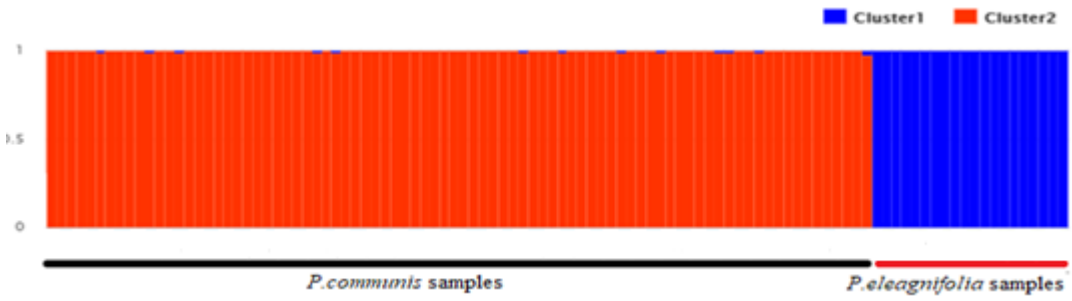


Figure 3.9: Graphical representation of cluster assignment of each studied individuals (Cluster 1 is composed of only *P. communis* samples, cluster 2 samples contains only *P. eleagnifolia* population).

3.3.4. Minimum Spanning Network (MSN) Analysis

To understand and visualize the distance between populations and individuals Minimum Spanning Network (MSN) analysis was carried out. Similar to STRUCTURE analysis, individuals are separated from each other not related with their originated populations. Only Panta population members were closely grouped together at the bottom of the graph in Figure 3.10. (Individuals with number between 70-75).

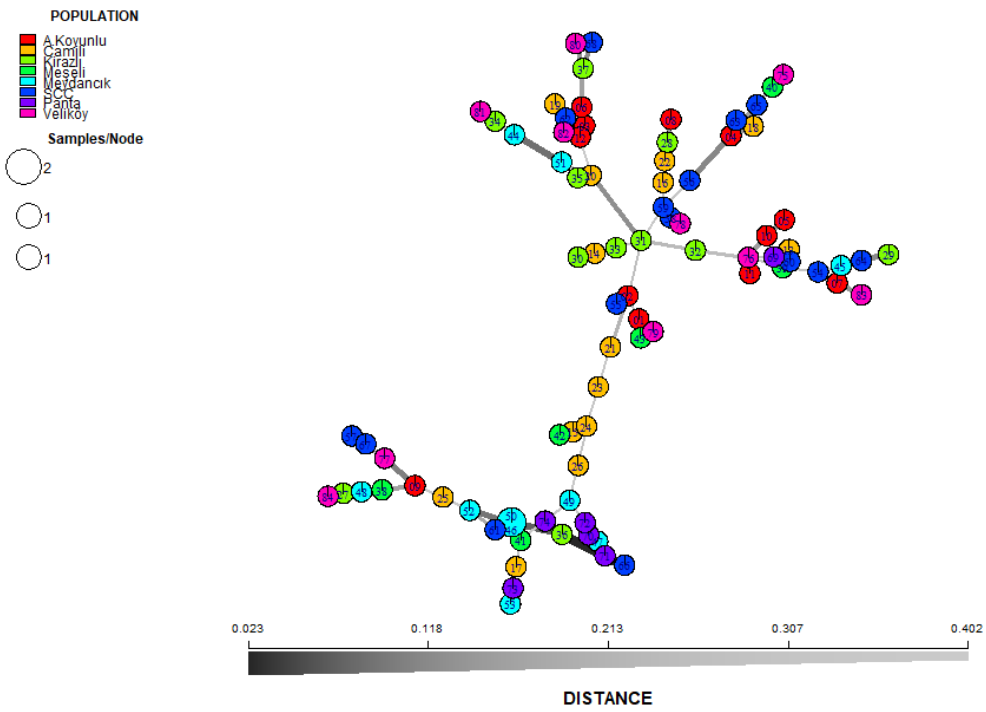


Figure 3.10: Minimum spanning network analysis of the 84 *P. communis* individuals.

Beside the population differentiation, MSN analysis was performed once more for the samples with known local names given by local farmers. It is found that there is significant correlation between the genotype of samples and their corresponding local names. While most of the individuals within the same group were closely grouped together, individuals in the groups named “Acara”, “Işık Mehmet”, and “Ruma” were found to be genotypically distant to each other. Moreover, individuals in the “Panta” population were found in the center of the other populations. In Figure 3.11, MSN map of samples and their corresponding names can be found.

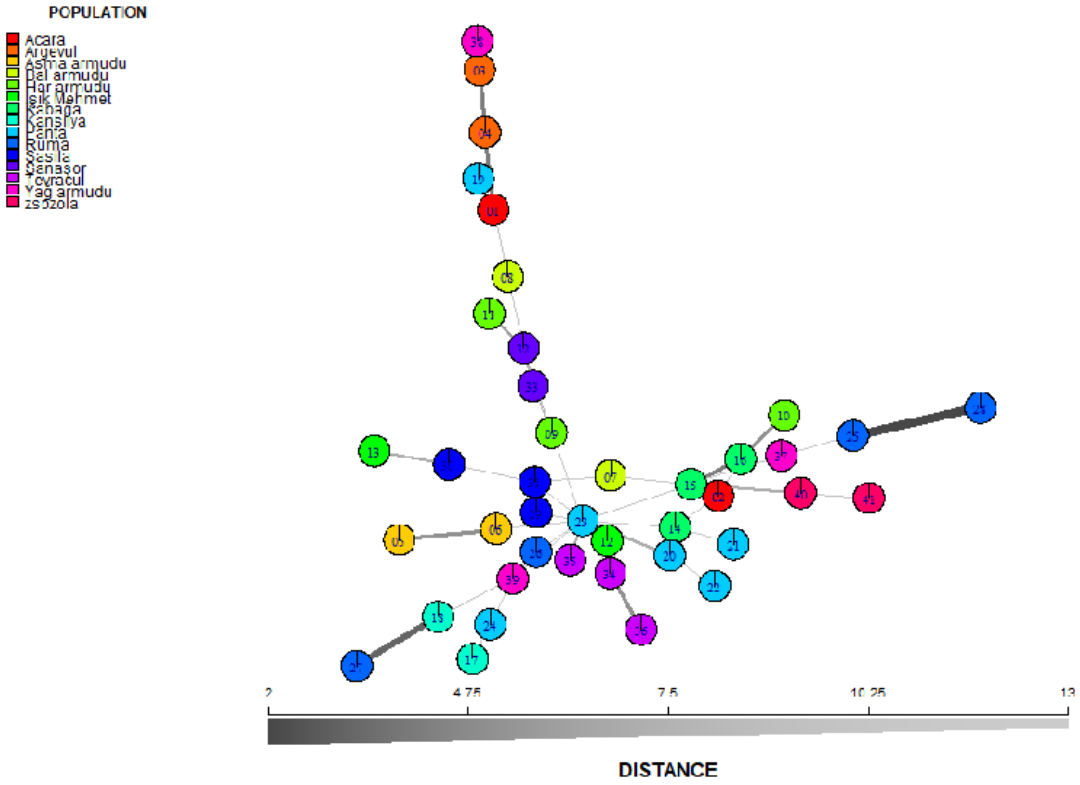


Figure 3.11: MSN map of samples with known local names

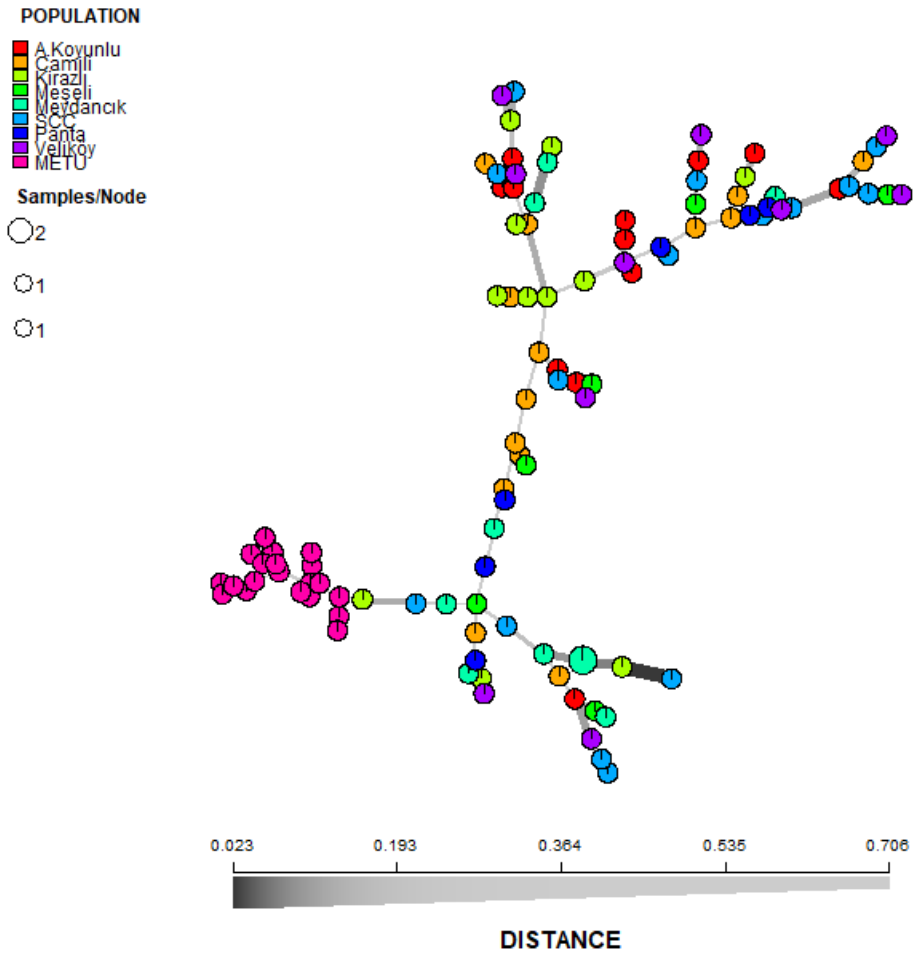


Figure 3.12: MSN map of *P. communis* and *P. eleagnifolia* individuals.

