# THE EFFECTS OF SEX RATIO IN ESTIMATION OF GENETIC DIFFERENTIATION IN *POPULUS NIGRA* POPULATIONS

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

BY

BURAK YELMEN

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

JUNE 2017

Approval of the thesis:

**THE EFFECTS OF SEX RATIO IN ESTIMATION OF GENETIC DIFFERENTIATION IN *POPULUS NIGRA* POPULATIONS**

submitted by **BURAK YELMEN** in partial fulfillment of the requirements for the degree of **Master of Science in Biology Department, Middle East Technical University** by,

Prof. Dr. Gülbin Dural Ünver
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. Orhan Adalı
Head of the Department, **Biological Sciences** _____

Prof. Dr. Zeki Kaya
Supervisor, **Biological Sciences Dept., METU** _____

**Examining Committee Members:**

Prof. Dr. Ayşegül Çetin Gözen
Biological Sciences Dept., METU _____

Prof. Dr. Zeki Kaya
Biological Sciences Dept., METU _____

Prof. Dr. Sertaç Önde
Biological Sciences Dept., METU _____

Prof. Dr. İrfan Kandemir
Biology Dept., Ankara Uni. _____

Yrd. Doç. Dr. Fatih Temel
Faculty of Forestry, Düzce University _____

**Date:** 27.07.2017

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

Name, Last name:

Signature :

## ABSTRACT

### THE EFFECTS OF SEX RATIO IN ESTIMATION OF GENETIC DIFFERENTIATION IN *POPULUS NIGRA* POPULATIONS

Yelmen, Burak

M.S., Department of Biology

Supervisor: Prof. Dr. Zeki KAYA

June 2017, 84 pages

Effective population size is an important concept in conservation biology. Biased sex ratio lowers effective population size, consequently causing loss of genetic variation. In this study, available microsatellite DNA marker data of 121 *Populus nigra* clones originated from 5 geographical regions were analyzed to evaluate genetic diversity of genders and to investigate possible effects of sex ratio on differentiation in these populations in Turkey.

There was an abundance of identical genotypes in the dataset. The same genotypes were observed both in males and females that might be suggesting a rare occurrence of deviation from dioecism. Three clusters were estimated in structure analysis with k-means clustering. Significant differentiation was found among clusters. However, they did not correlate with geographical assignment of clones. Also, there was no variation between pre-determined five geographical populations. Overall allelic richness was found to be similar for both genders whereas heterozygosity was higher in males. Both male and female group had private alleles on all studied loci. Although

there was no significant variation between genders, combination of four loci showed a slight empirical differentiation. Additionally, a simulation software prototype was developed to see effects of sex ratio on diversity and differentiation in future generations given the available molecular data. Results showed that if biased sex ratio persists, allele loss and fixation might occur in a higher rate, causing a loss of variation and faster differentiation in case of isolation.

**Key Words:** *Populus, microsatellite marker, sex ratio, genetic differentiation, simulation*

# ÖZ

## *POPULUS NIGRA* POPÜLASYONLARINDA EŞEY ORANININ GENETİK FARKLILAŞMAYA OLAN ETKİSİ

Yelmen, Burak
Yüksek lisans, Biyoloji Bölümü
Tez Yöneticis: Prof. Dr. Zeki Kaya
Haziran 2017, 84 sayfa

Etkili popülasyon büyüklüğü koruma çalışmaları için önemli bir kavramdır. Erkek-dişi oranındaki sapma, etkili popülasyon büyüklüğünü düşürerek genetik varyasyonda kayba neden olur. Bu çalışmada, 5 coğrafik bölgeden 121 klona ait mikrosatellit DNA belirteç verileri kullanılarak, Türkiye'deki *Populus nigra* populasyonlarında cinsiyetler arasındaki genetik çeşitlilik değerlendirilmiş ve erkek-dişi oranındaki sapmanın genetik farklılaşmaya etkisi araştırılmıştır.

Mevcut veri setinde çok sayıda özdeş genotip tespit edildi. Aynı genotip hem dişi hem erkek bireylerde gözlemlendi. Bu durum nadir görülen, ikievcikli üreme yapısından sapmaya işaret ediyor olabilir. Yapı analizinde k-means gruplama yöntemi kullanılarak üç farklı küme tahmin edildi ve bu kümeler arasında kayda değer farklılaşma tespit edildi. Ancak bu kümeler, bireylerin bölgesel sınıflandırılması ile benzerlik göstermedi. Ayrıca, önceden tanımlanmış 5 coğrafik popülasyon arasında varyasyon görülmedi. Toplam ortalama alelik zenginlik erkek ve dişi grubu için aynı bulunurken, heterozigotluğun erkek grubunda daha fazla olduğu görüldü. Çalışılan lokuslarda, erkek ve dişilere özel alleller tespit edildi.

Erkek ve dişiler arasında anlamlı bir genetik varyasyon görülmemesine rağmen dört lokus kombinasyonu ayrı incelendiğinde empirik bir farklılaşma gözlemlendi. Bunlara ek olarak, bir simülasyon yazılımı prototipi gelişitiridi ve eldeki veriler kullanılarak erkek-dişi oranındaki eşitsizliğin gelecek nesillerde genetik çeşitlilik ve farklılaşma üzerindeki etkileri incelendi. Sonuçlar, eğer eşitsizlik devam ederse, alel kaybı ve fiksasyonunun daha çabuk gerçekleşebileceğini ve genetik varyasyon kaybı görülebileceğini, izolayson oluşması durumundaysa farklılaşmanın daha hızlı olabileceğini gösterdi.

**Anahtar Kelimeler:** *Populus,* mikrosatellit markörü, cinsiyet oranı, genetik farklılaşma, simülasyon

To My Mother

## ACKNOWLEDGEMENTS

I sincerely want to thank my supervisor Prof. Dr. Zeki Kaya for his patience and guidance throughout the study.

I also specifically thank to Asiye Çiftçi along with my supervisor for their help and providing the data for this thesis.

I would like to express my gratitude to Alev Ateş, Bircan Taşkıran, Çiğdem Kansu, Funda Özdemir Değirmenci, Pelin Keske Acar and Yasin Tokdemir for their support during my MSc studies.

I want to express my thanks to all jury members for their valuable comments and criticism.

Data used in this thesis was obtained from a previous study which was financially supported by TUBITAK TOVAG 110O570, BAP-08-11-DPT2002K12510 and Ministry of Forest and Water Affairs research funds.

## TABLE OF CONTENTS

## LIST OF TABLES

TABLES

## LIST OF FIGURES

FIGURES

## LIST OF ABBREVIATIONS

| | |
|---|---|
| AMOVA | Analysis of Molecular Variance |
| ANOVA | Analysis of Variance |
| DAPC | Discriminant Analysis of Principal Components |
| HWE | Hardy-Weinberg Equilibrium |
| LD | Linkage Disequilibrium |
| PCA | Principal Component Analysis |
| PCR | Polymerase Chain Reaction |

# CHAPTER 1

# INTRODUCTION

*Populus nigra* (black poplar) (Figure 1) is a widely spread dioecious tree (having male and female flowers on different individuals) which has ecological and economical importance for Europe and Turkey (Smulders *et al.,* 2008; Kaya, 2013). It is dispersed widely over large portions of Eurasia as a pioneer species mainly in riparian areas. Along with being dioecious, it can also reproduce by vegetative propagation naturally (Arens *et al.,* 1998).

European black poplar populations can exist in a variety of types as isolated trees and large mixed or pure stands (Lefevre *et al.,* 1998). Metapopulation formations occur by colonization of open regions via seeds or stem cuttings and root suckers (Herpka, 1986). *P. nigra*, having economic importance, is broadly planted in Turkey for domestic use. Over 60 000 ha planted as a pure species in row plantations whereas over 70 000 ha planted in hybrid plantations (Tunçtaner, 1995). It is also planted for soil protection and afforestation of polluted areas. Aside from its use as a pure species, black poplar is mostly used as a genetic source for breeding programmes for production of more adaptive hybrids (Lefevre *et al.,* 1998).




**Figure 1.** *Populus nigra*

**A)** Black poplar tree (www.onlineplantguide.com, retrieved on 28.07.2017)

**B)** Male flowers (www.ornamental-trees.co.uk, retrieved on 28.07.2017)

**C)** Female flowers (www.bgflora.net, retrieved on 28.07.2017)

An important part of the genetic resource of the European black poplars in Turkey is held in a clone bank located in Ankara. Trees preserved in this clone bank were studied by Çiftçi *et al.* (2017) using polymorphic microsatellites. In addition to the six geographical populations from this clone bank, two newly found natural populations were also analyzed. The study found that clone collections are significantly differentiated from natural populations whereas little differentiation was observed among clone bank populations. Two clusters were estimated by population structure analysis, one cluster covering the natural populations and the other one clone bank populations. The researchers stated that high admixture, clonal duplication and reduced genetic differentiation observed in the clone collection populations are all indicators of the deterioration in the genetic resource of black poplars in Turkey, which is probably caused by human activity. A portion of the clone bank samples used in this study had gender label and sex ratio of this subset of data was unbalanced in favor of the females.

Male and female biased sex ratios are common in plant studies (Delph, 1999). Difference between the production cost of male and female flowers have been the primary explanation for male-biased sex ratios in dioecious plants, as female plants generally spend higher amount of resources for reproduction than males (Stehlik *et al.,* 2005). On the other hand, several models have been offered for female bias, but none with definitive experimental results (Barret *et al.,* 2010). Sex ratios in *Populus nigra* populations are also reported to have high variability. One suggested explanation is that different genders have different responses to environmental conditions such as availability of water (Hughes *et al*., 2000). The study by Hughes *et al.* (2000) found that female black poplars might tend to prefer wetter regions with more nutrients compared to males, which in return, might effect the spatial distribution of genders and overall sex ratios based on available regional resources. They also concluded that flow regimes can cause habitats to become more favorable for one gender, thus give rise to distorted sex ratios in floodplains. In another study,

gender frequencies were found to differ at different elevations and in seeds from open-pollinated flowers for female biased dioecious plant species (Stehlik *et al.,* 2005). All frequencies were in favor of females but different factors seem to affect the scope of the bias.

Sex determination mechanisms in plants are widely diverse, including sex chromosomes, autosomal traits and external factors. However, most of the molecular mechanisms are mainly unknown or not well understood. (Moneger, 2007). Sex is considered to be an autosomal trait in black poplar, mapped onto chromosome XIX as sex locus. Male is heterogametic whereas female is homogametic, meaning sex locus is only mapped on male individuals (Gaudet *et al.,* 2007). Complete sex determination mechanism in *Populus nigra* is thought to involve multiple loci and possibly environmental effects (Mcletchie & Tuskan*,* 1994).

Real population size usually differs from census size since every individual's probability of contributing to the genetic pool differs. Biased sex ratio causes the effective population size ($N_e$) to decrease while the rate of inbreeding and drift increase accordingly (Kliman *et al.,* 2008). Inbreeding and drift are important contributors to fragmentation. Therefore, population structure can significantly change when sex ratio differs from 1:1 in demes (Frankham, 2015). Various methods have been proposed to estimate effective population size with available molecular data using heterozygote excess, linkage disequilibrium and genetic variation. Each one of these approaches have their own biases and advantages (Wang, 2005). In an idealized population, Robertson (1965) showed that heterozygote excess in the progeny given the male and female numbers of the parents is;

$$H_e = -\frac{1}{8N_m} - \frac{1}{8N_f} = -\frac{1}{2N_e}$$

which is then transformed to the formula for estimation of effective population size for populations with biased sex ratio;

$$N_e = \frac{4N_m N_f}{N_m + N_f}$$

where $N_m$ is the number of males and $N_f$ is the number of females. The formula suggests, as the biased ratio increases, number of individuals contributing to the genetic pool decreases. Though, it is important to state that this approach is only valid under some unrealistic assumptions like complete random mating and equal chance of reproduction for all individuals. Different modifications of this generalized formula to estimate effective population size under different assumptions have been proposed (Caballero, 1994). In this study, effects of unbalanced sex ratio on European black poplars in Turkey were examined along with possible reasons behind the bias and other related findings from the available data.

# CHAPTER 2

## OBJECTIVES OF THE STUDY

Main objective of the study is to assess possible effects of unbalanced sex ratio on genetic diversity and differentiation of *Populus nigra* species in Turkey through available microsatellite data. The broad objective includes:

- Detecting possible genetic structure to analyze contributions from both genders,
- Inspecting genetic variation between genders,
- Simulating future effects of unbalanced sex ratio on genetic diversity and differentiation.

# CHAPTER 3

# MATERIALS AND METHODS

## 1. Data Assessment

Main data was constructed from a subset of microsatellite data used in the study by Ciftci *et al.* (2017). Only individuals with gender information were selected. The data from 121 clones originated from 5 geographic regions (Table 1) which included 12 diploid loci data were combined into different formats. Formats used for analysis of the data are *GENEPOP* (Raymond & Rousset, 1995; Rousset, 2008) and *NEXUS* (Maddison *et al.*, 1997) file formats, *genind* (for individual genotypes) and *genpop* (for population allele counts) classes from R package *adegenet* (Jombart, 2008; Jombart & Ahmed, 2011). For conversion of file formats, mostly a developed Python (Python Software Foundation, https://www.python.org) script was used (Appendix A). Majority of applications were done by R programming language (R Core Team, 2016). Genotype and allele data were further combined with gender, coordinate, altitude and provincial information of individuals for analysis of possible correlations. Missing data analysis and construction of genotype accumulation curve were performed by using R *poppr* package (Kamvar *et al.*, 2014; Kamvar *et al.*, 2015). Genotype accumulation curve was constructed by randomly sampling loci without replacement up to n-1 = 11, and counting the unique genotypes at every addition of new locus until the real number of unique genotypes is reached. When necessary, clone correction was done following population/gender hierarchy (unless stated otherwise) so that after clone correction, each gender within each population would have no genotypic copies. To be able to see the dispersal of individuals and

populations, and to determine possible clones, QGIS Open Source Geographic Information System (QGIS Development Team, 2009) was used. All individuals which have coordinate and provincial data were placed on the retrieved Google satellite map (Google Earth, 2008). A table of individuals in each geographical population with their available data (city, longitude, latitude, altitude, gender and genotype) is available in Appendix B.

**Table 1.** Sample sizes with respect to population and gender

| | **Male** | **Female** | **Total** |
|---|---|---|---|
| **Central A.** | 8 | 18 | 33 |
| **Eastern A.** | 10 | 18 | 29 |
| **Aegean** | 4 | 20 | 16 |
| **Blacksea** | 8 | 17 | 34 |
| **Marmara** | 6 | 12 | 17 |
| **Total** | 36 | 85 | 121 |

## 2. Locus Quality Control

### 2.1. Hardy – Weinberg Equilibrium

The Hardy – Weinberg principle states that allele frequencies of a sexually reproducing populations will not change in the absence of evolutionary forces (Gillespie, 2004). Analysis of loci for Hardy – Weinberg equilibrium state was performed using R *pegas* package (Paradis, 2010). An exact test approach using Monte Carlo permutations and a chi-squared approach which calculates genotype frequencies from allele frequencies were applied for significance testing.

## 2.2. Identifying Loci with Possible Null Alleles

For identification of loci with possible null alleles, implementation of formula by Brookfield (1996) was used via R *PopGenReport* package (Adamack & Gruber 2014; Gruber & Adamack 2015). Formula is basically;

$$r = \frac{H_e - H_o}{1 + H_e}$$

where r is estimated frequency of null alleles, $H_e$ is expected heterozygosity and $H_o$ is observed heterozygosity. Median, $2.5^{th}$ and $97.5^{th}$ percentile values were estimated via bootstrapping results.