Lastly, MSN assay was carried out by using all studied genotypes. *P. elegnifolia* genotypes were closely grouped at the bottom-left of the MSN map (Figure 3.12). Besides, three main groups were observed: METU (*P. elegnifolia*), Meydancık and SCC (*P. communis*) and the rest of the populations (*P. communis*).

3.3.5. Principle Coordinate Analysis (PCoA) and Discriminat Analysis of of Principle Components (DAPC)

Principle Coordinate analysis shows that 95% of total variation is explained by the first three axes with 48%, 36%, and 11%. The first coordinate separated Kirazlı,

Meydancık, Meşeli, and Special clone collection populations from Panta, Veliköy, Camili, and A.Koyunlu populations while the second coordinate distinguished Panta, Meydancık, and Kirazlı populations from A.Koyunlu, Camili, Veliköy, Meşeli, and SCC populations. Finally, the same dataset was used to create graphical representation of discriminant analysis (Figure 3.13). Similar to the result of STRUCTURE analysis, DAPC analysis was failed to cluster individuals to distinct groups. However, as with PCoA results, the DAPC results showed that Meydancık and Panta populations were most distinct populations while the other populations were closely grouped together. (Figure 3.14)

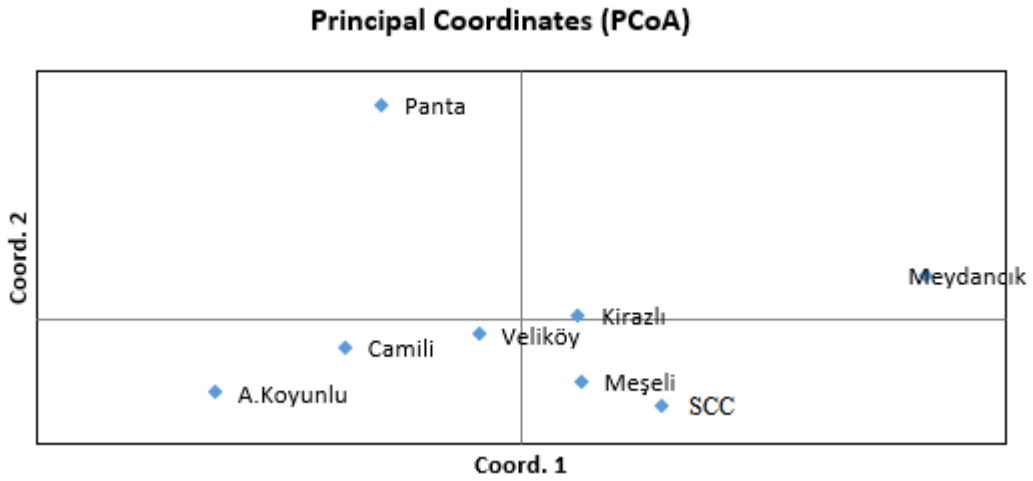


Figure 3.13: Principle Coordinate analysis based on Nei's distance of studied 8 populations

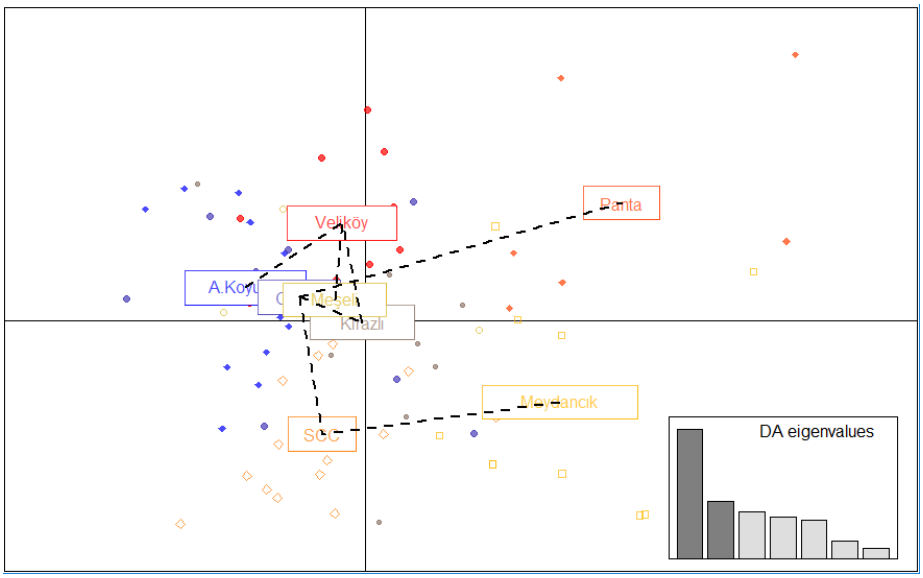


Figure 3.14: DAPC analysis of studied populations

3.4. Analyses of Molecular Variance (AMOVA)

AMOVA was carried out two times: One with two species and the other was including only *P. communis* populations. In the first analysis, great percentage of the total variation was found to be among genotypes between populations as 85.53%. 12.85% of total variation was found to be among species. Lastly, small portion of the total variation (1.61%) was found to be among population between species (Table 3.12).

Table 3.12: Analysis of molecular variance results of *P. communis* and *P. eleagnifolia* populations

Variation	Sum of Squares	Sigma	Percentage of total variation
Among Species	51.10	0.67	12.85 %
Among Populations Between Species	43.44	0.08	1.61 %
Among Genotypes Within Populations	883.90	4.49	85.53 %
Total	977.88	5.24	100 %

On the second AMOVA, putative groups of *P.communis* which were derived STRUCTURE, DAPC, and MSN analysis were tested. Most of the variation defined among genotypes within populations as 97.78%. Small portion of the variation was found to be among groups as 1.87%. Lastly, little variation was found to be among populations between groups as 0.35% (Table 3.13).

Table 3.13: AMOVA results of *P. communis* populations

Variation	**Sum of Squares**	**Sigma**	**Percentage of total variation**
Among Groups	22.50	0.08	1.87 %
Among Populations Between Groups	18.77	0.01	0.35 %
Among Genotypes Within Populations	696.0	4.35	97.78 %
Total	737.27	4.43	100 %

CHAPTER 4

DISCUSSION

4.1. SSR Markers Quality

The study was conducted by using 11 microsatellite loci markers previously developed for the genus *Pyrus* and the genus *Malus*. KU10, Bgt23b, NH013a, NB113a, TsuEnh008, NH007b, and NH008b were developed for and previously used in the genus *Pyrus* (Yamamoto *et al.,* 2002a, 2002b, 2002c; Nishitani *et al.,* 2009). CH03G06, CH02B10, CH02F06, and CH01F02 were developed for the genus *Malus* (Liedhard *et al.,* 2002; Gianfranceschi *et al.,* 1998).

During scoring microsatellite loci, null alleles, allelic drop-outs, stuttering and incorrect allele sizes due to misreading cause errors. In the studied populations, no significant null alleles, large allelic drop-outs, stuttering were not detected. Alleles with incorrect allele sizes were read again and corrected. Only in some populations, some loci showed null alleles. Generally, null alleles are resulted from mutations in the flanking region, poor quality/quantity of DNA and primers during PCR, amplification of shorter alleles, and slippage during PCR (Chapuis and Estoup, 2006). However, deficiency of heterozygosity due to inbreeding or Wahlund effect can be misinterpreted as null allele evidences (Chakraborty *et al*., 1992). In the study, no significant null allele was found.

In studied 11 SSR loci, allele size ranges highly differ between *Pyrus communis* and *Pyrus eleagnifolia* samples. Furthermore, observed allele sizes of *P. communis* samples were within the limits of previously reported studies (Bao *et al.*, 2007; Katayama *et al.,* 2007). However, in some loci, some alleles were found to be different from previous studies (Kimura *et al.,* 2002; Cao *et al.,* 2012). Those

differences could be arisen due to the geographical distribution of the species. Since there are environmental differences between studied and previously reported populations, environmental factors could alter the allele sizes. Besides, due to economic importance of pear species, anthropogenic factors are also important for this species' genotypic variations; different sets of genotypes could be used and propagated by humans around the world.

Differentiation among populations due to isolation, genetic drift, asexual reproduction, and linkage between alleles cause linkage disequilibrium (Agapow and Burt, 2001). In the conducted study, no significant linkage disequilibrium was found among loci. According to previous studies, NB113a, NH013a, NH007b, NH008b, and CH01F02 were found on different chromosomes of 3, 8, 1, 10, and 11, respectively in an interspecific hybrid of *P. communis* and *P. pyrifolia* Nakai (Yamamoto *et al.,* 2002c). In another study conducted on the same species, NB113a, NH013a, NH008b, CH03G06, NH007b, and CH01F02 are found of different chromosomes of 3, 1, 17, 14, 16, and 10 (Yamamoto *et al*, 2004). Bgt23b and CH02B10 are found on same chromosome in both studies. TsuEnh008b is found on 13^{th} chromosome of apple and 9^{th} chromosome of pear in another study (Celton *et al.*, 2009; Chen *et al.*, 2015). Similar to the literature data, no significant linkage between loci was found.

4.2. Genetic Diversity of Loci and Populations

4.2.1. Genetic Diversity of Loci

He, Ar, and PIC are important parameters for detection of a primer's suitability to use in genetic diversity analysis. In studied SSR loci, all loci had high Ar (>0.3). Besides, polymorphic information content was also higher than 0.5 in all loci that make them be useful for genetic diversity analysis. Therefore, it is suggested that all markers studied can be used effectively in further genetic diversity analysis in *P.communis.*

Some loci showed slightly higher/lower values of diversity parameters than previous studies (Yamamoto *et al.,* 2001; Bao *et al.,* 2007; Brini *et al.,* 2008). Those differences in the values of expected heterozygosity, allelic richness, and polymorphic information content can be explained by geographical isolation and human activities. Since those studies were performed in distinct areas around the world, the evolutionary histories of the samples are distinct as well. Therefore, different alleles can be found in distinct populations. Besides, since pear has agricultural value, human activities play significant role in its dispersal.

While 7 of 11 loci had a positive fixation index, 4 loci showed excess of heterozygosity. One of those four, CH02B10, was significantly deviated from HWE since some excess of heterozygotes can affect Hardy-Weinberg equilibrium. Excess of heterozygosity in those four loci can be resulted from disassortative mating in pear cultivars. For genus *Pyrus,* it was shown that there is self-incompatibility as a type of disassortative selection (Kim *et al.,* 2002; 2006). As for the other seven loci with positive fixation indices, one of those, TsuEnh008, was also deviated from HWE. This deviation can be a result of inbreeding. However, small population size, artificial selection, and/or mutation can be also the reason of this deviation.

Locus wise F-statistics analysis is an important assay to understand population differentiation and inbreeding among populations. Loci showed negative fixation indices (Fis): NB113a, CH03G06, CH02B10, and CH01F02. Besides, NH013a, NH008b, NH007b, TsuEnh008, Bgt23b, and CH02F06 had relatively higher Fst values with lower Nm values. Therefore, it can be said that differentiation of the population was mostly determined by those loci.

4.2.2. Genetic Diversity of Populations

Genetic diversity analysis was performed to the studied populations to reveal the conversational and evolutional properties of the populations. Polymorphism, private alleles, expected and observed heterozygosity (thus, fixation index) were calculated. Percentage of polymorphic loci was calculated as 100% for all studied populations which states that all populations were genetically diverse in terms of studied loci.

Private alleles count is another genetic diversity parameters for populations. Private alleles show uniqueness of the samples. According to Slatkin (Slatkin, 1985), private alleles are also indicators of migration. Logarithm of the number of private alleles and logarithm of number of migrants are linearly related to each other. Besides, private alleles are also important for the future of the population since they define an individual as unique. In the studied populations, number of private alleles ranged between one to four. Camili and Panta populations had relatively higher private alleles (4 and 3, respectively) than the other populations. Therefore, those populations are more important than the others in terms of conservation. Since Camili is the remotest village among studied villages, it is expected to observe more private alleles in this population. Besides, since Panta population is composed of wild cultivars of *Pyrus communis*, it is not surprising that it has different alleles than other populations. While commercially valuable cultivars are spread between villages, individuals in Panta population are not cultivated and spread between the villages. Therefore, it has more distinct individuals and private alleles.

Observed heterozygosity values of the populations were estimated between 0.64 and 0.77 (Panta and Veliköy, respectively) with a mean of 0.76. Interestingly, observed heterozygosity of the wild population named Panta is lower than all of the other populations. It can be explained as hybridization must have occurred during domestication of the cultivated *P. communis* individuals since hybridization is common in Pear species (Volk *et al.,* 2006). Since wild *Pyrus communis* individuals

are grown naturally and others are spread village to village, it is expected that others have hybridized more to get more yield. However, since Panta population has few number of individuals, this outcome may be wrong due to sampling errors. This putative hybridization can be tested by analyzing more individuals in Panta population and some other possible hybrid sources in the region.

4.3. Genetic Differentiation and the Structure of the Populations

Pairwise Fst values of the populations were ranged between 0.023 and 0.068. Accordingly, Nm values between population were varied between 10.64 and 3.42. Most of the Fst values among populations were found to be small which states that differentiation between those populations is not much. Since there is natural predisposition to hybridization in genus *Pyrus* and the human factors (grafting, artificial hybridization), these populations are expected to be genetically close with each other (Silva *et al.*, 2014). However, Fst values between Panta and any other populations were found significant (>0,05). Since Panta population is the wild cultivars of *P. communis*, differentiation between Panta and other populations is not confounding. While wild cultivars are exposed to mostly natural selection, others are exposed to artificial selection. This artificial selection over generations increased the genetic distance between wild and domestic cultivars. Moreover, these results are consisted with other studies conducted on wild and domesticated cultivars of genus *Pyrus* (Iketani *et al.*, 2010; Volk *et al.*, 2010; Cao *et al.*, 2012).

As for Unweighted Pair-Group Method with Arithmetic Mean (UPGMA) phenogram, two phenograms were created in the study. One for *P. communis* populations studied in Artvin region and one included *P. eleagnifoila* population collected in METU. As pairwise Fst values, differentiation of Panta population from other populations was clearly visualized in the first phenogram. Beside Panta and Meydancık populations, all of the other populations are clustered together. Since Meydancık village is located on different distributary of Çoruh river than other

villages, it was an expected result. When *P. eleagnifolia* population was added into the analysis, expectedly, METU population separated from all other populations mostly due to the comparison of two different species. Besides, geographical isolation between those populations prevented hybridization.

The STRUCTURE analysis using only *P. communis* genotypes revealed the presence of 2 gene pools. Even though delta K was calculated as 2, these two clusters failed to group individuals into two accordingly their population information. In the previous assays, differentiation between populations is found possibly due to hybridization and human factors such as grafting and vegetative growth of samples in different populations. However, STRUCTURE analysis showed that due to mostly hybridization between samples and populations, each individual in each population felled into different clusters. Only the members of wild population, Panta, were included into cluster 1 with a 100% membership value. Therefore, it can be said that the members of cluster 1 is closer to their wild relatives. Furthermore, minimum spanning network (MSN) analysis was confirmed this result visually. According to MSN results, most of the individuals are found to be more related to the individuals from different populations rather than the individuals in their own populations. These results confirm that there are significant gene flow between populations. Therefore, populations cannot be differentiated. However, the second MSN analysis showed that individuals which were classified and grouped by local farmers based on their phenotypic variations, were found to be closely associated with each other. Most of the clusters were grouped and differentiated from other clusters. However, some pre-defined clusters were failed to re-cluster in the MSN analysis. Even if the results increased the resolution of the clustering, due to gene flow between populations/clusters, there were no clear differentiation. Lastly, MSN analysis showed that there are three main gene pool when *P. eleagnifolia* population is included. One is *P. eleagnifolia* population, the second one is Meydancık and the SCC, and the other pool includes the rest of the populations.

When it comes to the *P. eleagnifolia* population, expectedly, this population was separated with all other *P. communis* populations. Even though populations are composed of two different species, interspecific hybrids are common in genus *Pyrus.* (Van der Zwet *et al.,* 1974; Montanari *et al.,* 2013; Rubstov, 1944) Therefore, geographical isolation of these population is another reason for this differentiation.

Same with UPGMA analysis, Principle Coordinate Analysis (PCoA) and Discriminant Analysis of Principle Coordinates (DAPC) analyses showed the distribution of the populations. Even if STRUCTURE analysis showed there is not a population wise distribution of the samples, it can be seen that the wild population Panta and Meydancık population were discriminated from other populations. This result confirms the previous result of that Panta is the most diverse population from the others.

Lastly, to reveal the source of differentiations, Analysis of Molecular Variance (AMOVA) assay was performed. In the first AMOVA assay, populations from two species were used. 85.53% of all variance were found to be among genotypes within populations. It is not surprising to attain this result considering the previous assays' results. Besides, 12.85% of total variation was found to be among species. Lastly, a small portion of the variation, 1.61%, was found to be among populations between species. On the second AMOVA assay, 97.78% of the total variation was found to be among genotypes within population. Besides 1.87% and 0.35% of the total variation were found to be among groups and among populations between groups, respectively. These results supported previous assays' results that there is gene flow between populations.

It is important to state that all of the assays indicate gene flow between populations, possibly as hybridization. It is known that hybridization between individuals is common in pear species (Culley and Hardiman, 2009; Hardiman and Culley, 2010;

Zheng *et al.*, 2014). Therefore, geographically close populations as studied ones show high hybridization. Besides, vegetative growth such as grafting is common in agricultural activities for pear species. Thus, grafting an individual outside of its original population facilitates gene flow between populations. Since Şavşat is the most populated center of the studied region, human factors on distribution of commercially valuable plants are important. Gene flow between individuals and populations, therefore, is an expected result of human activity on agriculture and trade. However, to assess the genetic source of those hybridizations, it is important to extend the study area. Since pear species can be hybridized inter- and intraspecifically, during extending study area, not only cultivars of species of genus *Pyrus* but also genus *Malus* should be collected and analyzed.

CHAPTER 5

CONCLUSION

Species of genus *Pyrus* are one of the most important tree fruit group in terms of their economic value and consumption rate. Overall, pears are the second most consumed and produced tree fruit after apples. As for Turkey, similarly, pear is the second most produced pome fruit after apple. Therefore, due to its economic value, genetic structure analysis and conservation of genetic sources of pears are important.

In this study, 8 population of *P. communis* samples from Artvin Province (84 samples in total) and 1 population of *P. eleagnifolia* samples from METU campus (Ankara Province, 20 samples in total) were analyzed with 11 SSR markers. No null allele was found in the loci studied. Only some loci showed null alleles in some specific populations. Linkage disequilibrium analysis showed that there is not significant linkage between loci. Besides, out of 84 samples, 83 samples were found to be multilocus genotypes. According to locus-wise parameters as Allelic Richness (Ar), Polymorphic Information Content (PIC), Expected Heterozygosity (He), and Shannon's index of Identity (I), all markers used in the study (NB113a, NH013a, KU10, NH008b, CH03G06, NH007b, TsuEnh008, Bgt23b, CH02B10, CH01F02, and CH02F06) were found to be informative and can be used further analyses in *P. communis.*

Population differentiation analyses (pairwise Fst, DAPC, PCoA, MSN, and UPGMA) reveal that the wild population Panta is significantly differentiated from other populations. However, population structure analyses (STRUCTURE and MSN) and AMOVA showed that all the individuals in the populations highly admixed with each other. Even though MSN and STRUCTURE suggest that there are two main clusters, those clusters are not related with the original source of population. Therefore, these

results show that individuals in the populations are hybrids. Since the villages are too close to each other, especially the ones other than Meydancık, hybridization between individuals is an outcome of human activities.