## 2.3. Linkage Disequilibrium

Linkage disequilibrium suggests non-random association of alleles on different loci, and can occur due to phenomenon like differentiation among populations, asexual reproduction, linkage between alleles, selection and genetic drift (Agapow and Burt, 2001).

Linkage disequilibrium was estimated using a modification of index of association (Brown *et al.* 1980) proposed by Agapow and Burt, and implemented in R *poppr* package (Kamvar *et al.*, 2014; 2015). Index of association is;

$$I_a = \left(\frac{V_o}{V_e}\right) - 1$$

where $V_o$ is observed variance of pairwise distances and $V_e$ is expected variance when there is no linkage disequilibrium. Estimations were done after complete clone correction so that only one representative of each genotype would remain.

## 3. Diversity Measures

### 3.1. Allelic Richness

Allelic richness for each population and each locus was estimated via the methods proposed by Mousadik & Petit (1996) and implemented in R *PopGenReport* package (Adamack & Gruber 2014; Gruber & Adamack 2015). Basically, allelic richness is an estimation of mean observed number of alleles.

### 3.2. Genetic Diversity

Genetic diversity was assessed with expected heterozygosity as provided by Nei (1973). Average gene diversity per locus is;

$$H_s = \frac{1}{k} * \sum_{s=1}^{k} [1 - q_s^2 - (1 - q_s)^2]$$

where k is the loci number and $q_s$ is frequency of one allele at the $s^{th}$ locus.
The unbiased $H_s$ by Nei (1978) is;

$$H_s' = H_s * \frac{2n}{2n - 1}$$

where n is the number of individuals.

## 4. Population Structure Analysis

### 4.1. Identification of Population Structure

For cluster identification without any prior population information, data was transformed by principal component analysis (PCA) and k-means clustering method was performed via tools in R adegenet package (Jombart, 2008; Jombart & Ahmed, 2011). Theoretical optimum cluster number (k) was chosen based on suggestions by Jombart *et al.* (2010).

Furthermore, STRUCTURE software (Pritchard *et al.*, 2000; Falush *et al.,* 2007; Hubisz *et al.,* 2009) was used to estimate population structure with following parameters; 50000 burning length, 100000 MCMC repeats after burning, admixture ancestry model and correlated allele frequencies. Best K (true number of clusters) was determined with the method described by Evanno *et al.* (2005) using STRUCTURE HARVESTER software (Earl *et al*. 2012).

Found clusters were further described by discriminant analysis of principal components. DAPC is a method developed by Jombart *et al*. (2010) to describe genetic clusters via a few synthetic variables. These variables are obtained from alleles transformed into principal components as discriminant functions. Main rationale is to identify clusters by alleles with largest between group and smallest within group variances so that discriminant functions are inclined to show best variation among groups. DAPC was performed with R adegenet package implementation (Jombart, 2008; Jombart & Ahmed, 2011).

### 4.2. Genetic Differentiation, F-statistics and G-statistics

Weir & Cockerham (1984) $F_{ST}$ and $F_{IS}$ (inbreeding coefficient), Nei's $G_{ST}$ (Nei, 1973) and Hedrick's standardized $G'_{ST}$ (Hedrick, 2005) were used to evaluate population structure and genetic differentiation. Calculations of these indices were implemented in R mmod package (Winter, 2012) and R hierfstat package (Goudet, 2005). First proposed by Wright (1943), $F_{ST}$ and its modern derivatives are indices for estimation of genetic differentiation among demes based on allele frequencies. Theoretically, a minimum value 0.0 of $F_{ST}$ analogues indicates same allelic composition between demes and maximum value 1.0 indicates no shared alleles. For highly polymorphic markers like microsatellites, differentiation is usually underestimated by indices (Hedrick, 1999). There is still an ongoing debate on which $F_{ST}$ analogue and approach to use for a better estimation of differentiation (Bird *et al.,* 2011).

### 4.3. Analysis of Molecular Variance

For analysis of molecular variance to study molecular variation in a hierarchical manner, R poppr package (Kamvar *et al.*, 2014; 2015) implementation of the original method was used (Excoffier *et al.,* 1992). After transformation of molecular data to Euclidian distances from Boolean vectors, variance of hierarchical structures in a population is calculated via nested analysis of variance (ANOVA) approach. Therefore, within and between group differences at many hierarchical levels can be estimated. Significance of variance is determined by resampling the data with several permutations.

## 5. Correlation Analysis

In analysis of variance (ANOVA), equality of two or more group means is tested by comparison of variance among groups and variance within groups (Larson, 2008). ANOVA was used to test correlation between different categorical groups which have numerical variables. Chi squared test of independence (Agresti, 2007) was used to test the independence of two categorical variables and p-values were inferred by Monte Carlo significance test (Hope, 1968). Both tests were performed via R core functions (R Core Team, 2016).

## 6. Simulating Effects of Sex Ratio

An R software prototype was developed to simulate subsequent generations and provide population genetics parameters (APPENDIX C). The tool can be used to observe effects of migration, isolation, genetic drift, sex ratio and population size on genetic diversity and differentiation over generations. It's possible to subset any range of generations or a particular generation for other applications. Migration rates between all population pairs can be provided for better estimations. Also, it is possible to introduce isolation at a given generation. Principle assumptions are no linkage between alleles, no selection, constant migration rate, non-overlapping generations and no introduction of new alleles via mutation or migration. Tool uses functions from R *adegenet* (Jombart, 2008; Jombart & Ahmed, 2011), *poppr* (Kamvar *et al.*, 2014; 2015) and *hierfstat* (Goudet, 2005) packages for storing data and calculating differentiation parameters.

Since $F_{ST}$ based calculation of migration assumes same population size and constant migration among all populations, for the estimation of migration, MIGRATE software was used (Beerli 2009). Bayesian implementation provided by the software

was chosen to infer mutation-scaled population sizes and mutation-scaled immigration rates (Beerli 2006). Mutation-scaled population size is $\Theta = x * N_e * \mu$ where x is a multiplier depending on the ploidy, $N_e$ is effective population size and $\mu$ is mutation rate. Mutation-scaled immigration rate is $M = m / \mu$ where $m$ is immigration rate and $\mu$ is mutation rate. To estimate number of individual migrants per each population pair, these two parameters were multiplied.

# CHAPTER 4

# RESULTS

## 1. Quality of Data

### 1.1. Missing Data

Marmara and Central Anatolia populations had missing data on WPMS12 locus, Eastern Anatolia and Aegean populations had missing data on PMGC2163 locus. Mean missing allele data percentage over all loci and all individuals was 0.57 % (Figure 2). Individuals with missing data were contained in analyzes without including their missing alleles.

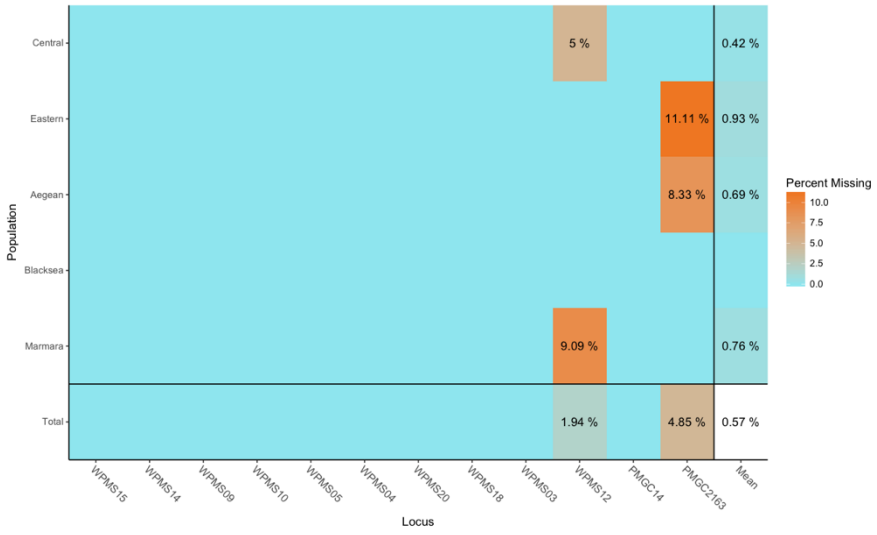



**Figure 2.** Percentage of missing data with respect to loci and populations

### 1.2. Genotype Accumulation Curve

Random sampling of loci for 1000 times provided that actual unique genotype number can be reached for 11 loci (Figure 3). This statistically indicates that loci count is enough for discrimination of individual trees.

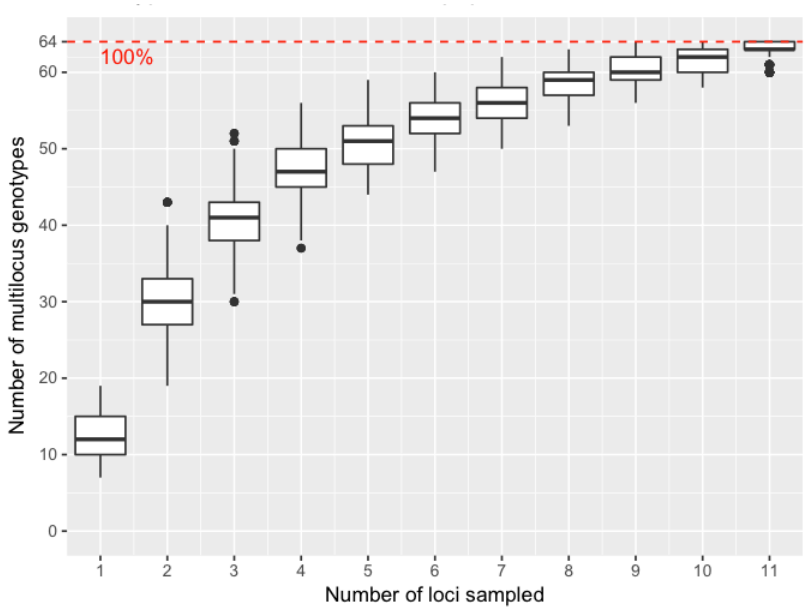



**Figure 3.** Genotype accumulation curve

## 2. Geographic Distribution

Two sample pairs had same coordinates and genotypes. Therefore, one set of pairs was excluded. Trees from the Mediterranean region were also discarded due to low sample size and distant locations. Additionally, from clones which have the same genotypes and are in close proximity, only one clone was taken to eliminate possible duplicated clones. These reductions resulted in a total of 103 clones (Table 2). Geographic assignment of populations was done by considering coordinate and provincial information (provided as city names) along with geographical characteristics. For instance, geographically adjacent clones around Eğirdir Lake would belong to Mediterranean, Aegean and Central Anatolia regions only due to provincial borders. These clones were grouped under the same region. There was no apparent visual clustering of individuals with the same gender or genotype (Figure 4).

**Table 2.** Sample sizes with respect to population and gender after excluding close proximity genotypic duplicates

| | **Male** | **Female** | **Total** |
|---|---|---|---|
| **Central A.** | 7 (8) | 13 (18) | 20 (33) |
| **Eastern A.** | 9 (10) | 18 (18) | 27 (29) |
| **Aegean** | 4 (4) | 20 (20) | 24 (16) |
| **Blacksea** | 8 (8) | 13 (17) | 21 (34) |
| **Marmara** | 6 (6) | 5 (12) | 11 (17) |
| **Total** | 34 (36) | 69 (85) | 103 (121) |

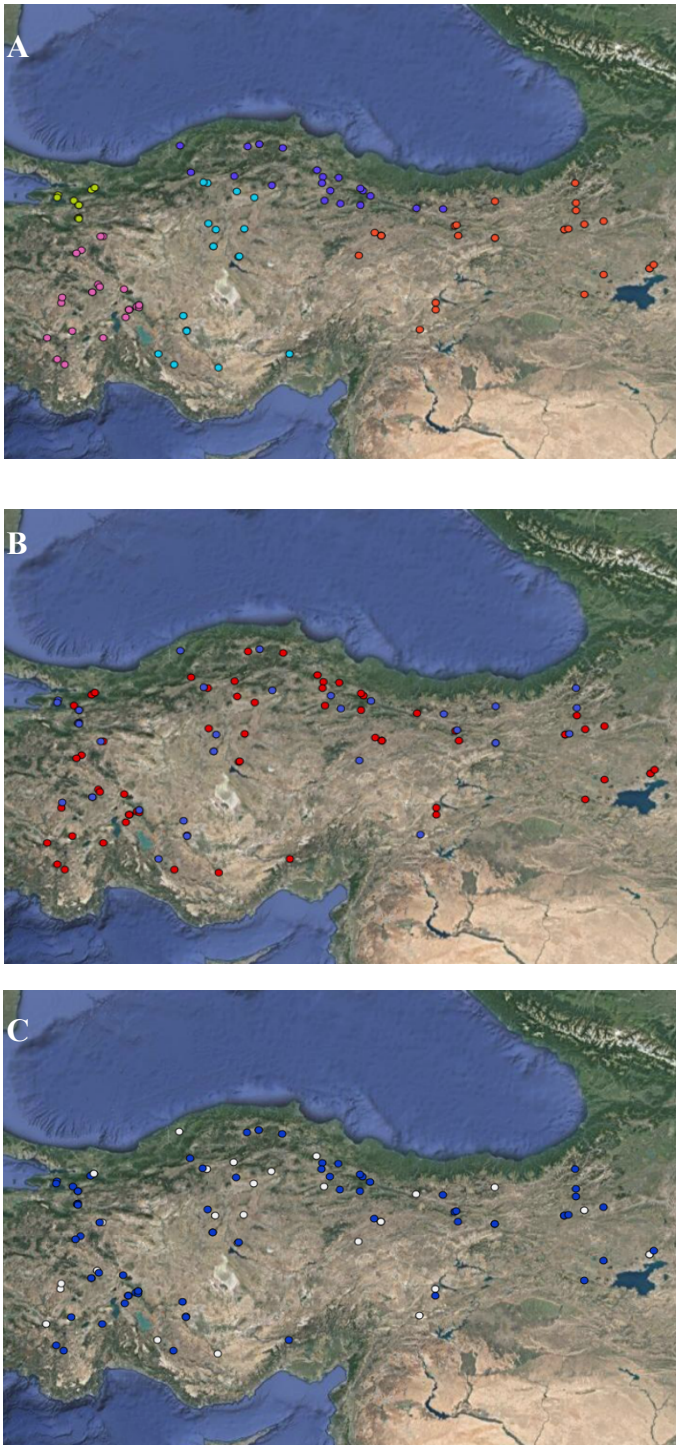



**Figure 4. A)** Geographical classification of populations. Dots are individuals and colors represent populations (dots may contain more than 1 individual)

**B)** Gender distribution. Females are represented as red dots

**C)** Occurrence of the same genotype across regions, represented as white dots

## 3. Loci Analysis

### 3.1. Hardy – Weinberg Equilibrium

Both chi-squared analysis based on calculating expected genotypic frequencies from allele frequencies and Monte Carlo permutation based approach showed that none of the loci follow Hardy – Weinberg equilibrium ($p < 0.05$). Results did not change after clone correcting the data by excluding all identical clones having the same genotype.