The study reveal that the cultivars of *P. communis* are highly hybridized. Even though sampling area is narrow, and number of samples are few, it was seen that domesticated cultivars are highly heterozygous than the wild population. Therefore, it is important for the further studies to extent the sampling area and to increase number of wild populations to understand the domestication process of Anatolian pear. Besides, by doing so, the original source of this hybrids can be found and conserved.

REFERENCES

Agarwal, M., Shrivastava, N., & Padh, H. (2008). Advances in molecular marker techniques and their applications in plant sciences. Plant cell reports, 27(4), 617-631.

Agapow, P. M., & Burt, A. (2001). Indices of multilocus linkage disequilibrium. Molecular Ecology Notes, 1(1-2), 101-102.

Akçay, M. E., Burak, M., Kazan, K., Yüksel, C., Mutaf, F., Bakir, M., ... & Ergül, A. (2014). Genetic analysis of Anatolian pear germplasm by simple sequence repeats. Annals of applied biology, 164(3), 441-452.

Altınbay, Hatice (2012), "Doğu Anadolu'da Yetişen Bazı Armut Gen Kaynaklarının SSR's (Simple Sequence Repeats) Markörlere Dayalı Genetik Karakterizasyonu", MS thesis, Ankara University, Institute of Biotechnology

Arnaud-Haond S, Belkhir K. (2007), Genclone: a computer program to analyse genotypic data, test for clonality and describe spatial clonal organization. Molecular Ecology Notes, 7:15–17.

Bao, L., Chen, K., Zhang, D., Cao, Y., Yamamoto, T., & Teng, Y. (2007). Genetic diversity and similarity of pear (Pyrus L.) cultivars native to East Asia revealed by SSR (simple sequence repeat) markers. Genetic Resources and Crop Evolution, 54(5), 959.

Bassil, N. V., Postman, J. D., & Neou, C. (2004, February). Pyrus microsatellite markers from GenBank sequences. In IX International Pear Symposium 671 (pp. 289-292).

Bassil, N., & Postman, J. D. (2010). Identification of European and Asian pears using EST-SSRs from Pyrus. Genetic resources and crop evolution, 57(3), 357-370.

Bell RL (1990) Pears (Pyrus). In: Moore JN, Ballington JR Jr (eds) Genetic resources of temperate fruit and nut crops I. International Society for Horticultural Science, Wageningen, The Netherlands, pp 655–697

Bell, R.L., H.A. Quamme, R.E.C. Layne & R.M. Skirvin, 1996. Pears. In: Janick, J. & J.N. Moore, (Eds), Fruit Breeding, Vol I: Tree and Tropical Fruits, pp. 441–514. John Wiley & Sons, Inc., New York, NY

Brini, W., Mars, M., & Hormaza, J. I. (2008). Genetic diversity in local Tunisian pears (*Pyrus communis L*.) studied with SSR markers. Scientia horticulturae, 115(4), 337-341.

Brookfield JFY. (1996), A simple new method for estimating null allele frequency from heterozygote deficiency. Molecular Ecology, 5:453–455

Browicz, K. (1993). Conspect and chorology of the genus Pyrus L. Arboretum Kórnickie, 38, 17-33.

Cao, Y., Tian, L., Gao, Y., & Liu, F. (2012). Genetic diversity of cultivated and wild Ussurian Pear (*Pyrus ussuriensis Maxim.*) in China evaluated with M13-tailed SSR markers. *Genetic resources and crop evolution*, *59*(1), 9-17.

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

Chakraborty R, De Andrade M, Daiger SP, Budowle B (1992). Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. Ann Hum Genet 56: 45–57.

Chapuis, M. P., & Estoup, A. (2006). Microsatellite null alleles and estimation of population differentiation. Molecular biology and evolution, 24(3), 621-631.

Chen, H., Song, Y., Li, L. T., Khan, M. A., Li, X. G., Korban, S. S., ... & Zhang, S. L. (2015). Construction of a high-density simple sequence repeat consensus genetic map for pear (*Pyrus spp.*). Plant Molecular Biology Reporter, 33(2), 316-325.

Christenhusz, M. J., & Byng, J. W. (2016). The number of known plants species in the world and its annual increase. Phytotaxa, 261(3), 201-217.

Celton, J. M., Chagné, D., Tustin, S. D., Terakami, S., Nishitani, C., Yamamoto, T., & Gardiner, S. E. (2009). Update on comparative genome mapping between Malus and Pyrus. BMC research notes, 2(1), 182.

Culley, T. M., & Hardiman, N. A. (2009). The role of intraspecific hybridization in the evolution of invasiveness: a case study of the ornamental pear tree *Pyrus calleryana*. Biological Invasions, 11(5), 1107-1119.

Doyle, J.J. and Doyle, J.L. (1987), A rapid DNA isolation procedure for small quantities of fresh leaf tissue, Phytochem. Bull., 19: 11-15.

Earl Dent A, Vonholdt Bridgett M (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conser Genet Resour 4(2): 359–361

Excoffier, L., Lischer, H.E.L. (2010), Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 10:564–567, doi: 10.1111/j.1755-0998.2010.02847.x.

FAOSTAT (2010), FAO statistics database on the World Wide Web. http://www.fao.org

FERNÁNDEZ-FERNÁNDEZ, F., Harvey, N. G., & James, C. M. (2006). Isolation and characterization of polymorphic microsatellite markers from European pear (*Pyrus communis L.*). Molecular Ecology Notes, 6(4), 1039-1041.

Francis RM (2016) POPHELPER: an R package and web app to analyse and visualise population structure. Mol Ecol ResourI. doi:10.1111/ 1755-0998.12509

Garza J, Williamson E. (2001), Detection of reduction in population size using data from microsatellite loci. Mol Ecol, 10:305–318. doi: 10.1046/j.1365-294x.2001.01190.x.

Gemayel, R., M. D. Vinces, M. Legendre, and K. J. Verstrepen (2010), Variable tandem repeats accelerate evolution of coding and regulatory sequences. Annu. Rev. Genet. 44: 445–477.

Gemayel, R., J. Cho, S. Boeynaems, and K. J. Verstrepen (2012), Beyond junk-variable tandem repeats as facilitators of rapid evolution of regulatory and coding sequences. Genes 3: 461–480

Gianfranceschi, L., Seglias, N., Tarchini, R., Komjanc, M., & Gessler, C. (1998). Simple sequence repeats for the genetic analysis of apple. Theoretical and Applied Genetics, 96(8), 1069-1076.

Greenbaum, G., Templeton, A. R., Zarmi, Y., & Bar-David, S. (2014). Allelic richness following population founding events–a stochastic modeling framework incorporating gene flow and genetic drift. PloS one, 9(12), e115203.

Grünwald NJ., Goss EM. 2011. Evolution and population genetics of exotic and re-emerging pathogens: Novel tools and approaches. Annual Review of Phytopathology 49:249–267

Guo, X., & Elston, R. (1999). Linkage information content of polymorphic genetic markers. Human Heredity, 49(2), 112-118.

Hardiman, N. A., & Culley, T. M. (2010). Reproductive success of cultivated *Pyrus calleryana* (Rosaceae) and establishment ability of invasive, hybrid progeny. American Journal of Botany, 97(10), 1698-1706.

Hartl, D. L., Clark, A. G., & Clark, A. G. (1997). Principles of population genetics (Vol. 116). Sunderland: Sinauer associates.

Hummer, K. E., & Janick, J. (2009). Rosaceae: taxonomy, economic importance, genomics. In Genetics and genomics of Rosaceae (pp. 1-17). Springer, New York, NY.

Iketani, H., Yamamoto, T., Katayama, H., Uematsu, C., Mase, N., & Sato, Y. (2010). Introgression between native and prehistorically naturalized (archaeophytic) wild pear (*Pyrus spp.*) populations in Northern Tohoku, Northeast Japan. Conservation genetics, 11(1), 115-126.

Inoue, E., Matsuki, Y., Anzai, H., & Evans, K. (2007). Isolation and characterization of microsatellite markers in Japanese pear (*Pyrus pyrifolia Nakai*). Molecular Ecology Notes, 7(3), 445-447.

Jakobsson M, Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. Bioinformatics 23:1801–1806

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics, 24(11), 1403-1405.

Jombart T., Devillard S., Balloux F. 2010. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. BMC genetics 11:94

Kamvar, Z. N., Brooks, J. C., & Grünwald, N. J. (2015). Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. Frontiers in Genetics, 6, 208.

Katayama, H., Adachi, S., Yamamoto, T., & Uematsu, C. (2007). A wide range of genetic diversity in pear (*Pyrus ussuriensis var. aromatica*) genetic resources from Iwate, Japan revealed by SSR and chloroplast DNA markers. Genetic resources and crop evolution, 54(7), 1573.

Keller, L. F., & Waller, D. M. (2002). Inbreeding effects in wild populations. Trends in ecology & evolution, 17(5), 230-241.

Kim, H. T., Hirata, Y., & Nou, I. S. (2002). Determination of S-genotypes of pear (*Pyrus pyrifolia*) cultivars by S-RNase sequencing and PCR-RFLP analyses. Molecules and cells, 13(3), 444-451.

Kim, H. T., Hirata, Y., Kim, H. J., & Nou, I. S. (2006). The presence of a new S-RNase allele (S 10) in Asian pear (*Pyrus pyrifolia (Burm; Nakai)*). Genetic resources and crop evolution, 53(7), 1375.

Kimura, T., Shi, Y. Z., Shoda, M., Kotobuki, K., Matsuta, N., Hayashi, T., ... & Yamamoto, T. (2002). Identification of Asian pear varieties by SSR analysis. Breeding Science, 52(2), 115-121.

Kutzelnigg, H., & Silbereisen, R. (1995). Pyrus. Illustrierte Flora von Mitteleuropa, 4(2B), 278-298.

Lewis, P. O., and Zaykin, D. (2001), Genetic Data Analysis: Computer program for the analysis of allelic data. (Available at http://en.bio-soft.net/dna/gda.html Retrieved: 07.17)

Li, X., Ding, X., Chu, B., Zhou, Q., Ding, G., & Gu, S. (2008). Genetic diversity analysis and conservation of the endangered Chinese endemic herb Dendrobium officinale Kimura et Migo (Orchidaceae) based on AFLP. Genetica, 133(2), 159-166.

Liebhard, R., Gianfranceschi, L., Koller, B., Ryder, C. D., Tarchini, R., Van de Weg, E., & Gessler, C. (2002). Development and characterisation of 140 new microsatellites in apple (Malus x domestica Borkh.). Molecular Breeding, 10(4), 217-241.

Liu, Q., Song, Y., Liu, L., Zhang, M., Sun, J., Zhang, S., & Wu, J. (2015). Genetic diversity and population structure of pear (*Pyrus spp.*) collections revealed by a set of core genome-wide SSR markers. *Tree genetics & genomes*, *11*(6), 128.

Mayer, C., Leese, F., & Tollrian, R. (2010). Genome-wide analysis of tandem repeats in Daphnia pulex-a comparative approach. BMC genomics, 11(1), 277.

Montanari, S., Saeed, M., Knäbel, M., Kim, Y., Troggio, M., Malnoy, M., ... & Perchepied, L. (2013). Identification of Pyrus single nucleotide polymorphisms (SNPs) and evaluation for genetic mapping in European pear and interspecific Pyrus hybrids. PLoS One, 8(10), e77022.

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

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

Nei, M., & Chesser, R. K. (1983). Estimation of fixation indices and gene diversities. Annals of human genetics, 47(3), 253-259.

Nishitani, C., Terakami, S., Sawamura, Y., Takada, N., & Yamamoto, T. (2009). Development of novel EST-SSR markers derived from Japanese pear (*Pyrus pyrifolia*). Breeding Science, 59(4), 391-400.

Okasha, Samir, "Population Genetics", The Stanford Encyclopedia of Philosophy (Winter 2016 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/win2016/entries/population-genetics/>.

Paetkau, D., Calvert, W., Stirling, I., & Strobeck, C. (1995). Microsatellite analysis of population structure in Canadian polar bears. Molecular ecology, 4(3), 347-354.

Peakall, R. and Smouse P.E. (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28, 2537-2539.

Petit, R. J., A. El Mousadik, and O. Pons. 1998. Identifying populations for conservation on the basis of genetic markers. Conservation Biology 12:844-855.

Potter, D., Eriksson, T., Evans, R. C., Oh, S., Smedmark, J. E. E., Morgan, D. R., ... & Campbell, C. S. (2007). Phylogeny and classification of Rosaceae. Plant systematics and evolution, 266(1-2), 5-43.

Pritchard, J.K., Stephens, M., Donnelly, P. (2000). Inference of population structure using multilocus genotype data. Genetics 155(2):945-59.

Rana, J. C., Chahota, R. K., Sharma, V., Rana, M., Verma, N., Verma, B., & Sharma, T. R. (2015). Genetic diversity and structure of Pyrus accessions of Indian Himalayan region based on morphological and SSR markers. Tree genetics & genomes, 11(1), 821.

Raymond, M., Rousset, F. (1995), GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J. Heredity, 86:248-249.

Rousset, F. (2008), Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Mol. Ecol. Resources 8: 103-106.

Rubtsov, G. A. (1944). Geographical distribution of the genus Pyrus and trends and factors in its evolution. The American Naturalist, 78(777), 358-366.

Sax, K. (1931). The origin and relationships of the Pomoideae. Journal of the Arnold arboretum, 12(1), 3-22.

Schlötterer, C. (2004). The evolution of molecular markers—just a matter of fashion?. Nature reviews genetics, 5(1), 63.

Sehic, J., Garkava-Gustavsson, L., Fernández-Fernández, F., & Nybom, H. (2012). Genetic diversity in a collection of European pear (*Pyrus communis*) cultivars determined with SSR markers chosen by ECPGR. Scientia horticulturae, 145, 39-45.

Silva, G. J., Souza, T. M., Barbieri, R. L., & Costa de Oliveira, A. (2014). Origin, domestication, and dispersing of pear (*Pyrus spp.*). Advances in Agriculture, 2014.

Slatkin, M. (1985). Rare alleles as indicators of gene flow. Evolution, 39(1), 53-65.

Terpó, A. (1985). Studies on taxonomy jand grouping of *Pyrus* species. Feddes Repertorium.

Thomas, E. E. (2005). Short, local duplications in eukaryotic genomes. Current opinion in genetics & development, 15(6), 640-644.

Türkiye istatistik kurumu (TUIK), Haber Bülteni, Sayı: 24581, 29.09.2017, http://www.tuik.gov.tr/PreHaberBultenleri.do?id=24581

Van der Zwet, T., Oitto, W. A., & Westwood, M. N. (1974). Variability in degree of fire blight resistance within and between *Pyrus* species, interspecific hybrids, and seedling progenies. Euphytica, 23(2), 295-304.

van Oosterhout C., Hutchinson W. F., Wills D. P. M., Shipley P. 2004. MICRO-CHECKER: Software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4: 535–538

Vavilov NI (1951) The origin, variation, immunity and breeding of cultivated plants. Chronica Botanica, 13:1–366 (translated by K. Starr)

Verstrepen, K. J., A. Jansen, F. Lewitter, and G. R. Fink, 2005 Intragenic tandem repeats generate functional variability. Nat. Genet. 37: 986–990.

Volk, G. M., Richards, C. M., Henk, A. D., Reilley, A. A., Bassil, N. V., & Postman, J. D. (2006). Diversity of wild *Pyrus communis* based on microsatellite analyses. Journal of the American Society for Horticultural Science, 131(3), 408-417.

Weeden, N. F., & Lamb, R. C. (1987). Genetics and linkage analysis of 19 isozyme loci in apple. Journal of the American Society for Horticultural Science (USA).

Wright, S. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution, 19(3), 395-420.

Yamamoto, T., Kimura, T., Sawamura, Y., Kotobuki, K., Ban, Y., Hayashi, T., & Matsuta, N. (2001). SSRs isolated from apple can identify polymorphism and genetic diversity in pear. Theoretical and Applied Genetics, 102(6-7), 865-870.

Yamamoto, T., Kimura, T., Sawamura, Y., Manabe, T., Kotobuki, K., Hayashi, T., ... & Matsuta, N. (2002a). Simple sequence repeats for genetic analysis in pear. Euphytica, 124(1), 129.

Yamamoto, T., Kimura, T., Shoda, M., Ban, Y., Hayashi, T., & Matsuta, N. (2002b). Development of microsatellite markers in the Japanese pear (*Pyrus pyrifolia Nakai*). Molecular Ecology Resources, 2(1), 14-16.

Yamamoto, T., Kimura, T., Shoda, M., Imai, T., Saito, T., Sawamura, Y., ... & Matsuta, N. (2002c). Genetic linkage maps constructed by using an interspecific cross between Japanese and European pears. Theoretical and Applied Genetics, 106(1), 9-18.

Yamamoto, T., Kimura, T., Saito, T., Kotobuki, K., Matsuta, N., Liebhard, R., ... & Hayashi, T. (2004). Genetic linkage maps of Japanese and European pears aligned to the apple consensus map. Acta Horticulturae, 663, 51-56.

Yamamoto, T., Kimura, T., Terakami, S., Nishitani, C., Sawamura, Y., Saito, T., ... & Hayashi, T. (2007). Integrated reference genetic linkage maps of pear based on SSR and AFLP markers. Breeding Science, 57(4), 321-329.

Yao, L., Zheng, X., Cai, D., Gao, Y., Wang, K., Cao, Y., & Teng, Y. (2010). Exploitation of Malus EST-SSRs and the utility in evaluation of genetic diversity in Malus and Pyrus. Genetic resources and crop evolution, 57(6), 841-851.

YMS değerlendirme raporu Ocak – Aralık 2017, 12.2017, Türkiye Yaş Meyve Sebze İhracatçı Birliği, http://www.yms.org.tr/tr/istatistikler-degerlendirme-raporlari.html

Yue, X. Y., Liu, G. Q., Zong, Y., Teng, Y. W., & Cai, D. Y. (2014). Development of genic SSR markers from transcriptome sequencing of pear buds. Journal of Zhejiang University Science B, 15(4), 303-312.

Zhang, R. P., Wu, J., Li, X. G., Khan, M. A., Chen, H., Korban, S. S., & Zhang, S. L. (2013). An AFLP, SRAP, and SSR genetic linkage map and identification of QTLs for fruit traits in pear (*Pyrus L.*). Plant Molecular Biology Reporter, 31(3), 678-687.

Zheng, X., Cai, D., Potter, D., Postman, J., Liu, J., & Teng, Y. (2014). Phylogeny and evolutionary histories of *Pyrus L.* revealed by phylogenetic trees and networks based on data from multiple DNA sequences. Molecular phylogenetics and evolution, 80, 54-65.

Zohary, D. (1997). Wild apples and wild pears. In Proceedings of the workshop on conservation of the wild relatives of European cultivated plants. Bocconea (Palermo) (Vol. 7, pp. 409-416).