**Table 3.** Hardy – Weinberg tests

| Loci | $\chi^2$ | Degrees of freedom | *P*-value ($\chi^2$) | *P*-value (exact) |
|---|---|---|---|---|
| **WPMS15** | 96.86 | 10 | 0.00 | 0.00 |
| **WPMS14** | 97.89 | 36 | 0.00 | 0.00 |
| **WPMS09** | 139.13 | 10 | 0.00 | 0.00 |
| **WPMS10** | 384.97 | 36 | 0.00 | 0.00 |
| **WPMS05** | 229.20 | 21 | 0.00 | 0.00 |
| **WPMS04** | 399.37 | 36 | 0.00 | 0.00 |
| **WPMS20** | 371.54 | 15 | 0.00 | 0.00 |
| **WPMS18** | 83.10 | 10 | 0.00 | 0.00 |
| **WPMS03** | 192.40 | 28 | 0.00 | 0.00 |
| **WPMS12** | 97.79 | 28 | 0.00 | 0.00 |
| **PMGC14** | 165.51 | 45 | 0.00 | 0.00 |
| **PMGC2163** | 427.30 | 91 | 0.00 | 0.00 |

When loci were examined in geographical population level, again most loci are not in HWE with $p < 0.05$ (Figure 5).

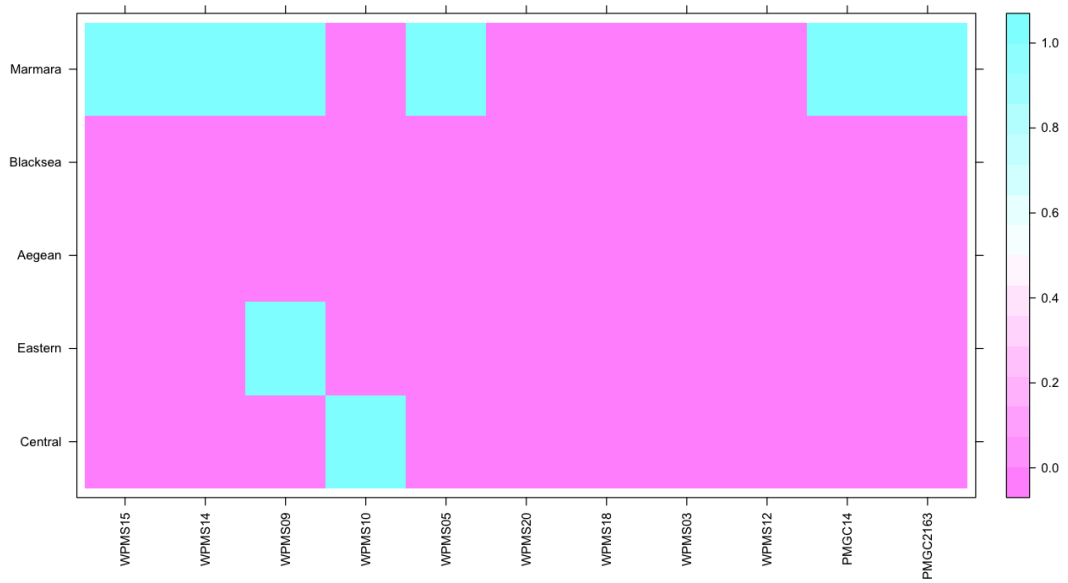



**Figure 5.** Hardy – Weinberg tests by population estimated by Monte Carlo permutations. Pink loci are highly likely to be not in HWE ($p < 0.05$)

### 3.2. Null Alleles

WPMS09, WPMS10, WPMS05 and WPMS04 loci showed positive values for null allele frequency estimations (Table 4). The highest value was estimated for WPMS04 ($r \approx 0.45$). This finding also correlates with the findings in the study by Çiftçi *et al.* (2017). Therefore, this locus was discarded in further applications.

**Table 4.** Estimation of the frequency of null alleles. WPMS04 is suspected of having high null allele frequency.

| | **WPMS15** | **WPMS14** | **WPMS09** | **WPMS10** | **WPMS05** | **WPMS04*** |
|---|---|---|---|---|---|---|
| **Observed frequency** | -0.16 | -0.13 | 0.11 | 0.10 | 0.15 | 0.45 |
| **Median frequency** | -0.16 | -0.13 | 0.11 | 0.09 | 0.15 | 0.44 |
| **2.5th percentile** | -0.19 | -0.16 | 0.05 | 0.03 | 0.08 | 0.36 |
| **97.5th percentile** | -0.13 | -0.09 | 0.19 | 0.16 | 0.24 | 0.53 |
| | **WPMS20** | **WPMS18** | **WPMS03** | **WPMS12** | **PMGC14** | **PMGC2163** |
| **Observed frequency** | -0.06 | -0.18 | -0.11 | -0.16 | -0.14 | -0.13 |
| **Median frequency** | -0.06 | -0.18 | -0.11 | -0.16 | -0.14 | -0.13 |
| **2.5th percentile** | -0.12 | -0.20 | -0.16 | -0.18 | -0.16 | -0.17 |
| **97.5th percentile** | -0.00 | -0.16 | -0.06 | -0.13 | -0.12 | -0.09 |

### 3.3. Linkage Disequilibrium

Overall linkage disequilibrium (LD) was found to be significant after 1000 permutations (Figures 6, 7). When each population is analyzed separately, all of them showed significant LD, Marmara having the lowest and Central Anatolia having the highest value (Table 5).

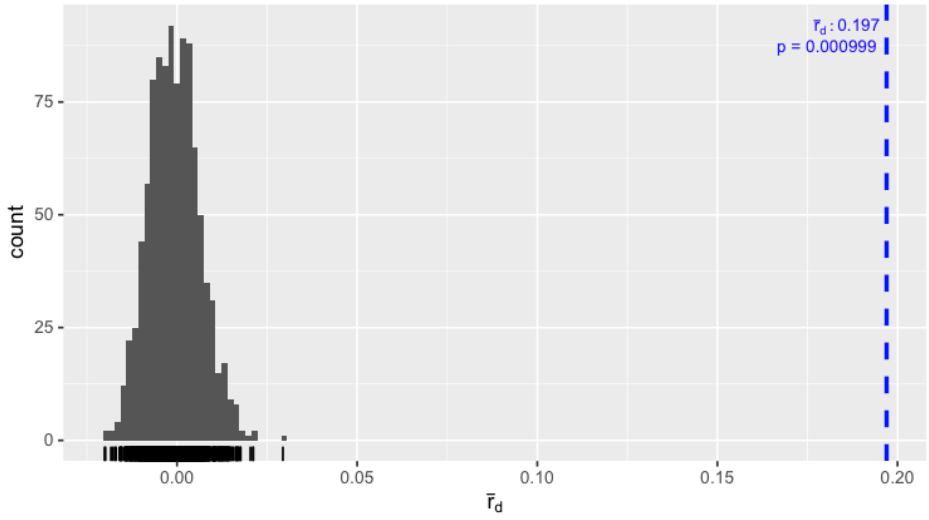



**Figure 6.** Estimation of overall linkage disequilibrium. p is *P*-value from 1000 permutations (<< 0.05), $\bar{r}_d$ is standardized index of association. $\bar{r}_d$ value does not fall into the expected range from permutations

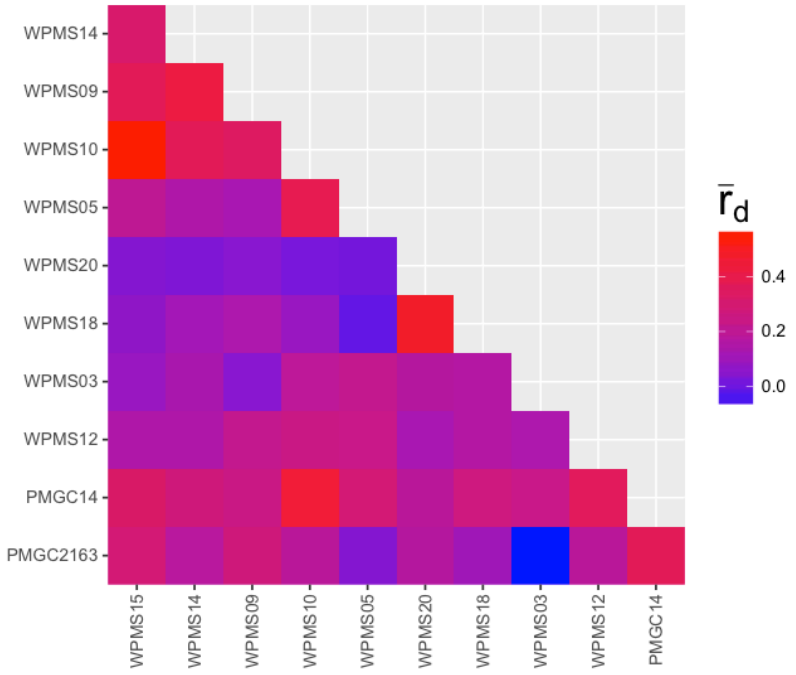



**Figure 7.** Heatmap of pairwise standardized index of association for 11 loci

**Table 5.** Index of association for all geographic populations

| | $\bar{r}_d$ | *P*-value |
|---|---|---|
| **Central** | 0.27 | 0.00 |
| **Eastern** | 0.18 | 0.00 |
| **Aegean** | 0.16 | 0.00 |
| **Blacksea** | 0.24 | 0.00 |
| **Marmara** | 0.08 | 0.02 |

Locations of markers on chromosomes were checked through genetic linkage maps (Gaudet *et al.,* 2007; Paolucci *et al.,* 2010) to see if any physical linkage exists. The loci WPMS14 and WPMS15 are located on chromosome V, WPMS09 and WPMS12 on chromosome VI, WPMS03 and WPMS05 on chromosome XII, PMGC14 and WPMS20 on chromosome XIII. None of the other markers were on the same chromosome and none of the markers were close to the sex determining locus. After excluding one marker from each shared chromosome, there was still overall significant LD.

LD calculations of 3 clusters (explained in "Population Structure" section under "Clustering") from clone corrected structure estimation provided much lower $\bar{r}_d$ values 0.03 ($p = 0.06$), 0.15 ($p = 0.00$) and 0.09 ($p = 0.00$), one of the results being insignificant at $p \leq 0.05$.

## 4. Diversity & Richness

### 4.1. Allelic Richness

There were several private alleles specific to males or females (Table 6). Overall, males had less private alleles (11) than females did (21). These are correlated with sex ratio (34:69). This results in very close private allele frequencies for both genders. None of the private alleles in either gender had frequencies higher than 0.06. Sample size is not enough to significantly detect any relation between alleles and gender. Mean allelic richness value over all loci was very close for males and females (Table 7).

**Table 6.** Frequency of private alleles in each gender

| | | Male | Female | | | Male | Female |
|---|---|---|---|---|---|---|---|
| **Locus** | **Allele (bp)** | **Frequency** | **Frequency** | **Locus** | **Allele (bp)** | **Frequency** | **Frequency** |
| **WPMS15** | 195 | - | 0.02 | **WPMS12** | 152 | - | 0.02 |
| **WPMS14** | 234 | 0.02 | - | | 169 | - | 0.02 |
| | 237 | - | 0.01 | | 177 | - | 0.01 |
| | 252 | - | 0.02 | | 189 | - | 0.01 |
| | 260 | 0.02 | - | **PMGC14** | 195 | 0.02 | - |
| | 261 | - | 0.01 | | 211 | 0.06 | - |
| **WPMS10** | 231 | - | 0.01 | **PMGC2163** | 220 | - | 0.01 |
| | 240 | - | 0.01 | | 238 | 0.03 | - |
| | 246 | 0.02 | - | | 244 | 0.02 | - |
| | 250 | - | 0.01 | | 246 | - | 0.04 |
| | 256 | 0.02 | - | | 254 | 0.03 | - |
| **WPMS05** | 278 | - | 0.06 | | 258 | - | 0.02 |
| | 284 | - | 0.01 | | 263 | - | 0.01 |
| **WPMS20** | 237 | - | 0.01 | | 270 | - | 0.01 |
| **WPMS03** | 266 | - | 0.02 | | | | |
| | 270 | 0.02 | - | | | | |
| | 272 | - | 0.03 | | | | |
| | 280 | 0.03 | - | | | | |

**Table 7.** Allelic Richness

| Locus | Female | Male |
|---|---|---|
| **WPMS15** | 4.75 | 3.99 |
| **WPMS14** | 5.24 | 5.36 |
| **WPMS09** | 4.35 | 4.89 |
| **WPMS10** | 4.88 | 5.02 |
| **WPMS05** | 5.56 | 4.58 |
| **WPMS20** | 5.61 | 5.00 |
| **WPMS18** | 4.35 | 4.90 |
| **WPMS03** | 5.45 | 5.55 |
| **WPMS12** | 5.93 | 3.70 |
| **PMGC14** | 6.99 | 8.99 |
| **PMGC2163** | 7.75 | 8.09 |
| **Mean** | 5.53 | 5.46 |

## 4.2. Genetic Diversity

Overall, there was excess heterozygosity. Only WPMS09, WPMS10 and WPMS05 loci showed excess of homozygosity. After excluding duplicated genotypes, WPMS20 also showed a slight deficit in heterozygosity (Table 9). Though mean expected heterozygosity was improved as expected.

Nei's average gene diversity index was 0.70 for males and 0.66 for females. Males showed higher diversity in 8 loci (Figure 8).

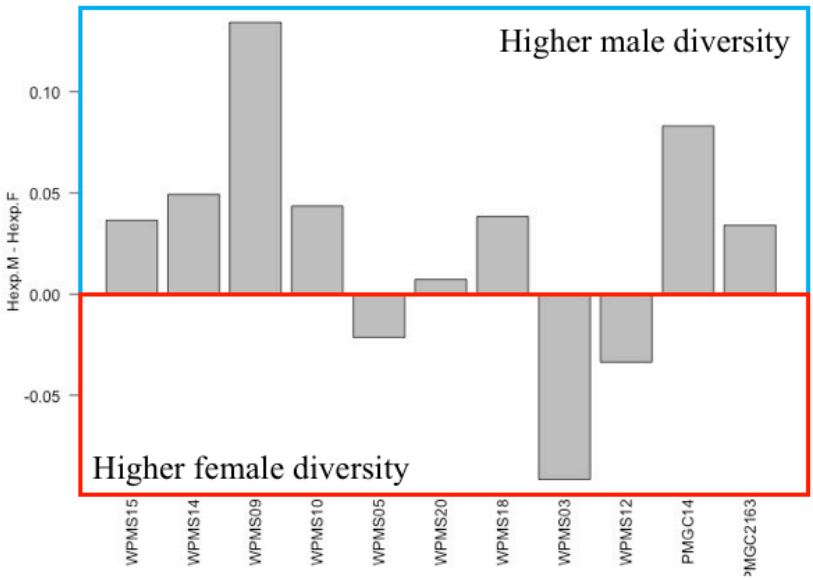



**Figure 8.** Difference between Male $H_{exp}$ and Female $H_{exp}$ over 11 loci

## 5. Population Structure

### 5.1. Clustering

3 clusters were estimated without clone correction by k-means clustering and principal component analysis (PCA) based on suggestions by Jombart *et al*. (2010) (Figure 9). Proposed structure contained identical copies in a single large cluster of 68 individuals and was not correlated with pre-determined 5 populations (Figure 10). Also, no significant explanation (a = 0.05) was found for clusters by one-way ANOVA considering altitude ($p \leq 0.20$), latitude ($p \leq 0.90$) and longitude ($p \leq 0.20$) (Figure 11). Additionally, Chi-Squared test of independence did not provide any significant correlation between gender and clusters ($p \leq 0.44$), though sample size was too small to detect a meaningful relation.

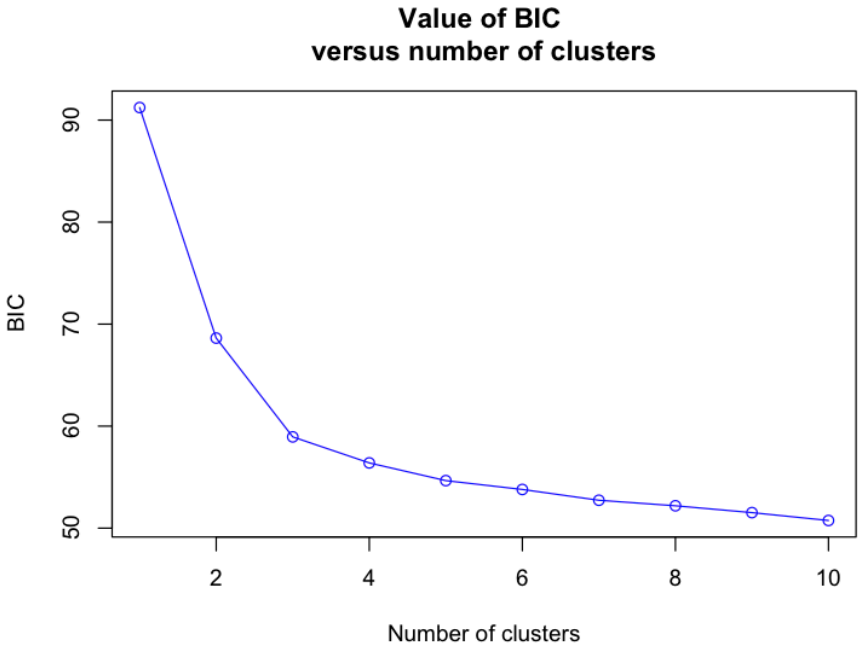



**Figure 9.** Bayesian information criterion (BIC) vs cluster number, steepness decreases after 3 clusters

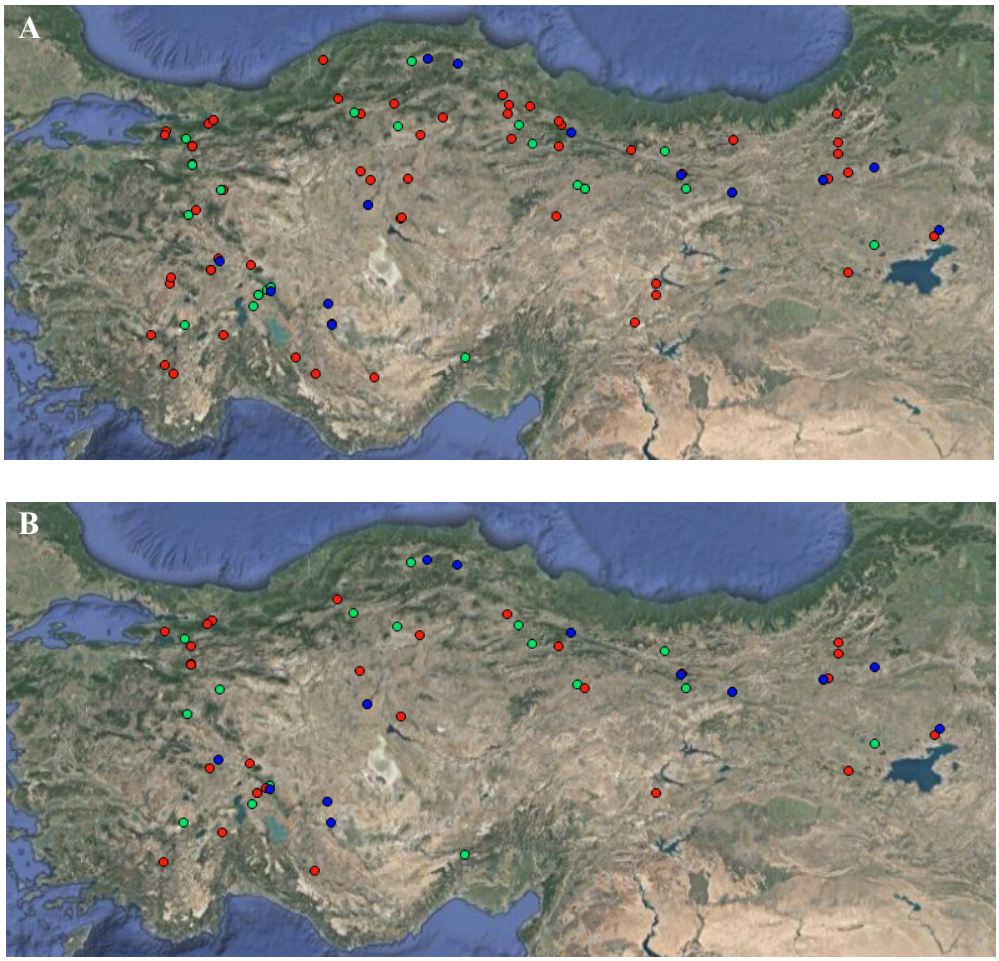



**Figure 10.** Dispersal of individuals in predicted 3 clusters, colors represent different clusters

**A)** Estimation before clone correction

**B)** Estimation after clone correction

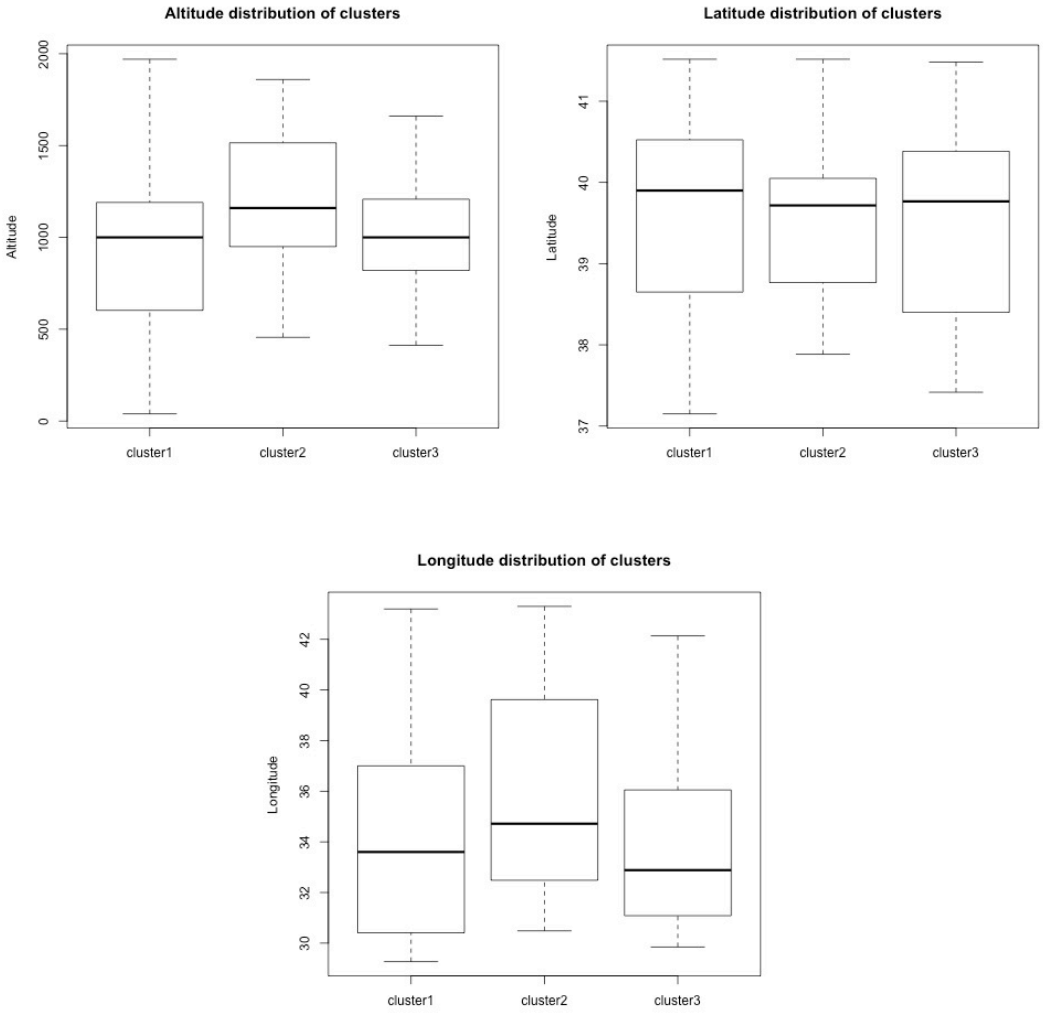



**Figure 11.** Altitude, latitude and longitude distribution of clusters

Although no significant correlation was found between clusters and longitude, Marmara and Aegean regions had lower frequency of clones belonging to Cluster 2 (Table 8).

**Table 8.** Frequency of clones belonging to 3 of the estimated clusters in each geographical population

| | **Cluster1 (N = 68)** | **Cluster2 (N = 22)** | **Cluster3 (N = 13)** |
|---|---|---|---|
| **Central A.** | 0.46 | 0.23 | 0.31 |
| **Eastern A.** | 0.44 | 0.31 | 0.25 |
| **Aegean** | 0.29 | 0.14 | 0.57 |
| **Blacksea** | 0.22 | 0.33 | 0.44 |
| **Marmara** | 0.75 | 0.00 | 0.25 |

After clone correction, 3 possible clusters were estimated (Figure 10). Both suggested structures grouped individuals very similarly, except that identical copies are not included in clone corrected structure. Estimated structure was further described by discriminant analysis of principal components (DAPC) and membership possibilities of all individuals were nearly 100% for each cluster (Figure 12).

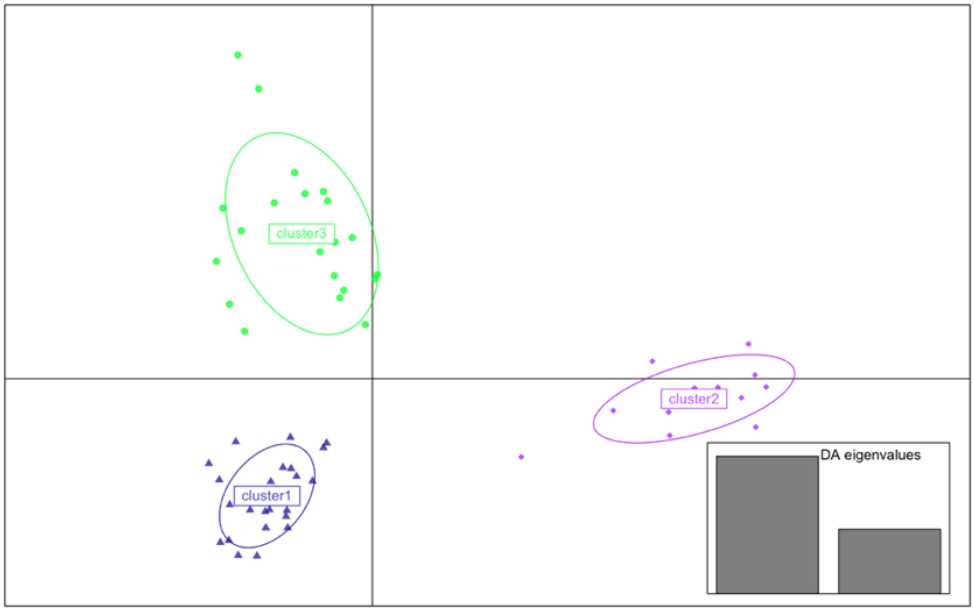



**Figure 12.** DAPC grouping of 3 estimated clusters without identical genotypes

Best K was estimated with highest delta K value after STRUCTURE (Pritchard *et al.*, 2000; Falush *et al.*, 2007; Hubisz *et al.,* 2009) software runs (Figure 13). Software was run without prior population information. Estimated 3 clusters fitted well with both K=2 and K=3 (Figure 14). No structure was detected for pre-determined 5 populations.

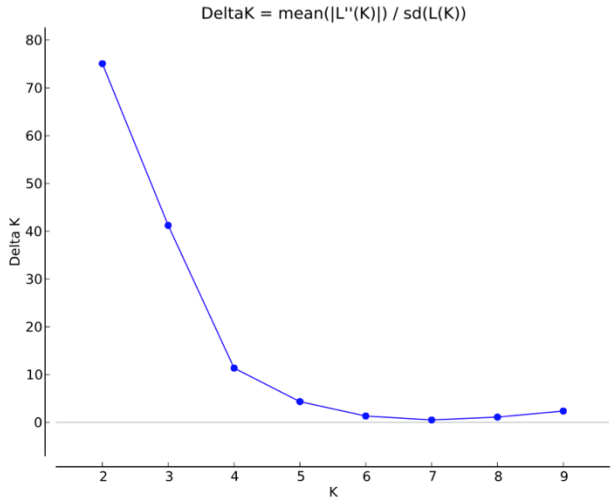



**Figure 13.** Estimated best K = 2 via STRUCTURE HARVESTER (Earl *et al.*, 2012)

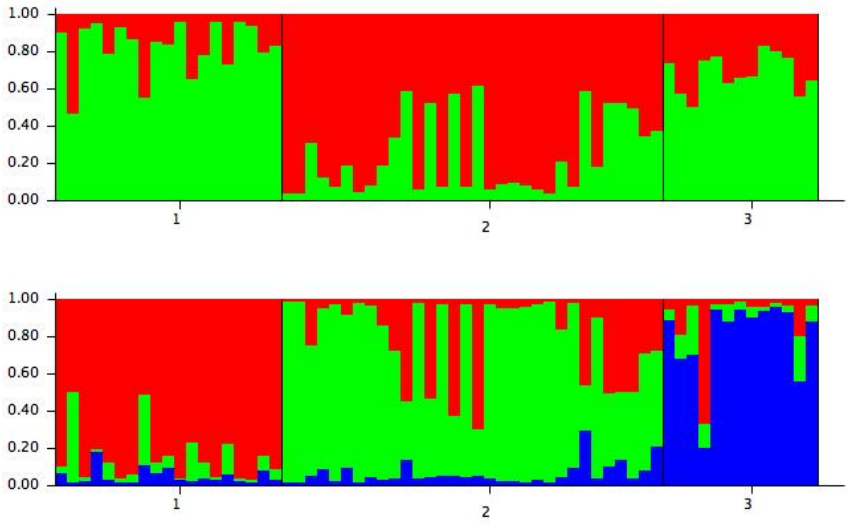



**Figure 14.** STRUCTURE clustering with K = 2 and K = 3, fitting well with estimated 3 PCA clusters

## 5.2. Differentiation Indices

There was no significant genetic differentiation among 5 pre-determined populations, based on $F_{ST}$, $G_{ST}$ and $G'_{ST}$. Significance of values were tested with bootstrap analysis by resampling demes 1000 times, providing a range. After clone correction, all indices of differentiation improved slightly as expected since identical genotypes are dispersed homogenously throughout the whole sampling regions, though, there was still no significant differentiation (Table 9). Negative values can be interpreted as no differentiation and may indicate more variation within population than variation between compared 2 populations (Willing *et al*., 2012).

Calculations of pairwise $G_{ST}$ and $G'_{ST}$ between all combinations of populations did not show significant differentiation between any of two populations (Table 10). Though, highest value was obtained between Aegean and Eastern Anatolia populations.