APPENDICES

APPENDIX A

Sampled *P.communis* genotypes

	Kuma	Meydancık
	Kuma	Meydancık
	Kuma	Meydancık
	anta(Yaban-Aşısız)	Meydancık
	Kabağa Armudu	Meydancık
	anta(Yaban-Aşısız)	Meydancık
	örstella	Meydancık
	anta(Yaban-Aşısız)	Meydancık
	Kansilya	Meydancık
,1	Aşı	Meydancık
0	anta(Yaban-Aşısız)	Meydancık
1	Kuma	Meydancık
2	Kasitava	Meydancık
3	Tevracul	Meydancık
00	sozola	Yeliköy
01	Acara	Yeliköy
02	Sozdoğan	Yeliköy
03	Usket	Yeliköy
04	Harğ Armudu	Yeliköy
05	Kış Armudu	Yeliköy
06	ahara	Yeliköy
07	Huleşiya	Yeliköy
08	simsiz-	Meşeli
09	anaşor - Harman	A.Koyunlu
10	Yağ Armudu	Meşeli
11	sozola	Meşeli
12	Yağ Armudu	Meşeli
13	simsiz-	Meşeli
14	anta(Yaban-Aşısız)	Meşeli

15	evracul	eliköy
16	aranşor	.Koyunlu
17	lyanak	.Koyunlu
18	abağa Armudu	.Koyunlu
19	rgevul	.Koyunlu
20	sma Armudu	.Koyunlu
21	evracul	.Koyunlu
22	al Armudu	.Koyunlu
23	eveli Armudu	.Koyunlu
24	rgevul	.Koyunlu
25	ez erişen	eliköy
26	al Armudu	.Koyunlu
27	eşil Armut	.Koyunlu
28	asila Armudu	irazlı
29	anga	irazlı
30	anizor	irazlı
31	ar Armudu	eşeli
32	oz Armut	irazlı
33	şık Mehmet	irazlı
34	utupasila	irazlı
135	Kuriştava	Kirazlı
136	Karasesela Armudu	Kirazlı
137	Arpa Armudu	Kirazlı
138	Asma Armudu	Kocabey
201	Salum Savuray	Camili
202	Caniviray	Camili
203	Panta(Yaban-Aşısız)	Camili
204	Bulducay	Camili
205	Şekeray	Camili
206	Bağda(Büyük bağ armudu)	Camili
207	Didivanay	Camili
208	Ğomay	Camili
209	Saselay	Camili
210	Kurssukanay	Camili
211	Seselay	Camili
212	S(Z)aray	Camili

213	Haçeçuray	Camili
214	Küçük Bağ Armudu	Camili
215	Karnushapi(Samaha)	Camili
301	Cetrula	SCC
302	Şeker	SCC
303	Harg	SCC
304	Bebera	SCC
305	Yağ Armudu	SCC
306	Goaha	SCC
307	Loğ	SCC
308	Kiraz	SCC
309	Bardak	SCC
310	Acara	SCC
311	Yarar	SCC
312	Kabağa Armudu	SCC
313	Kansilya	Kirazlı
314	Nağsit	SCC
315	Işık Mehmet	SCC
316	Goş	SCC

APPENDIX B

Genomic DNA Isolation

DNA isolation was performed by using an alternate version of CTAB (cetyl trimethylammonium bromide) protocol. (Doyle and Doyle, 1987) The whole procedure is given below:

1. 0.1 gram crushed leaf tissue was put in a mortar and grounded with 1000 µl pre-heated 2X CTAB solution.
2. 800 µl of the grounded solution was poured into 1.5 ml Eppendorf tube, and 200 µl β-mercaptoethanol and 5 µl Proteinase K were added into the solution.
3. Tubes, then, were placed on water bath. In 65 ºC, they were incubated for one hour.
4. After incubation, tubes were centrifuged at 15000 rpm, 4 ºC for 20 minutes.
5. Supernatant were placed into another tube and mixed with phenol up to 80% of its volume.
6. Solutions were centrifuged at 15000 rpm, 4 ºC for 15 minutes.
7. Supernatant were placed into another tube and chloroform:isoamyl alcohol (24:1) were added into it up to 80% of its volume.
8. Tubes were centrifuged at 15000 rpm, 4 ºC for 17 minutes.
9. Supernatant were placed into another tube and equal volume of cold isopropanol added into it.
10. Tubes were placed into -20 ºC for overnight.
11. Samples were centrifuged at 13000 rpm, 4 ºC for 8 minutes.
12. The pellet was washed with 500 µl, 70% cold ethanol.
13. The pellet was let dry onto blotter for about 60 minutes.
14. The pellet was resuspended with 75 µl TE buffer.

APPENDIX C

BUFFERS, CHEMICALS AND EQUIPMENTS

Buffers and solutions for DNA isolation

2X CTAB: 2 gr CTAB (Cetyl Trimethyl Ammonium Bromide), (SIGMA)

4 ml (pH:8) 0.5 M EDTA, (FLUKA)

10 ml (pH:8) Tris HCL, (SIGMA)

28 ml NaCl is completed with 100 mL distilled water

Phenol, (AMRESCO): Pure phenol

Chloroform isoamyl alcohol, (FLUKA) : (24/1)

Ethanol: 70% in distilled water

β mercapto ethanol, (SIGMA): 17,5 ml β mercapto ethanol is completed with 250 ml with distilled water

TE buffer: 10mm Tris HCL (pH:7) 10mm ethylene diamine tetra acetic acid disodium salt (EDTA)

Isopropanol, (FLUKA): Pure Isopropanol, ice cold

Buffers and solutions for PCR

Sterile water

Taq DNA Polymerase (SIGMA Red Taq): 1U/µl

10X PCR buffer including MgCl2 (SIGMA)

dNTPs (SIGMA): 10mM

DNA: 10ng/ µl

Primer Pairs: 10µM

Reaction mixture: 5x HOT FIREPol® Blend Master Mix Ready to Load (Solis BioDyne, Tartu, Estonia)

Agarose Gel Electrophoresis Buffers and Gel System

10X TBE Buffer: 108 gr Trizma Base, (SIGMA), 55 gr Boric Acid, (SIGMA)

Running Buffers: X TBE prepared in distilled water

Ethydium Bromide (SIGMA):4 mg/ ml

Agarose, (SIGMA): 3 % Agarose Gel

40 ml EDTA, (FLUKA) (0.5 M, pH:8) completed with 1000 ml with distilled water

Low molecular weight DNA Ladder (SIGMA)

Equipments

Autoclave: Yamato

Centrifuge: Nüve- NF048

Electrophoresis System: Thermo Scientific

Thermocyclers: Eppendorf- Mastercycler

Deep-freezer: UĞUR- Freezer

Magnetic Stirrer: Labor Brand – Hotplate L-81

Refrigerator: Siemens

UV Transilluminator: Vilbor Lourmant

Vortex: Nüve- NM110

Water Bath: Memmert

Oven: Dedeoğlu

Micropipettes: Gilson

pHmeter: Hanna Inst.

APPENDIX D

Analysis which was performed by BM Labosis Company

1. PCR product registration by customer.

2. PCR product + Hi-Di formamide + size standard * size standard type : 120LIZ, 350ROX, 400HD, 500LIZ, 600LIZ, 1200LIZ

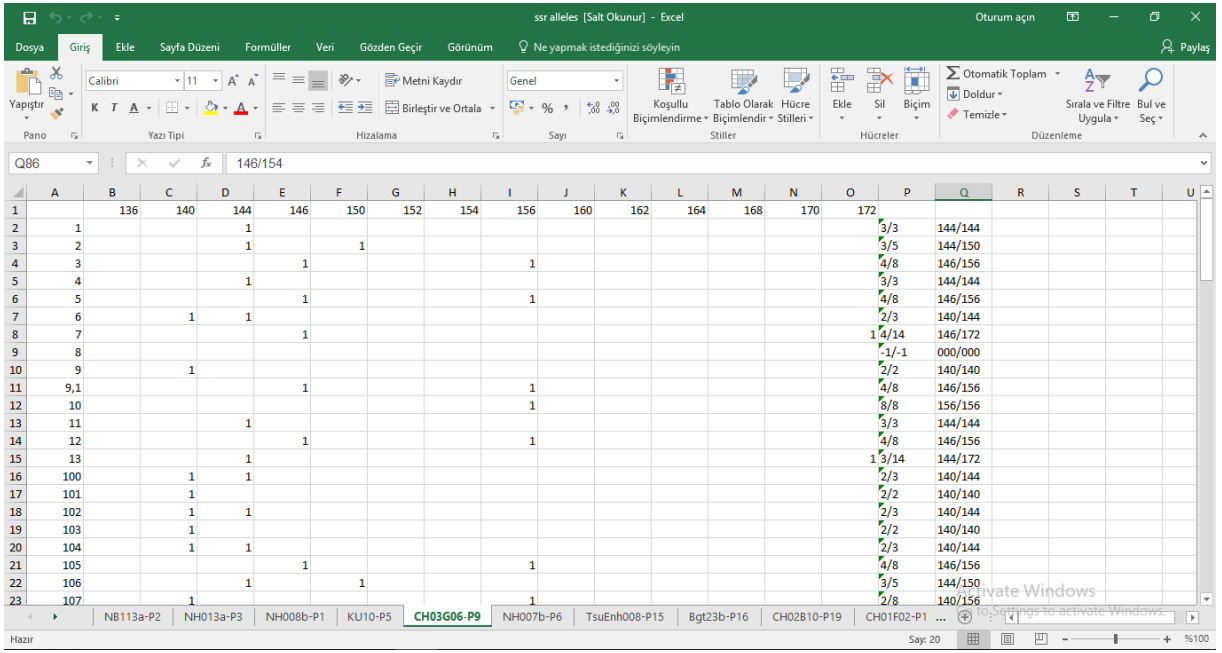
3. Denaturation

4. 3730xl running by using Dye set: DS-30 set for internal standard size marker 400HD , DS-33 set for internal standard size marker 400HD

5. Genemapper v.5 analysis

APPENDIX E

Raw Data Sample



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1		136	140	144	146	150	152	154	156	160	162	164	168	170	172		
2	1			1												3/3	144/144
3	2			1		1										3/5	144/150
4	3				1				1							4/8	146/156
5	4			1												3/3	144/144
6	5				1				1							4/8	146/156
7	6		1	1												2/3	140/144
8	7				1										1	4/14	146/172
9	8															-1/-1	000/000
10	9		1													2/2	140/140
11	9,1				1				1							4/8	146/156
12	10								1							8/8	156/156
13	11			1												3/3	144/144
14	12				1				1							4/8	146/156
15	13			1											1	3/14	144/172
16	100		1	1												2/3	140/144
17	101		1													2/2	140/140
18	102		1	1												2/3	140/144
19	103		1													2/2	140/140
20	104		1	1												2/3	140/144
21	105				1				1							4/8	146/156
22	106			1		1										3/5	144/150
23	107		1						1							2/8	140/156