**Table 9.** Genetic diversity and differentiation estimators - 5 populations

| Locus | Allele # | Effective Allele # | $H_{exp}$ | $H_{obs}$ | $G_{ST}$ | $G'_{ST}$ | $F_{ST}$ | $F_{IS}$ |
|---|---|---|---|---|---|---|---|---|
| **WPMS15** | 5 | 2.70 | 0.68 | 0.93 | -0.01 | -0.06 | 0.00 | -0.36 |
| **WPMS14** | 9 | 3.09 | 0.72 | 0.89 | -0.01 | -0.07 | -0.01 | -0.22 |
| **WPMS09** | 5 | 1.66 | 0.53 | 0.37 | 0.01 | 0.03 | 0,01 | 0.31 |
| **WPMS10** | 9 | 1.33 | 0.35 | 0.21 | -0.01 | -0.02 | -0.03 | 0.42 |
| **WPMS05** | 7 | 1.6 | 0.53 | 0.31 | 0.04 | 0.09 | 0.03 | 0.41 |
| **WPMS20** | 6 | 3.22 | 0.75 | 0.73 | 0.00 | 0.01 | 0.00 | 0.03 |
| **WPMS18** | 5 | 2.61 | 0.66 | 0.96 | -0.02 | -0.06 | 0.00 | -0.44 |
| **WPMS03** | 8 | 2.55 | 0.64 | 0.73 | -0.01 | -0.05 | -0.01 | -0.12 |
| **WPMS12** | 8 | 2.59 | 0.67 | 0.90 | -0.02 | -0.07 | -0.01 | -0.33 |
| **PMGC14** | 10 | 3.30 | 0.77 | 0.99 | 0.00 | -0.02 | 0.00 | -0.28 |
| **PMGC2163** | 14 | 2.97 | 0.74 | 0.89 | -0.01 | -0.05 | -0.01 | -0.20 |
| **Overall** | 7.97 | 2.53 | 0.64 | 0.72 | -0.01 | -0.02 | 0.00 | -0.11 |

**Table 10.** Pairwise $G'_{ST}$ calculations

| | Central | Eastern | Aegean | Blacksea |
|---|---|---|---|---|
| **Central** | 0.00 | | | |
| **Eastern** | -0.04 | | | |
| **Aegean** | -0.03 | 0.02 | | |
| **Blacksea** | -0.03 | 0.01 | -0.02 | |
| **Marmara** | -0.05 | -0.01 | 0.00 | -0.01 |

When dataset was rearranged as 2 populations (males and females), genetic differentiation estimators were higher, but again it did not provide significant structure (Table 11). After only highest scoring loci WPMS14, WPMS09, WPMS03 and PMGC14 were selected, global estimates produced slight, but significant $G_{ST}$ and $G'_{ST}$ values through bootstrapping ($G_{ST}$ = 0.0104 and between 0.003 – 0.018, $G'_{ST}$ = 0.0667 and between 0.020 and 0.113 with 95% confidence). These results may be an indication of different allelic compositions between genders on those loci.

**Table 11.** Genetic differentiation estimators - 2 populations (Male and Female)

| Locus | $G_{ST}$ | $G'_{ST}$ | $F_{ST}$ |
|---|---|---|---|
| **WPMS15** | 0.00 | 0.03 | 0.01 |
| **WPMS14** | 0.01 | 0.10 | 0.03 |
| **WPMS09** | 0.01 | 0.04 | 0.01 |
| **WPMS10** | 0.00 | -0.02 | -0.02 |
| **WPMS05** | 0.00 | -0.01 | -0.01 |
| **WPMS20** | -0.01 | -0.06 | -0.01 |
| **WPMS18** | 0.00 | 0.00 | 0.01 |
| **WPMS03** | 0.01 | 0.08 | 0.03 |
| **WPMS12** | 0.00 | -0.03 | 0.00 |
| **PMGC14** | 0.01 | 0.07 | 0.02 |
| **PMGC2163** | 0.00 | 0.00 | 0.00 |
| **Overall** | 0.00 | 0.02 | 0.01 |

Estimated 3 clustered structure showed differentiation with $G_{ST}$ = 0.13, $G'_{ST}$ = 0.42 and $F_{ST}$ = 0.21 and all values confirmed to be positive with 95% confidence interval via bootstrapping. Clone corrected structure also provided similar results with $G_{ST}$ = 0.10, $G'_{ST}$ = 0.37 and $F_{ST}$ = 0.15 (Table 12).

**Table 12.** Genetic differentiation estimators for estimated 3 clusters

| Locus | $G_{ST}$ | $G'_{ST}$ | $F_{ST}$ |
|---|---|---|---|
| **WPMS15** | 0.05 | 0.2 | 0.09 |
| **WPMS14** | 0.07 | 0.33 | 0.12 |
| **WPMS09** | 0.16 | 0.45 | 0.26 |
| **WPMS10** | 0.25 | 0.47 | 0.34 |
| **WPMS05** | 0.09 | 0.28 | 0.15 |
| **WPMS20** | 0.12 | 0.55 | 0.15 |
| **WPMS18** | 0.08 | 0.33 | 0.12 |
| **WPMS03** | 0.10 | 0.36 | 0.13 |
| **WPMS12** | 0.13 | 0.46 | 0.17 |
| **PMGC14** | 0.07 | 0.41 | 0.11 |
| **PMGC2163** | 0.07 | 0.35 | 0.10 |
| **Overall** | 0.10 | 0.37 | 0.15 |

### 5.3. Analysis of Molecular Variance

There was no significant variation between pre-defined regional populations. All estimated variance was associated with differences within populations. AMOVA tests provided a significant variation between estimated clusters (Figure 15). Variation between clusters was responsible for 48% of total variation. After excluding duplicates, there was still significant variation, which was 30% of total (Table 13).

Also different hypothetical hierarchical levels were tested. Between genders within cluster, between regions within cluster and between genders within region results did not yield any significant variation.

No significant variation was detected between genders. When only loci with highest differentiation values (WPMS14, WPMS09, WPMS03 and PMGC14) were selected a significant variation was observed with p = 0,007. 5.12% of total variation was attributed to variation between genders (Table 14). Also, in hierarchical analysis, variation between regions within gender was found to be low (4% of total variation). Additionally, creating this type of subgroups resulted in very small sample sizes which is not enough for decisive conclusions.

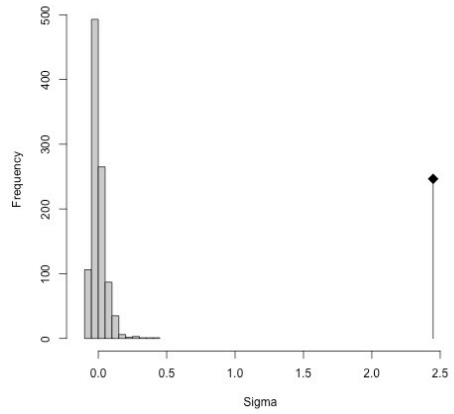



**Figure 15.** Histogram of significance. Expected range produced by permutations. Simulated p-value ≈ 0

**Table 13.a.** AMOVA results for 3 cluster structure

| | **Sigma** | **%** |
|---|---|---|
| **Between clusters** | 2.45 | 48.32 |
| **Within cluster** | 2.62 | 51.68 |
| **Total variations** | 5.07 | 100.00 |

**Table 13.b.** AMOVA results for 3 cluster structure (clone corrected)

| | **Sigma** | **%** |
|---|---|---|
| **Between clusters** | 1.68 | 29.77 |
| **Within cluster** | 3.97 | 70.23 |
| **Total variations** | 5.65 | 100.00 |

**Table 14.** AMOVA results for gender comparison for selected 4 loci

| | Sigma | % |
|---|---|---|
| **Between genders** | 0.11 | 5.12 |
| **Within gender** | 2.07 | 94.88 |
| **Total variations** | 2.18 | 100.00 |

## 6. Correlation Analysis

Fisher's exact test and chi-squared statistics did not provide significant correlation between gender and region ($p \leq 0.23 – 0.24$), though sample size was not enough for a reliable result. There was no significant correlation between gender and altitude ($p \leq 0.13$), gender and latitude ($p \leq 0.07$), gender and longitude ($p \leq 0.90$) based on one-way ANOVA tests. Region and altitude showed correlation as expected (Figure 16).

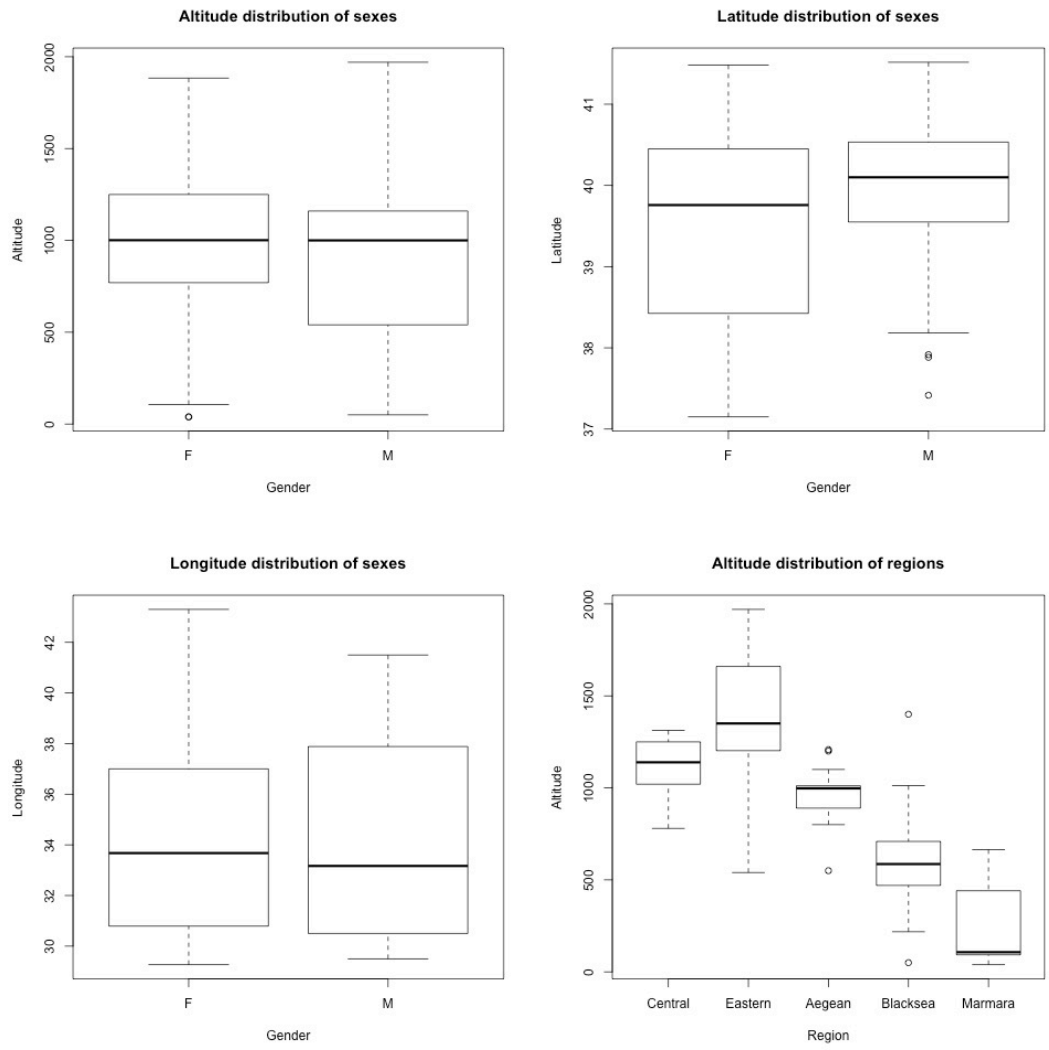
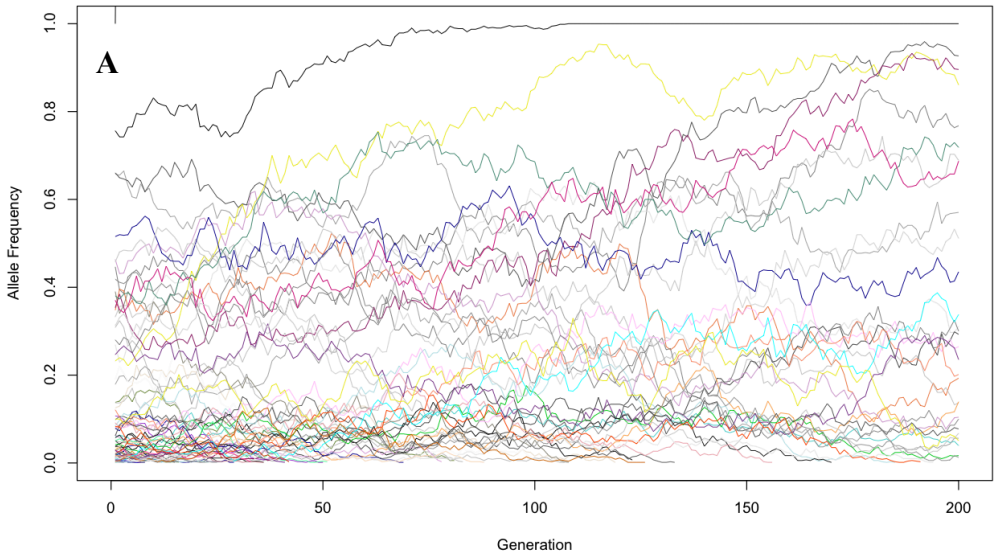



**Figure 16.** Boxplots of different variables as explained in headings

## 7. Simulations for Future Generations

Simulations of successive generations based on allelic and migration data were performed with different combinations of conditions several times. Clone numbers of populations were kept as constant in all simulations. To reduce the effect of drift, individual number of first progeny of each population was increased by four fold. Equal vs unequal sex ratio cases were examined under different gene flow (migration) rates.

With estimated migration rates between populations and a sex ratio of 3:7 which continued over generations, alleles got converged to fixation faster and range of allele frequencies broadened (Figure 17). Also, more alleles with initial low frequency (< 0.05) disappeared compared to the undisturbed 1:1 ratio (Figure 18), resulting in a sharper decline of genetic diversity (Figure 19). Mean number of lost alleles after 200 generations over 20 iterations was 41.7 for continuous 1:1 ratio scenario and 54.4 for continuous 3:7 biased ratio scenario. Genetic differentiation indices $G_{ST}$ and $F_{ST}$ did not change over generations, an expected result as there was no initial significant differentiation.



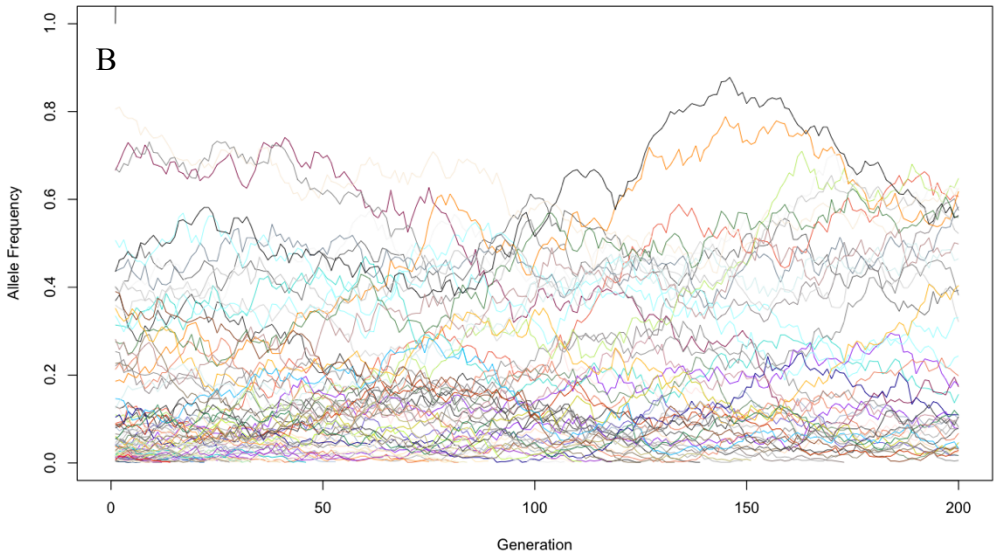
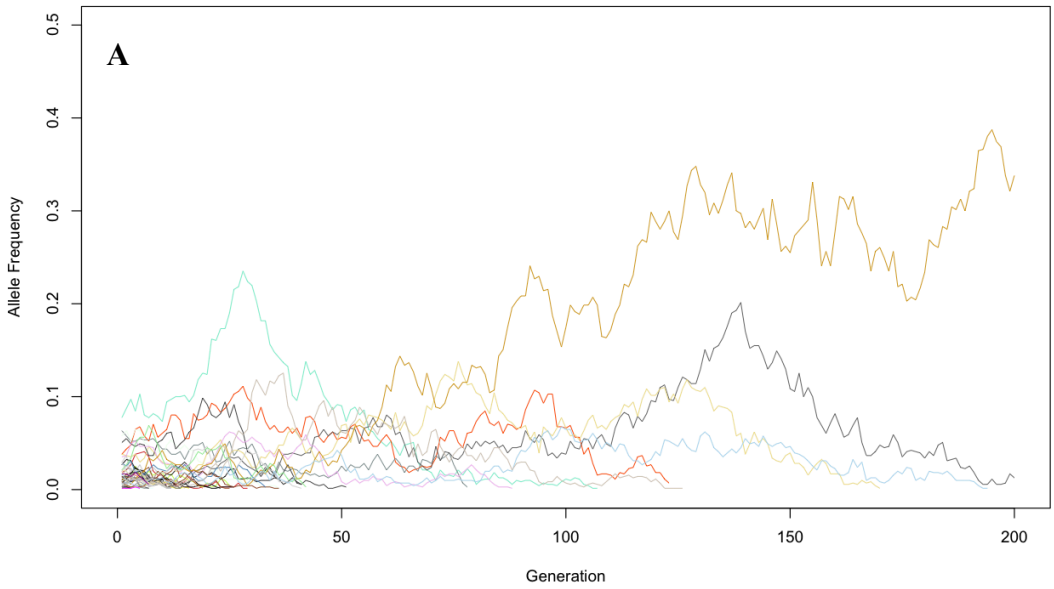




**Figure 17. A)** Allele frequencies over generations with 3:7 sex ratio. More alleles are closer to fixation. One allele in WPMS10 with initial 0.81 frequency gets fixed for whole populations. Average fixation rate over 20 iterations was 0.41

**B)** Allele frequencies over generations with 1:1 sex ratio. Average fixation rate over 20 iterations was 0.05



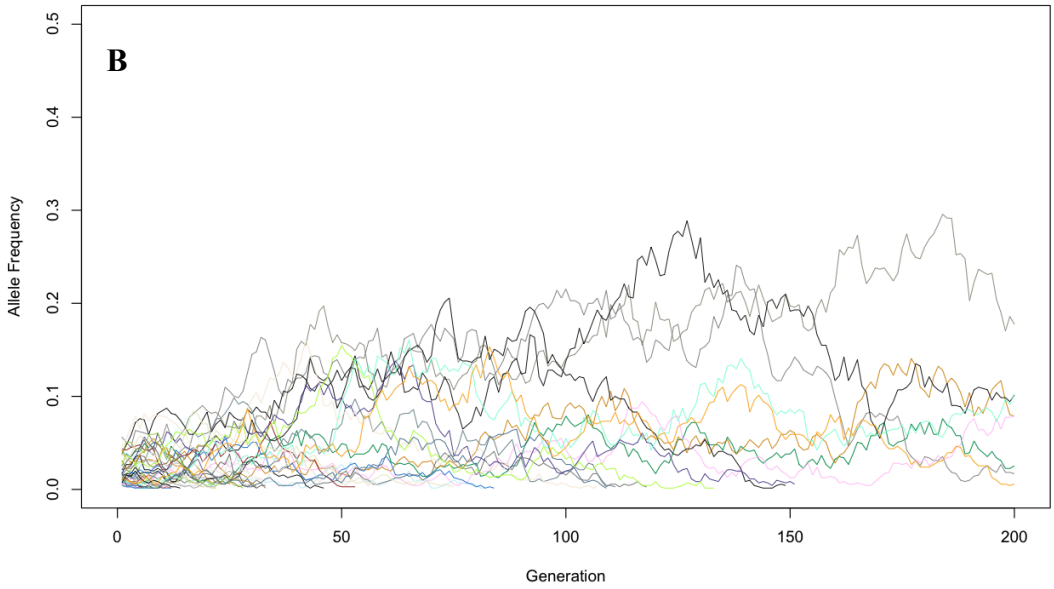
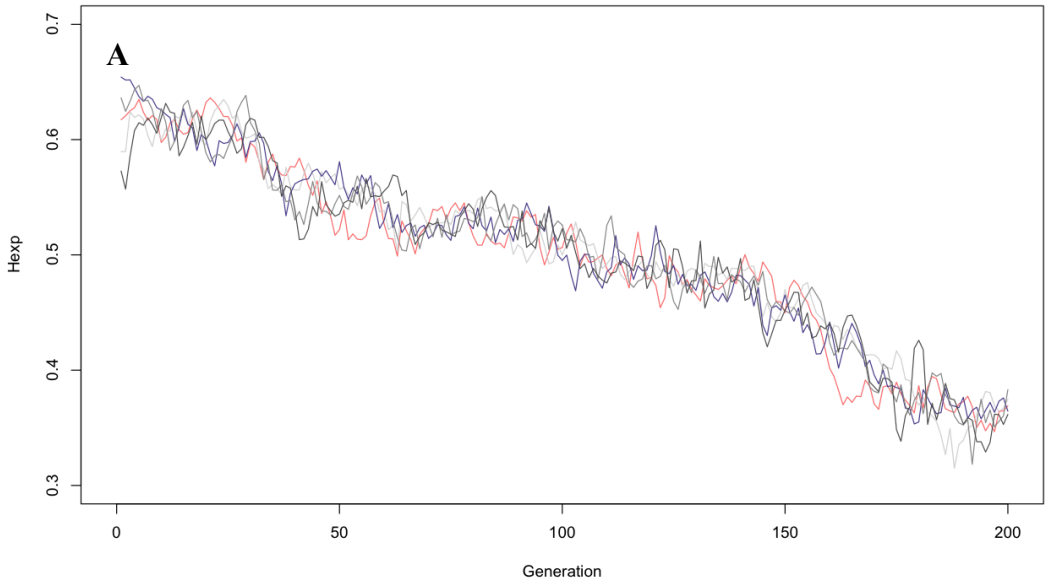




**Figure 18. A)** Alleles with initial frequencies < 0.05, 3:7 sex ratio. Only 2 alleles remained (survived) after 200 generations. Mean survival rate after 20 iterations was 2.4

**B)** Alleles with initial frequencies < 0.05, 1:1 sex ratio. Mean survival rate after 20 iterations was 9.3



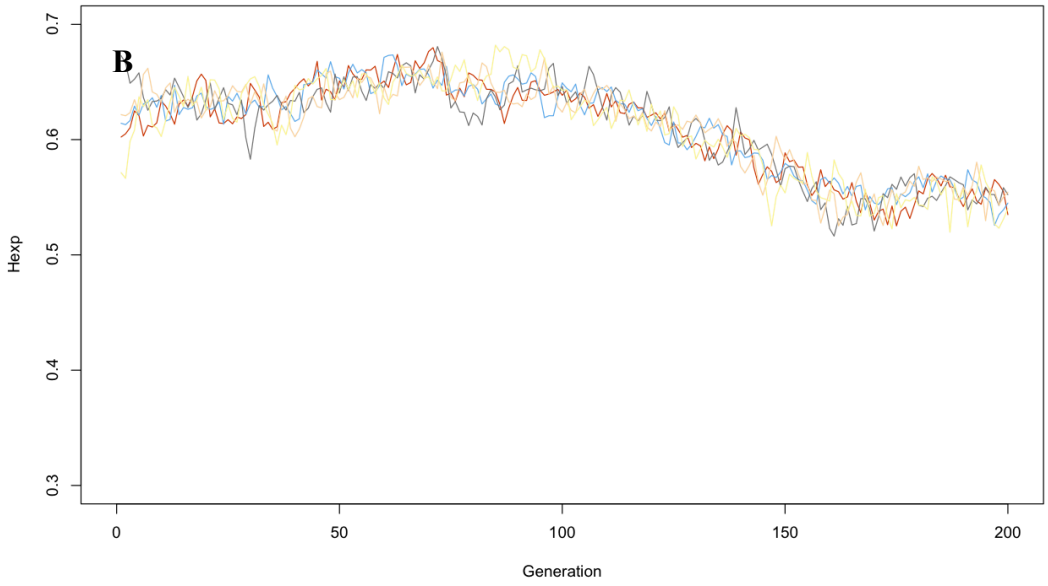




**Figure 19. A)** Expected heterozygosity of 5 populations, 3:7 sex ratio. Patterns follow the overall $H_{exp}$ due to gene flow

**B)** Expected heterozygosity of 5 populations, 1:1 sex ratio

Simulations were also performed under complete isolation starting from 2nd generation. Without the gene flow, effects of biased ratio on diversity could be seen more clearly. Same initial 5 populations were tested under assumptions of 1:1, 3:7, 1:4, 1:9 and 1:20 sex ratios. Higher bias resulted in higher differentiation and lower diversity (Figure 20).

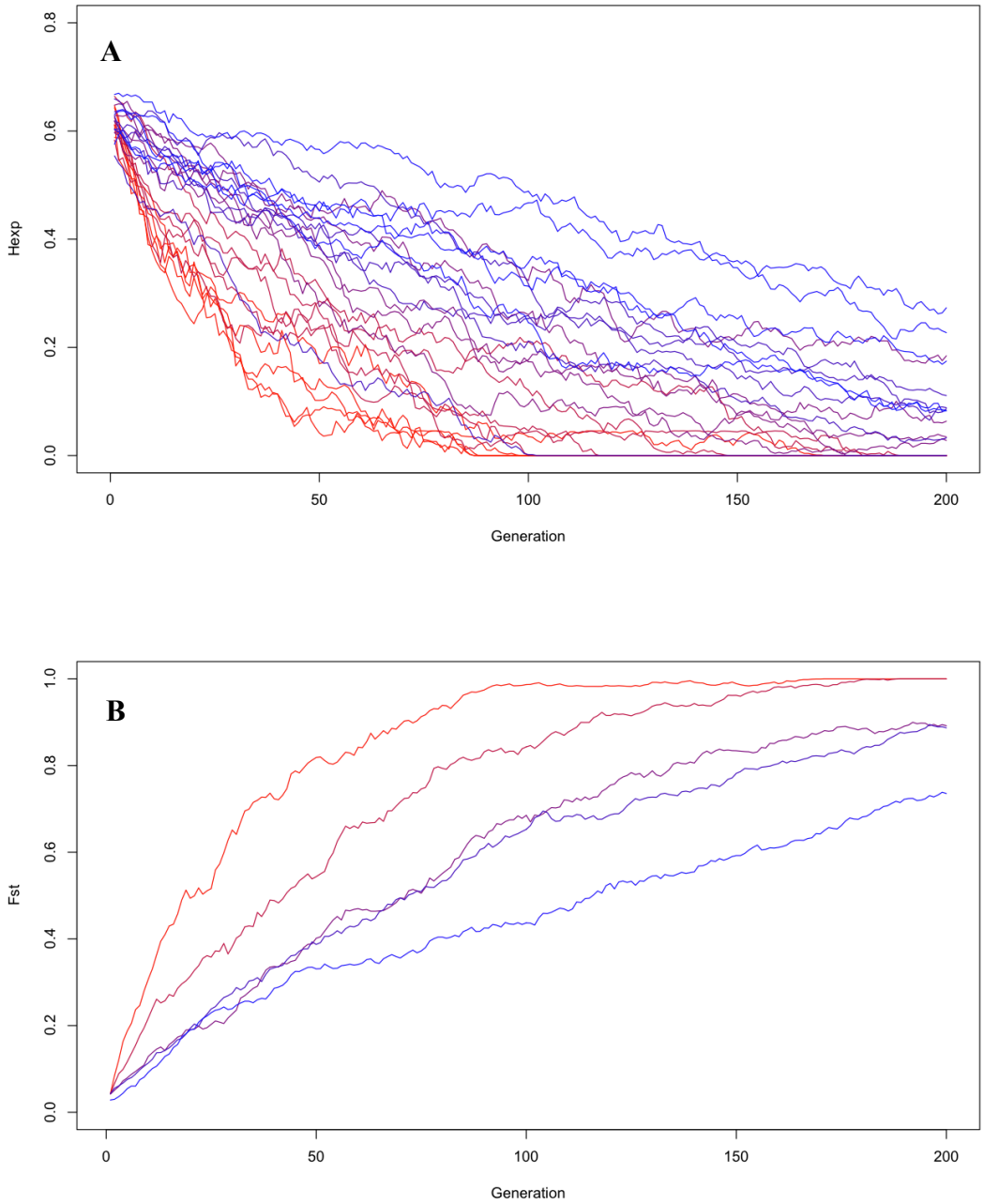



**Figure 20. A)** $H_{exp}$ of 5 populations over generations without gene flow. Sex ratios from blue to red, 1:1 – 3:7 – 1:4 – 1:9 – 1:20

**B)** $F_{ST}$ over generations when there is no gene flow. Sex ratios from blue to red, 1:1 – 3:7 – 1:4 – 1:9 – 1:20

# CHAPTER 5

# DISCUSSION

## 1. Data Quality & Clone Issue

There were still four duplicated genotypes after manual exclusion of clones having close coordinates and same genotypes. One genotype was observed in 32 clones which were coded as different clones. 22 of these clones were female and 10 were male. Another duplicated genotype was observed in 4 females and 2 males (all were coded as different clones). *P. nigra* is a dioecious plant, but Novotna & Stochlova (2013) reported finding six monoecious trees in trial plots. Controlled pollinations in their study produced viable seeds. Trees were able to produce offspring as a male, female and by self-fertilization. Another study found a similar occurrence in *Populus deltoides,* that trees produce male and subsequently female flowers (Rowland *et al.,* 2002). Although sex variation in genus *Populus* seems to occur rarely*,* these previous findings suggest that the the same genotype observed in this study might have both male and female flowers which resulted in misidentification of genders of some clones.

Another explanation may lie in the resolution of current markers. Microsatellites generally provide reliable results for individual discrimination due to their highly variable nature (Gerber *et al.,* 2000). However, despite overall high heterozygosity, certain allele combinations were very common. Alleles in WPMS09, WPMS10 and WPMS05 are very close to fixation. Thus their discriminative power is reduced. Also, WPMS04 locus was estimated to have high null allele frequency (~44%). Null alleles

can be caused by poor primer annealing related to mutations in nucleotide sequence in flanking regions (since SSR markers are detected by PCR), alleles with higher length and poor DNA template (Dakin, 2004). Since null allele estimations are based on deviations from HWE, it is possible to include false null alleles due to phenomenon like inbreeding (Chakraborty *et. al* 1992). Most common two duplicated genotypes in the dataset only differ in this locus and both are homozygous, increasing the possibility of a null allele.

There was no correlation between distribution of duplicated genotypes and any of the other available data. Homogenous and wide spatial dispersal of this genotype is most probably related to human activity.

After extraction of duplicated genotypes, remaining number was not enough to accurately estimate allele compositions of each population (Hale *et al.,* 2012). For further analysis, more appropriate sampling techniques should be applied.

**2. Linkage & Selective Neutrality**

Population structure, asexual reproduction, linkage between alleles, selection and genetic drift can all produce LD on a given locus (Agapow & Burt, 2001). Marker loci must be assessed for LD since if presence of significant LD on a marker is related to physical linkage or selection, that marker can not provide reliable information about population structure and allelic compositions (Selkoe & Toonen, 2006). LD found in this study was not specific to a few markers. Most loci pairs showed significant association. One interpretation might be existence of a population structure. LD values were substantially lower for structures estimated by k-means clustering and STRUCTURE software (Pritchard *et al.*, 2000; Falush *et al.*, 2007; Hubisz *et al.,* 2009). It is important to note that methods used by STRUCTURE

software estimates clusters by minimizing LD between loci and Hardy-Weinberg disequilibrium among individuals. At this point, it is not easy to determine whether the significant LD values were obtained because of a structure or a pseudo-structure is found because of significant LD related to another phenomenon like selection or physical linkage.