APPENDIX F

Used Data Types

Arlequin:

```
[Profile]
        Title="Title line:  Pyrus.txt"

        NbSamples=7
        DataType=MICROSAT
        GenotypicData=1
        LocusSeparator=WHITESPACE
        GameticPhase=0
        RecessiveData=1
        RecessiveAllele="000"
        MissingData="?"

[Data]
        [[Samples]]
          SampleName="1_"
          SampleSize=12
          SampleData= {
1     1   138  195  230  200  140  132  134  197  128  165  154
          150  205  276  205  144  142  144  205  130  171  154
2     1   138  195  230  195  140  132  134  209  130  171  154
          150  205  260  205  144  142  138  209  130  180  154
3     1   148  197  232  191  140  134  134  211  130  171  172
          156  199  232  200  144  134  134  211  130  178  172
4     1   138  191  248  195  140  122  134  211  126  171  154
```

GDA:

```
#nexus
[!Data from Pyrus
]

begin gdadata;
       dimensions nloci=11 npops=8;
       format tokens missing = ? datapoint = standard;
              locusallelelabels
              1 'NB113a'      [/ 1 2 3 4 5 6 7 8 9 10 11],
              2 'NH013a'      [/ 1 2 3 4 5 6 7 8 9 10 11 12],
              3 'KU10'        [/ 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18],
              4 'NH008b'      [/ 1 2 3 4 5 6 7 8 9 10],
              5 'CH03G06'     [/ 1 2 3 4 5 6 7 8 9 10 11 12 13],
              6 'NH007b'      [/ 1 2 3 4 5 6 7 8 9 10],
              7 'TsuEnh008'   [/ 1 2 3 4 5 6],
              8 'Bgt23b'      [/ 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18],
              9 'CH02B10'     [/ 1 2 3 4 5 6 7],
              10 'CH01F02'    [/ 1 2 3 4 5 6 7 8 9 10 11],
              11 'CH02F06'    [/ 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15]
;

MATRIX



AKoyunlu:

109     2/7     4/7     2/17    6/7     2/3     5/10    1/6     6/9     5/6     3/5     2/2
116     2/10    4/7     2/11    4/7     2/3     5/10    1/3     11/11   6/6     5/10    2/2
117     6/10    5/6     3/3     2/6     2/3     6/6     1/1     12/12   6/6     5/9     7/7
118     2/7     2/4     7/15    4/6     2/12    2/6     1/1     12/12   4/6     5/8     2/14
119     2/6     4/7     14/14   4/7     2/2     2/6     1/1     6/11    4/6     9/10    3/8
120     6/9     6/6     5/6     2/6     4/7     6/7     2/4     12/12   4/6     1/10    8/8
121     2/7     6/11    5/8     4/6     3/10    1/10    4/6     2/16    4/6     1/5     3/8
122     6/10    6/7     5/5     2/6     2/9     2/6     1/1     2/2     5/6     3/7     2/2
123     7/7     3/12    2/3     2/5     2/3     7/7     3/5     9/18    2/6     1/8     8/13
```

Genepop:

```
Title line:  "Pyrus.txt"
NB113a
NH013a
KU10
NH008b
CH03G06
NH007b
TsuEnh008
Bgt23b
CH02B10
CH01F02
CH02F06
Pop
1_      ,       138150  195205  230276  200205  140144  132142  134144  197205  128130  165171  154154
1_      ,       138150  195205  230260  195205  140144  132142  134138  209209  130130  171180  154154
1_      ,       148156  197199  232232  191200  140144  134134  134134  211211  130130  171178  172172
1_      ,       138150  191195  248270  195200  140170  122134  134134  211211  126130  171176  154194
1_      ,       138148  195205  268268  195205  140140  122134  134134  197209  126130  178180  156174
1_      ,       148154  199199  236240  191200  146154  134136  136140  211211  126130  161180  174174
1_      ,       138150  199219  236250  195200  144162  120128  140144  189223  126130  161171  156174
1_      ,       148156  199205  236236  191200  140160  122134  134134  189189  128130  165174  154154
1_      ,       150150  193225  230232  191197  140144  136136  138142  205229  122130  161176  174192
1_      ,       138148  195205  230268  195205  140140  122134  134138  211211  126130  178180  156174
1_      ,       138152  185205  232252  205209  140140  132142  134138  205205  128130  165180  154176
1_      ,       148154  193193  232236  193205  140146  134140  136136  211211  126130  165171  168174
Pop
2_      ,       146150  195195  232274  197197  140160  122134  134134  197209  124130  165178  154174
2_      ,       148154  193205  232248  193205  144150  132132  134134  203209  126130  171176  174194
2_      ,       138156  191197  236258  191191  140160  132132  138138  209209  126130  176178  154196
2_      ,       142150  197197  260266  191211  140150  122134  136136  207207  126130  165178  154174
2_      ,       150150  197197  234234  195209  146156  136136  134138  197207  130130  163165  174190
2_      ,       138142  191195  248270  191195  140172  122134  134134  211211  130132  171176  154194
2_      ,       138148  191211  268268  191200  140140  132142  136136  205205  130130  171182  154194
2_      ,       150158  193211  232260  191191  140144  132140  136136  207209  128130  165171  154174
2_      ,       138150  195211  230232  200209  140144  132140  134134  207209  128130  171178  154154
2_      ,       142150  185199  230260  191200  140160  122134  136136  205211  128130  165171  154174
2_      ,       138152  195199  268268  195200  140144  120132  136136  207209  128130  163178  154154
```

Structure:

		NB113a	NH013a	KU10	NH008b	CH03G06	NH007b	TsuEnh008		Bgt23b	CH02B10	CH01F02	CH02F06
1	1	138	195	230	200	140	132	134	197	128	165	154	
1	1	150	205	276	205	144	142	144	205	130	171	154	
2	1	138	195	230	195	140	132	134	209	130	171	154	
2	1	150	205	260	205	144	142	138	209	130	180	154	
3	1	148	197	232	191	140	134	134	211	130	171	172	
3	1	156	199	232	200	144	134	134	211	130	178	172	
4	1	138	191	248	195	140	122	134	211	126	171	154	
4	1	150	195	270	200	170	134	134	211	130	176	194	
5	1	138	195	268	195	140	122	134	197	126	178	156	
5	1	148	205	268	205	140	134	134	209	130	180	174	
6	1	148	199	236	191	146	134	136	211	126	161	174	
6	1	154	199	240	200	154	136	140	211	130	180	174	
7	1	138	199	236	195	144	120	140	189	126	161	156	
7	1	150	219	250	200	162	128	144	223	130	171	174	
8	1	148	199	236	191	140	122	134	189	128	165	154	
8	1	156	205	236	200	160	134	134	189	130	174	154	
9	1	150	193	230	191	140	136	138	205	122	161	174	
9	1	150	225	232	197	144	136	142	229	130	176	192	
10	1	138	195	230	195	140	122	134	211	126	178	156	
10	1	148	205	268	205	140	134	138	211	130	180	174	
11	1	138	185	232	205	140	132	134	205	128	165	154	
11	1	152	205	252	209	140	142	138	205	130	180	176	
12	1	148	193	232	193	140	134	136	211	126	165	168	
12	1	154	193	236	205	146	140	136	211	130	171	174	
1	2	146	195	232	197	140	122	134	197	124	165	154	
1	2	150	195	274	197	160	134	134	209	130	178	174	
2	2	148	193	232	193	144	132	134	203	126	171	174	
2	2	154	205	248	205	150	132	134	209	130	176	194	
3	2	148	195	228	193	136	122	134	205	126	163	154	
3	2	156	195	232	200	140	134	138	209	130	176	174	
4	2	138	191	236	191	140	132	138	209	126	176	154	
4	2	156	197	258	191	160	132	138	209	130	178	196	
5	2	142	197	260	191	140	122	136	207	126	165	154	
5	2	150	197	266	211	150	134	136	207	130	178	174	

APPENDIX G

A part of Python data conversion scripts

```
fh = open("allele.txt")
allele_names = fh.readline().strip().split("\t")
all_data = fh.readlines()

#genepop
def genepop(all_data):
    file = open("genepop.txt", "w")
    number_of_pops = {}
    for i in all_data:
        line = i.strip().split("\t")
        if line[1] not in number_of_pops:
            number_of_pops[line[1]] = str(len(number_of_pops) + 1)
+ "_"
    file.write("Title line: genepop.txt"+ "\n")
    for i in number_of_pops:
        file.write(i +"\n")

    for i in range(len(all_data)):
        if i < 1:
            file.write("Pop\n")
            x = all_data[i].strip().split("\t")[1:]
            x[0] = number_of_pops[x[0]]
            x.insert(1, ",")
            file.write("\t".join(x)+"\n")
        else:
            if all_data[i].strip().split("\t")[1] != all_data[i-
1].strip().split("\t")[1]:
                file.write("Pop\n")
                x = all_data[i].strip().split("\t")[1:]
                x[0] = number_of_pops[x[0]]
                x.insert(1, ",")
                file.write("\t".join(x)+"\n")
            else:
                x = all_data[i].strip().split("\t")[1:]
                x[0] = number_of_pops[x[0]]
                x.insert(1, ",")
                file.write("\t".join(x)+"\n")


#structure_popwise
def structure_popwise(all_data):
    file = open("structure_popwise.txt", "w")
    file.write("\t" + "\t" + "\t".join(allele_names) + "\n")
    number_of_pops = {}
    for i in all_data:
        line = i.strip().split("\t")
        if line[1] not in number_of_pops:
            number_of_pops[line[1]] = str(len(number_of_pops) + 1)
    for i in all_data:
        x = i.strip().split("\t")
        x[1] = number_of_pops[x[1]]
        list1 = []
```

```
        list2 = []
        for j in range(len(x)):
            if j < 2:
                list1.append(x[j])
                list2.append(x[j])
            else:
                if x[j][:3] == "000":
                    list1.append("-9")
                    list2.append("-9")
                else:
                    list1.append(x[j][:3])
                    list2.append(x[j][3:])
        file.write("\t".join(list1) + "\n")
        file.write("\t".join(list2) + "\n")

#structure_all_one

def structure_all_one(all_data):
    file = open("structure_all_one.txt", "w")
    file.write("\t" + "\t" + "\t".join(allele_names) + "\n")
    for i in all_data:
        x = i.strip().split("\t")
        x[1] = "1"
        list1 = []
        list2 = []
        for j in range(len(x)):
            if j < 2:
                list1.append(x[j])
                list2.append(x[j])
            else:
                if x[j][:3] == "000":
                    list1.append("-9")
                    list2.append("-9")
                else:
                    list1.append(x[j][:3])
                    list2.append(x[j][3:])
        file.write("\t".join(list1) + "\n")
        file.write("\t".join(list2) + "\n")
```