An important assumption made for microsatellites is selective neutrality. Even though generally microsatellites are considered to be neutral, they can be seen within protein coding regions (Li, 2004). Also, mutations in microsatellites can be directed by external factors (Schmidt & Mitter, 2004). For this study, presence of outlier loci which might be under selection was checked by BayeScan software (Foll & Gaggiotti, 2008; Foll *et al.,* 2010; Fischer *et al.,* 2011). No outlier locus was detected, but findings are not included in the results since both sample size and number of markers of the dataset were very low to detect candidate loci. There was no study in the literature regarding neutrality of the markers in question. Therefore, it is an essential point to assess in future studies.

### 3. Population Structure

Estimations of population structure without prior knowledge suggested 2 or 3 clusters. The clusters were not associated with region, altitude and gender. Clusters estimated without clone correction placed all duplicated genotypes in the same cluster. Estimations with clone correction revealed that suggested clusters are the same except for duplicated genotypes. Additionally, high frequency alleles were mainly appointed to one single cluster. These findings may point to a single descent for this cluster, possibly spread artificially by humans. This possibility becomes stronger when commercially available trees are examined. Most common duplicated genotype (observed in 32 clones) differs only by three alleles from commercial Anadolu tree (Table 15). However, these clusters might also be falsely identified and

occur completely due to linkage between alleles (as discussed in "Data Quality & Clone Issue" section).

**Table 15.** Allele combinations

*2nd column: Most common combinations in whole dataset after excluding duplicated clones*

*3rd – 7th columns: Commercially available trees*

*8th -9th columns: Most abundant clones*

| Locus | Com | Anadolu | Kocabey | Gazi | Ata | Geyve | Clone 1 | Clone 2 |
|---|---|---|---|---|---|---|---|---|
| **W15** | 204/210 | 204/213 | 204/210 | 204/213 | 204/210 | 204/213 | 204/210 | 204/210 |
| **W14** | 210/243 | 243/252 | 243/252 | 243/252 | 243/252 | 243/252 | 210/243 | 210/243 |
| **W09** | 246/246 | 246/255 | 246/255 | 246/255 | 246/255 | 246/255 | 246/246 | 246/246 |
| **W10** | 248/248 | 248/248 | 248/248 | 248/248 | 248/248 | 248/248 | 248/248 | 248/248 |
| **W05** | 276/276 | 276/276 | 276/276 | 276/276 | 276/276 | 276/276 | 276/276 | 276/276 |
| **W04** | 274/274 | 274/274 | 258/274 | 278/278 | 258/274 | 260/274 | 274/274 | 250/250 |
| **W20** | 211/225 | 211/225 | 211/225 | 231/231 | 225/225 | 225/225 | 211/225 | 211/225 |
| **W18** | 222/235 | 222/235 | 222/235 | 218/235 | 218/235 | 218/235 | 222/235 | 222/235 |
| **W03** | 268/278 | 268/278 | 264/280 | 264/280 | 264/280 | 264/280 | 268/278 | 268/278 |
| **W12** | 165/174 | 165/174 | 165/165 | 167/174 | 165/165 | 165/167 | 165/174 | 165/174 |
| **P14** | 191/198 | 191/198 | 198/224 | 198/217 | 198/224 | 198/224 | 191/198 | 191/198 |
| **P2163** | 224/242 | 224/242 | 220/224 | ?/? | 224/224 | 220/224 | 224/242 | 224/242 |

## 4. Genetic Diversity & Variance

Observed and expected heterozygosity along with genetic differentiation estimations were very close to the findings from the main dataset in the recent study by Çiftçi *et al.* (2017). This is an indication that the subset of data used in this work is probably a good estimator of the whole dataset.

Allelic richness was nearly identical for males and females, but both genders had private alleles on 9 loci (1 locus excluded due to possibility of null allele). Private alleles had low frequencies. Therefore, it was not possible to find out a meaningful relation. These loci must be analyzed with higher sample sizes in future studies.

Expected heterozygosity over all loci was higher in males compared to females. In 8 loci, $H_{exp}$ was higher in males. Although this might indicate higher genetic diversity in males, it is important to note that all studied loci have alleles with very high frequencies and sample size is not very large. For a more reliable comparison of genetic diversity between genders, higher sample size is required to be able to account for low frequency alleles and better estimation of evenness.

Significant differentiation and estimated variance between genders for WPMS14, WPMS09, WPMS03 and PMGC14 loci might be an indication of different allelic compositions between genders for certain loci. In a recent study, sex separation loci were identified by high $F_{ST}$ values for those loci between genders (Jia *et al.,* 2015). Although $F_{ST}$ estimations were not too high in this study, these 4 loci might be examined for sex associated marker studies in the future.

For future analysis, also differentiation between male populations and female populations from the same location can be investigated for assessing direct effects of different genders on population structure. In this study, a hierarchical AMOVA was performed in a similar manner, but such subdivisions resulted in very low sample sizes. Thus, the results can not be considered as meaningful.

WPMS04 had the highest values for differentiation estimators both for among populations and genders. This finding further supports the existence of a null allele in that locus.

## 5. Biased Sex Ratio

Overall there was a female biased sex ratio with and without excluding duplicated genotypes (clones). Estimated clusters and regions also showed the same bias, except for Marmara region which had only 11 individuals (after clone correction). In this study, no significant relation to explain disturbed sex ratio was found with the available geographical data. Additionally, it was not possible to identify any possible gender specific alleles due to sample size.

## 6. Possible Effects of Biased Sex Ratio in Subsequent Generations

Performed simulations showed that bias in sex ratio, increases differentiation rate in case of isolation and reduces genetic diversity in the subsequent generations. Even with the estimated high migration rates which shuffle genetic resources frequently, it was observed that alleles can be fixed for the whole populations in the dataset. Especially WPMS09, WPMS10, WPMS05 and WPMS03 loci had alleles with frequencies > 0.5 and get fixed frequently in the simulations. This might be a sign that current high frequency of certain alleles might be related to the biased sex ratio.

*P. nigra* was listed as one of the priority species to be conserved in Turkey (Kaya *et al.,* 1997). Findings in this study suggest that genetic diversity of this species might be significantly effected by biased sex ratio in future generations. Therefore, further detailed investigations on the sex variation of European black poplars should be conducted. These studies could involve additional information like wetness and salinity. Also, experimental confirmation via morphological and molecular assessment on whether both male and female flowers occur on a single tree is essential. Sampling over seasons and years can be useful to detect any possible alternating sex variation. Another suggestion for future studies could be including

sex locus when selecting markers and increasing resolution with more loci to be able to differentiate individuals more accurately.

# CHAPTER 6

# CONCLUSION

Genotypic analysis showed that the most common genotypic duplicate occurs in both male and females. This might be an indication of deviation from dioecism in some of the *Populus nigra* species in Turkey. Also, high frequency alleles observed in all studied loci show that the resolution of these loci might not be enough to discriminate individuals. This might consequently lead to misrecognition of different individuals as genotypic copies.

There was no differentiation among 5 pre-assigned geographical populations. 3 clusters were found without prior information, indicative of an underlying genetic structure. Although no direct correlation was found, these estimated clusters may be related to 3 genetic source of black poplars dispersed throughout the regions by human activity. Estimated significant linkage disequilibrium might be related to this structuring. Though, the loci should also be studied further for selective neutrality.

Allelic richness and diversity measures were close for male and females. However, each gender had private alleles on 9 loci. All the found private alleles had low frequencies. Therefore, further investigation with more sample size is required to detect any loci possibly involved in sex determination. Additionally, WPMS14, WPMS09, WPMS03 and PMGC14 loci can be studied in the future, since combination of these loci showed a slight but significant empirical differentiation between male and females.

No explanation was found for the female biased sex ratio with the available geographical and molecular data. Subsequent studies should involve natural populations and molecular gender detection methods along with other variables like wetness and age.

Due to low sample sizes and no significant differentiation among geographical populations, direct contribution of genders to differentiation could not be assessed. However, by the help of a developed simulation software prototype, it was shown that biased sex ratio can reduce genetic diversity and increase differentiation in subsequent generations. This might also partially explain observed high frequency alleles if the biased sex ratio is common in natural populations.

Developed simulation software prototype can be used in studies to assess future direction of population structure given the molecular and migration data. Additionally, it can further be improved for more reliable predictions. For instance, mutation models like stepwise mutation model can be integrated.

## REFERENCES:


Adamack, A.T. & Gruber, B. (2014). PopGenReport: simplifying basic population genetic analyses in R. Methods in Ecology and Evolution, 5(4), 384-387.

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

Agresti, A. (2007). An Introduction to Categorical Data Analysis, 2nd ed., New York: John Wiley & Sons. Page 38.

Arens, P., Coops, H., Jansen, J., & Vosman, B. (1998). Molecular genetic analysis of black poplar (*Populus nigra L*). along Dutch rivers. Molecular Ecology,7(1), 11-18.

Barrett, S.C., Yakimowski, S. B., Field, D. L., & Pickup, M. (2010). Ecological genetics of sex ratios in plant populations. Philosophical Transactions of the Royal Society B: Biological Sciences,365(1552), 2549-2557.

Beerli, P. (2009). How to use migrate or why are markov chain monte carlo programs difficult to use? In G. Bertorelle, M. W. Bruford, H. C. Hauffe, A. Rizzoli, and C. Vernesi, editors, Population Genetics for Animal Conservation, volume 17 of Conservation Biology, pages 42–79. Cambridge University Press, Cambridge UK, 2009.

Beerli, P. (2006). Comparison of Bayesian and maximum likelihood inference of population genetic parameters. Bioinformatics, 22(3):341–345.

Bird, C., Karl, S., Mouse, P., & Toonen, R. (2011). Detecting and measuring genetic differentiation. Crustacean Issues Phylogeography and Population Genetics in Crustacea, 31-55.

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

Brown, A.H.D., Feldman, M.W., & Nevo, E. (1980). Multilocus structure of natural populations of *Hordeum spontaneum*. Genetics, 96, 523– 536.

Caballero, A. (1994). Developments in the prediction of effective population size. Heredity *73*(6), 657-679.

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

Chambers, J. M., Freeny, A, & Heiberger, R. M. (1992). Analysis of variance; designed experiments. Chapter 5 of Statistical Models in S eds J. M. Chambers and T. J. Hastie, Wadsworth & Brooks/Cole.

Ciftci, A., Karatay, H., Kücükosmanoğlu, F., Karahan, A., & Kaya, Z. (2017). Genetic differentiation between clone collections and natural populations of European black poplar (*Populus nigra L.*) in turkey. Tree Genetics & Genomes,13(3).

Dakin, E. E., & Avise, J. C. (2004). Microsatellite null alleles in parentage analysis. Heredity,93(5), 504-509.

Delph, L. F. (1999). Sexual dimorphism in life history. In Gender and sexual dimorphism in flowering plants (eds M. A. Geber, T. E. Dawson & L. F. Delph), pp. 152–155. Berlin, Germany: Springer.

Earl, D. A., & Vonholdt, B. M. (2011). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources,4(2), 359-361.

Excoffier, L., Smouse, P. E. and Quattro, J. M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics, 131, 479-491.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. Molecular Ecology,14(8), 2611-2620.

Falush D., Stephens M., Pritchard J.K. (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol Ecol Notes 7(4):574–578

Fischer, M.C., Foll, M., Excoffier, L., & Heckel, D. (2011). Enhanced AFLP genome scans detect local adaptation in high-altitude populations of a small rodent (*Microtus arvalis*). Molecular Ecology 20: 1450-1462

Foll, M. & Gaggiotti, O.E. (2008). A genome scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. Genetics 180: 977-993

Foll, M., Fischer, M.C., Heckel, G. & Excoffier, L. (2010). Estimating population structure from AFLP amplification intensity. Molecular Ecology 19: 4638-4647

Frankham, R., Ballou, J. D., & Briscoe, D. A. (2015). Introduction to conservation genetics. Cambridge: Univ. Press.

Gaudet, M., Jorge, V., Paolucci, I., Beritognolo, I., Mugnozza, G. S., & Sabatti, M. (2007). Genetic linkage maps of Populus nigra L. including AFLPs, SSRs, SNPs, and sex trait. Tree Genetics & Genomes,4(1), 25-36.

Gerber, S., Mariette, S., Streiff, R., Bodenes, C. & Kremer, A. (2000). Comparison of microsatellites and amplified fragment length polymorphism markers for parentage analysis. Mol. Ecol., 9, 1037–1048.

Goudet, J. (2005). Hierfstat, a package for r to compute and test hierarchical F-statistics. Molecular Ecology Notes,5(1), 184-186.

Gruber, B. & Adamack, A.T. (2015). Landgenreport: a new R function to simplify landscape genetic analysis using resistance surface layers. Molecular Ecology Resources, 15(5), 1172-1178.

Gillespie, J. H. (2004). Population genetics: a concise guide. Baltimore, Md: The Johns Hopkins University Press.

Google Earth. (2008). Turkey 39°00'N, 35°00'E. Viewed 05 December 2016. http://www.google.com/earth/index.html.

Hale, M. L., Burg, T. M., & Steeves, T. E. (2012). Sampling for microsatellite-based population genetic studies: 25 to 30 individuals per population is enough to accurately estimate allele frequencies. PloS one, 7(9), e45170.

Hedrick, P.W. (1999). Highly variable loci and their interpretation in evolution and conservation. Evolution 53: 313–318.

Hedrick, P.W. (2005). A standardized genetic differentiation measure. Evolution 59: 1633–1638.

Herpka I. (1986). A survey of development and possibilities of growing: natural forests of poplars and willows. Poplars and Willows in Yugoslavia, Poplar Research Institute, Novi Sad, pp. 21-36.

Hope, A. C. A. (1968). A simplified Monte Carlo significance test procedure. J. Roy, Statist. Soc. B 30, 582–598.

Hubisz, J.M., Falush, D., Stephens, M., Pritchard, J.K. (2009). Inferring weak population structure with the assistance of sample group information. Mol Ecol Resour 9(5):1322–1332

Jia, H., Jiao, Y., Wang, G., Li, Y., Jia, H., Wu, H., Chai, C., Dong, X., Guo, Y., Zhang, L., Gao, Q., Chen, W., Song, L., De Weg, E. V., Gao, Z. (2015). Genetic diversity of male and female Chinese bayberry (*Myrica rubra*) populations and identification of sex-associated markers. BMC Genomics,16(1).

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24: 1403-1405. doi:10.1093/bioinformatics/btn129

Jombart T. & Ahmed I. (2011). adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics. doi:10.1093/bioinformatics/btr521

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., Tabima, J.F., Grünwald, N.J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2:e281.

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. Front. Genet. 6:208.

Kaya, Z., Kun, E., Güner, A. (1997). National plan for in situ conservation ofplant genetic diversity in Turkey. Milli Egitim Basimevi, Istanbul 125p

Kaya, Z. (2013). SNP diversity of candidate genes encoding cellulose and lignin biosynthetic enzymes in *Populus nigra* clone bank in Turkey: its implications for conservation and breeding. In Plant and Animal Genome XXI Conference. Plant and Animal Genome.