APPENDIX H

R Script

```
install.packages("poppr")
library("poppr")
install.packages("pegas")
library("pegas")

without = read.genepop("genepop_pyrus_without_metu.gen", ncode = 3)
with = read.genepop("genepop_pyrus.gen", ncode = 3)


install.packages("diveRsity")
library("diveRsity")
divMigrate(without)

install.packages("polysat")
library("polysat")
PIC(without)


#allelic richness
install.packages("PopGenReport")
library("PopGenReport")
allel.rich(without)

gac_without <- genotype_curve(without, sample = 1000, quiet = TRUE)
gac_with <- genotype_curve(with, sample = 1000, quiet = TRUE)

#missing data
info_table(without, type = "missing", plot = TRUE)

#removing missing data
without_no_missing = without %>% missingno("loci") %>%
  info_table(plot = TRUE, scale = FALSE)

with = with %>% missingno("loci") %>%
  info_table(plot = TRUE, scale = FALSE)

library("magrittr")

#locuswise simpson index, evennes, nei's gene diversity
locus_table(without)
poppr(without)

locus_table(with)
poppr(with)

#index of association
ia(without, sample = 1000)
ia(with, sample = 1000)

mxpair <- without %>% pair.ia
mxpair
```

```
geneclone_without = as.genclone(without)
geneclone_without

geneclone_with = as.genclone(with)
geneclone_with

#clone correction
clonecorrected =  with %>% clonecorrect(strata= ~Continent/Country)
%>%
  ia(sample = 999)

clonecorrected =  without %>% clonecorrect() %>%
  ia(sample = 999)


withoutpair <- with %>% pair.ia
with

install.packages("mmod")
library("mmod")

#standardized gst
Gst_Hedrick(without)

#Phylogenetic tree
with %>%
  genind2genpop() %>%
  aboot(cutoff = 50, quiet = TRUE, sample = 1000, distance =
nei.dist)


without %>%
  genind2genpop() %>%
  aboot(cutoff = 50, quiet = TRUE, sample = 1000, distance =
nei.dist)


#cluster analysis

without_clust <- find.clusters(without)
with_clust <- find.clusters(with)
without_clust


withtree <- bruvo.boot(with, replen = c(2, 2, 2, 4, 2, 2, 2, 2, 2,
2, 4),
                       cutoff = 50, quiet = TRUE)
library("ape")
cols <- rainbow(4)
plot.phylo(withtree, cex = 0.8, font = 2, adj = 0, tip.color =
cols[with_clust$grp],
           label.offset = 0.0125)
nodelabels(withtree$node.label, adj = c(1.3, -0.5), frame = "n",
cex = 0.8,
           font = 3, xpd = TRUE)
axisPhylo(3)
```

```
#AMOVA

withoutamova <- poppr.amova(without, ~pop/Individual)
withamova

#DAPC

library(adegenet)

dapc.without <- dapc(without, var.contrib = TRUE, scale = FALSE,
                     n.pca = 30, n.da = nPop(without) - 1)
levels(dapc.without$grp) = c("A.Koyunlu", "Camili", "Kirazlı",
"Meşeli", "Meydancık",
                             "SCC", "Panta", "Veliköy")
scatter(dapc.without, cell = 0, pch = 18:23, cstar = 0, mstree =
TRUE,
        lwd = 2, lty = 2)


dapc.with <- dapc(with, var.contrib = TRUE, scale = FALSE,
                     n.pca = 30, n.da = nPop(with) - 1)
scatter(dapc.with, cell = 0, pch = 18:23, cstar = 0, mstree = TRUE,
        lwd = 2, lty = 2)


contrib_without <- loadingplot(dapc.without$var.contr,
                       axis = 1, thres = 0.05, lab.jitter = 1)


contrib_with <- loadingplot(dapc.with$var.contr,
                              axis = 1, thres = 0.07, lab.jitter =
1)


pramx_without <- xvalDapc(tab(without, NA.method = "mean"),
pop(without))

system.time(pramx_without <- xvalDapc(tab(without, NA.method =
"mean"),
                                      pop(without),
                             n.pca = 1:11, n.rep = 1000,
                             parallel = "snow", ncpus = 4L))
names(pramx_without)
pramx_without[-1]

scatter(pramx_without$DAPC, cex = 2, legend = TRUE,
        clabel = FALSE, posi.leg = "bottomleft", scree.pca = TRUE,
        posi.pca = "topleft", cleg = 0.75, xax = 1, yax = 2,
inset.solid = 1)

pramx_with <- xvalDapc(tab(with, NA.method = "mean"), pop(with))

system.time(pramx_with <- xvalDapc(tab(with, NA.method = "mean"),
                                   pop(with),
                                   n.pca = 9:15, n.rep = 1000,
                                   parallel = "snow", ncpus =
4L))
```

```
names(pramx_with)
pramx_with[-1]


scatter(pramx_without$DAPC, cex = 2, legend = TRUE,
        clabel = FALSE, posi.leg = "bottomleft", scree.pca = TRUE,
        posi.pca = "topleft", cleg = 0.75, xax = 1, yax = 2,
inset.solid = 1)

#Pairwise fst
aa = pairwise.fst(without)
ma = as.matrix(aa)
colnames(ma) = c("A.Koyunlu", "Camili", "Kirazlı", "Meşeli",
                 "Meydancık", "Özel Bahçe", "Panta", "Veliköy")
row.names(ma) = c("A.Koyunlu", "Camili", "Kirazlı", "Meşeli",
                  "Meydancık", "Özel Bahçe", "Panta", "Veliköy")
ma
heatmap(ma)
da = as.data.frame(ma)
ggplot(da)
ggplot(data = da)


install.packages("tidyverse")
library(tidyverse)

dt2 <- da %>%
  rownames_to_column() %>%
  gather(colname, value, -rowname)
head(dt2)


ggplot(dt2, aes(x = rowname, y = colname, fill = value)) +
  geom_tile()

#Minimum spanning networks
levels(without$pop) = c("A.Koyunlu", "Camili", "Kirazlı", "Meşeli",
                        "Meydancık", "SCC", "Panta", "Veliköy")
without$pop
imsn()


phenotype = read.genepop("based_on_phenotype.gen", ncode = 3)
levels(phenotype$pop) = c("Acara", "Argevul", "Asma armudu", "Bal
armudu", "Har armudu",
                          "Işık Mehmet", "Kabağa", "Kansilya",
"Panta", "Ruma",
                          "Sasila", "Şanaşor", "Tevracul", "Yağ
armudu",
                          "zsozola")

levels(with$pop) = c("A.Koyunlu", "Camili", "Kirazlı", "Meşeli",
                        "Meydancık", "SCC", "Panta", "Veliköy",
"METU")
imsn()
```

APPENDIX I

Statistical Calculations

MICRO-CHECKER:

Brookfield null allele estimation (1996):

$r = (He - Ho) / (1 + He)$

Where He is expected heterozygosity and Ho is observed heterozygosity.

GenAlEx and R (poppr Package)

Linkage Disequilibrium (Index of Association):

$I_A = (V_o / V_e) - 1$

Where V_o is observed variance and the V_e is the expected variance under no linkage.

Number of different alleles (Na):

Direct count of alleles.

Effective Number of Alleles (Ne):

$Ne = 1 / (1 - He)$

Where He is the expected heterozygosity.

Polymorphic Information Content (PIC):

$PIC = 1 - \sum(pi)^2$

Where pi is the frequency of i^{th} allele.

Observed Heterozygosity (Ho):

It is calculated by number of heterozygotes divided by sample size.

Ho = (Number of heterozygotes) / Sample Size

Expected Heterozygosity (He):

$He = 1 - \sum pi^2$

Where pi is the allele frequency

Fixation Index (F):

$F = (He - Ho) / He$

Where He is expected heterozygosity and Ho is observed heterozygosity.

F Statistics (Fis, Fst, Fit) and Number of Migrants (Nm):

$F_{ST} = (H_T – H_S) / H_T$

$F_{IS} = (H_S – H_I) / H_S$

$F_{IT} = (H_T – H_I) / H_T$

Where H_I is average heterozygosity in individuals over all populations, H_S is average of expected heterozygosity over all subpopulations, and H_T is expected heterozygosity ignoring population structure.

$Nm = ((1 / F_{ST}) – 1) / 4$

Private Allele (Pa):

Total number of unique alleles to of the population.

Percentage of Polymorphic Loci (P%):

$P = \sum(P_i / N)$

Where P_i is the proportion of loci polymorphic in a population and N is the number of populations.

APPENDIX J

Summary of Private Alleles

Pop	Locus	Allele	Freq
A.Koyunlu	KU10	252	0,042
A.Koyunlu	Bgt23b	229	0,042
Camili	NB113a	142	0,107
Camili	KU10	266	0,036
Camili	Bgt23b	203	0,036
Camili	CH01F02	183	0,036
Kirazlı	NH013a	215	0,045
Kirazlı	NH008b	207	0,136
Meşeli	CH03G06	168	0,083
Meydancık	CH02F06	150	0,050
SCC	CH03G06	152	0,067
SCC	CH01F02	173	0,067
Panta	KU10	228	0,100
Panta	KU10	262	0,100
Panta	Bgt23b	193	0,100
Veliköy	CH02F06	170	0,100