Kliman, R., Sheehy, B. & Schultz, J. (2008). Genetic Drift and Effective Population Size. Nature Education 1(3):3

Larson, M. G. (2008). Analysis of Variance. Circulation,117(1), 115-121. doi:10.1161/circulationaha.107.654335

Lefèvre, F., Légionnet, A., Vries, S. D., & Turok, J. (1998). Strategies for the conservation of a pioneer tree species, Populus nigra L., in Europe. Genetics Selection Evolution,30(Supplement). doi:10.1051/gse:19980711

Li, Y. (2004). Microsatellites Within Genes: Structure, Function, and Evolution. Molecular Biology and Evolution,21(6), 991-1007. doi:10.1093/molbev/msh073

Maddison, D. R., Swofford, D. L., & Maddison, W. P. (1997). NEXUS: An Extensible File Format for Systematic Information. Systematic Biology,46(4), 590. doi:10.2307/2413497

Mcletchie, D. N., Tuskan G.A. (1994). Gender determination in *Populus.* Norwegian Journal of Agriculture Science 18: 57–66.

Moneger, F. (2007). Sex Determination in Plants. Plant Signal Behav. 2(3): p178-179

Mousadik, A. E., & Petit, R. J. (1996). High level of genetic differentiation for allelic richness among populations of the argan tree [*Argania spinosa* (L.) Skeels] endemic to Morocco. TAG Theoretical and Applied Genetics,92(7), 832-839.

Nei, M. (1973). Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences, USA 70, 3321–3.

Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics, 89(3):583-590

Novotna, K. & Stochlova, P. (2013). Aspects of sexual reproduction in rare monoecious Populus nigra var. nigra trees. Silvae Genetica 62(3):117-124

Paolucci, I., Gaudet, M., Jorge, V., Beritognolo, I., Terzoli, S., Kuzminsky, E., Muleo, R., Mugnozza, G.S., Sabatti, M. (2010). Genetic linkage maps of *Populus alba L.* and comparative mapping analysis of sex determination across *Populus* species. Tree Genetics & Genomes,6(6), 863-875.

Paradis E. (2010). pegas: an R package for population genetics with an integrated-modular approach. Bioinformatics 26: 419-420.

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

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

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

Robertson, A. (1965). The interpretation of genotypic ratios in domestic animal populations. Animal Production,7(03), 319-324.

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

Rowland, D.L., Garner, E.R., Jespersen, M. (2002). A Rare Occurrence of Seed Formation on Male Branches of the Dioecious Tree, *Populus deltoides*. The American Midland Naturalist 147(1):185-187.

Schmidt, A. L., & Mitter, V. (2004). Microsatellite mutation directed by an external stimulus. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis,568(2), 233-243.

Selkoe, K. A., & Toonen, R. J. (2006). Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. Ecology Letters,9(5), 615-629.

Smulders, M. J. M., Cottrell, J. E., Lefèvre, F., Van der Schoot, J., Arens, P., Vosman, B., Tabbener, H., Grassi, F., Fossati, T., Castiglione, S., Fluch, S., Burg, K., Vornam, B., Pohl, A., Gebhardt, K., Alba, N., Agundez, D., Maestro, C., Notivol, E., Volosyanchuk, R., Popiskova, M., Bordacs, S., Bovenschen, J., Van Dam, B., Koelewijn, H., Halfmaerten, D., Ivens, B., Van Slycken, J., Vanden Broeck, A., Storme, V., Krystufek, V. (2008). Structure of the genetic diversity in black poplar (Populus nigra L.) populations across European river systems: consequences for conservation and restoration. Forest Ecology and Management, 255(5), 1388-1399.

Stehlik, I., & Barrett, S. C. (2005). Mechanisms Governing Sex-Ratio Variation In Dioecious *Rumex Nivalis*. Evolution,59(4), 814.

Tunctaner K. (1994). Conservation of genetic resources of Populus nigra in Turkey. *Populus nigra* Network. Report of the first meeting, Izmit, Turkey, pp. 41-44.

QGIS Development Team (2009). QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org.

Wang, J. (2005). Estimation of effective population sizes from data on genetic markers. Philosophical Transactions of the Royal Society B: Biological Sciences,360(1459), 1395-1409.

Weir, B. S., Cockerham, C., Clark (1984). Estimating F-Statistics for the Analysis of Population Structure. Evolution. 38 (6): 1358. ISSN 0014-3820.

Willing, E., Dreyer, C., & Oosterhout, C. V. (2012). Estimates of Genetic Differentiation Measured by FST Do Not Necessarily Require Large Sample Sizes When Using Many SNP Markers. PLoS ONE,7(8).

Winter, D. J. (2012). Mmod: an R library for the calculation of population differentiation statistics. Molecular Ecology Resources,12(6), 1158-1160.

Wright, S. (1943). Isolation by distance. Genetics 28: 114–138.

Wright, S. (1951). The genetical structure of populations. Annu. Eugen. 15: 323–354.

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

# APPENDIX A

# PYTHON SCRIPT FOR DATA CONVERSION

```
1.  import re
2.
3.  def popWriter(definition,popNum,locusList,lines,alleleNum,typ): #mainly to write th
    e data into the output file with appropriate parsing
4.
5.      letterDict = {}
6.      for i in range(alleleNum): #create a dictionary for number to letter conversion

7.          letterDict[str(i+1).zfill(2)] = chr(ord('a')+i) #increment letter along wit
    h the integer
8.      letterDict['00'] = '0'
9.
10.
11.     numDict = {}
12.     for i in range(alleleNum):
13.         numDict[str(i+1)] = chr(ord('a')+i)
14.     numDict['?'] = '0'
15.
16.     with open("POPGENE.txt","w") as fileObj:
17.         fileObj.write("/*"+ definition + "*/" + "\n" + "Number of populations = " +
    str(popNum) + "\n"\
18.                       + "Number of loci = " + str(len(locusList)) + "\n" + "Locus n
    ame:" + "\n")
19.         for i in range(len(locusList)):
20.             fileObj.write(locusList[i] + ' ')
21.         fileObj.write("\n\n" + "POP1" + "\n")
22.
23.         counter = 2
24.
25.         for i1 in range(len(lines)): #iterate over the lines containing the allele
    data
26.
27.             if len(lines[i1])>1:
28.                 for i2 in range(len(lines[i1])):
29.                     try:
30.                         if typ == 1: #if the input is GENEPOP type
31.                             fileObj.write(letterDict[(lines[i1][i2])[:2]] + letterD
    ict[(lines[i1][i2])[2:4]] + "\t")
32.                         elif typ == 2 and i2+1<len(lines[i1]): #if the input is GDA
    type
33.                             temp = lines[i1][i2+1].split('/')
34.                             temp[1] = re.sub('[,\n]','',temp[1])
35.                             fileObj.write(numDict[temp[0]] + numDict[temp[1]] + "\t
    ")
36.                     except KeyError:
37.                         print("\nError: Maximum possible allele types is exceeded.\
    n")
38.                         return 0
39.             else:
```

```
                if not lines[i1]:
                    break
                if re.search(('POP'or'Pop'or'pop'),lines[i1][0])!=None or re.search
(':',lines[i1][0])!=None:
                    fileObj.write("\n" + "POP" + str(counter))
                    counter+=1
                else:
                    continue
            fileObj.write("\n")

def genWriter(definition,popNum,locusList,lines,alleleNum): #mainly to write the da
ta into the output file with appropriate parsing

    numDict = {}
    for i in range(alleleNum): #create a dictionary for letter to number conversion

        numDict[chr(ord('a')+i)] = str(i+1).zfill(2) #increment letter along with t
he integer
    numDict['0'] = '00'

    with open("GENEPOP.txt","w") as fileObj:
        fileObj.write(definition + "\n")
        for elements in locusList:
            fileObj.write(elements + "\n")
        fileObj.write("\n")
        counter = 1

        for i1 in range(len(lines)): #iterate over the lines containing the allele
data
            if len(lines[i1])>=len(locusList):
                fileObj.write(str(counter) + ',\t')
                counter+=1
                for i2 in range(len(lines[i1])):
                    try:
                        fileObj.write(numDict[(lines[i1][i2])[:1]] + numDict[(lines
[i1][i2])[1:2]] + "\t")
                    except KeyError:
                        print("\nError: Maximum possible allele types is exceeded.\
n")
                        return 0
            else:
                fileObj.write("POP")
                counter = 1

            fileObj.write("\n")




def GENconv(): #conversion if the input is GENEPOP type (mainly to get the data fro
m the input file)
    locusList = []
    lines = []
    popNum = 0
    with open("input.txt","r") as fileObj:
        definition = next(fileObj)
        definition = re.sub("[\n\t]",'',definition)
        for line in fileObj:

```

```
            if line.startswith("POP" or "Pop" or "pop"):
                popNum+=1
                break
            if line.strip() == '':
                continue
            line = re.sub("[\n\t]",'',line)
            locusList.append(line)
        for line in fileObj: #iterate over the file object to create a list containing lines
            if line.startswith("POP" or "Pop" or "pop"):
                popNum+=1

            lineList = line.split()
            for i in range(len(lineList)):
                if re.search(',',lineList[i]):
                    del lineList[0:i+1]
                    break
            lines.append(lineList)

        typ = 1
        popWriter(definition,popNum,locusList,lines,alleleNum,typ) #call for writing into output txt


def POPconv(): #conversion if the input is POPGENE type (mainly to get the data from the input file)
    locusList = []
    lines = []
    popNum = 0
    with open("input.txt","r") as fileObj:
        definition = next(fileObj)
        definition = re.sub("[\n\t]",'',definition)
        popNum = int(re.search(r'\d+',next(fileObj)).group())
        next(fileObj)
        next(fileObj)
        locusList = next(fileObj).split()

        for line in fileObj: #iterate over the file object to create a list containing lines
            lineList = line.split()
            if lineList == []: #to remove empty lines
                continue
            lines.append(lineList)

        genWriter(definition,popNum,locusList,lines,alleleNum) #call for writing into output txt

def GDAconv(): #conversion if the input is GDA (mainly to get the data from the input file)
    locusList = []
    lines = []
    popNum = 0
    definition = ""
    seperator = "/"
    with open("input.txt","r") as fileObj:
        for line in fileObj:
            if line.startswith('['):
                line = re.sub('\n','',line)
                definition = line
            elif line.startswith("begin"):
```

```
                continue
            elif re.search(r"\d",line) and not re.search("(nloci)|(npops)",line):
                tempLine = line.split()
                tempLine[1] = re.sub('\'','',tempLine[1])
                locusList.append(tempLine[1])
            elif line.startswith("matrix"):
                break
        next(fileObj)

        for line in fileObj: #iterate over the file object to create a list containing lines
            if re.search(":",line):
                popNum+=1
            lineList = line.split()
            lines.append(lineList)

        typ = 2
        popWriter(definition,popNum,locusList,lines,alleleNum,typ) #call for writing into output txt

print("Current version can convert POPGENE to GENEPOP, GENEPOP to POPGENE and GDA to POPGENE diploid allele data.\nNote that if you enter the wrong file type, program will not terminate, so please check your output afterwards.\n")

while 1:

    fileType = str(input("Please name your txt file as 'input' and write down the type of your file below as GDA, POPGENE or GENEPOP:\n"))
    alleleNum = 26 #max different alleles that can be considered

    try:
        f = open("input.txt","r")
        f.close()
    except IOError:
        print("\nPlease place your input file into the program's directory.\n")
        continue

    if fileType == "GDA":
        GDAconv()
        break
    elif fileType == "POPGENE":
        POPconv()
        break
    elif fileType == "GENEPOP":
        GENconv()
        break
    else:
        print("\nPlease write down an appropriate file type.\n")


        print("\n\nProcess completed. Please check your output file.")
```

# APPENDIX B

# AVAILABLE DATA OF CLONE SAMPLES

**Table B**

| Sample | Region | City | Longitude | Latitude | Altitude | Gender | Genotype |
|---|---|---|---|---|---|---|---|
| **1257** | Central | Corum | 34° 03' | 40° 30' | 801 | F | 1 |
| **1** | Central | Kirsehir | 33° 42' | 39° 21' | 1140 | F | 1 |
| **17** | Central | Ankara | 33° 07' | 39° 33' | 1280 | F | 2 |
| **356** | Central | Kirsehir | 33° 43' | 39° 22' | 1090 | F | 4 |
| **366** | Central | Cankiri | 32° 59' | 40° 47' | 1181 | F | 1 |
| **441** | Central | Nigde | 34° 52' | 37° 25' | 830 | F | 5 |
| **481** | Central | Karaman | 33° 14' | 37° 09' | 1110 | F | 1 |
| **662** | Central | Ankara | 33° 50' | 39° 54' | 780 | F | 1 |
| **696** | Central | Konya | 32° 29' | 37° 54' | 1010 | F | 1 |
| **1031** | Central | Konya | 32° 12' | 37° 12' | 1200 | F | 6 |
| **1081** | Central | Ankara | 33° 00' | 40° 00' | 1250 | F | 7 |
| **1206** | Central | Nigde | #N/A | #N/A | #N/A | F | 8 |
| **1461** | Central | Cankiri | 33° 39' | 40° 37' | 1313 | F | 10 |
| **112** | Central | Kirsehir | 33° 10' | 39° 53' | 1300 | M | 1 |
| **191** | Central | Cankiri | 32° 53' | 40° 48' | 1138 | M | 12 |
| **296** | Central | Ankara | 33° 07' | 39° 33' | 1300 | M | 13 |
| **666** | Central | Konya | 31° 50' | 37° 25' | 1128 | M | 1 |
| **711** | Central | Konya | 32° 25' | 38° 11' | 1160 | M | 17 |
| **1091** | Central | Nigde | #N/A | #N/A | #N/A | M | 18 |
| **1156** | Central | Konya | 32° 29' | 37° 53' | 1020 | M | 19 |
| **66** | Eastern | Erzurum | 41° 14' | 39° 53' | 1884 | F | 20 |
| **76** | Eastern | Van | 43° 12' | 39° 07' | 1850 | F | 21 |
| **136** | Eastern | Elazig | 38° 47' | 39° 46' | 880 | F | 22 |
| **231** | Eastern | Erzurum | 42° 08' | 39° 00' | 1660 | F | 23 |
| **241** | Eastern | Sivas | 36° 51' | 39° 49' | 1250 | F | 24 |
| **346** | Eastern | Erzurum | 41° 14' | 39° 53' | 1860 | F | 25 |
| **351** | Eastern | Sivas | 37° 00' | 39° 46' | 1250 | F | 20 |
| **436** | Eastern | Mus | 41° 41' | 38° 37' | 1400 | F | 26 |

**Table B (Cont'd)**

| Sample | Region | City | Longitude | Latitude | Altitude | Gender | Genotype |
|---|---|---|---|---|---|---|---|
| **921** | Eastern | Erzurum | 41° 41' | 39° 59' | 1560 | F | 1 |
| **966** | Eastern | Malatya | 38° 15' | 38° 18' | 1000 | F | 27 |
| **972** | Eastern | Sivas | 37° 00' | 39° 46' | 1300 | F | 28 |
| **986** | Eastern | Erzurum | 41° 30' | 40° 15' | 1550 | F | 29 |
| **1021** | Eastern | Erzincan | 38° 42' | 39° 57' | 1515 | F | 30 |
| **1056** | Eastern | Van | 43° 12' | 39° 07' | 1850 | F | 1 |
| **1121** | Eastern | Van | 43° 18' | 39° 12' | 1750 | F | 31 |
| **1292** | Eastern | Sivas | 37° 00' | 39° 46' | 1300 | F | 1 |
| **1301** | Eastern | Malatya | 38° 15' | 38° 27' | 760 | F | 1 |
| **1366** | Eastern | Erzurum | 42° 08' | 40° 03' | 1660 | F | 32 |
| **36** | Eastern | Erzurum | 41° 19' | 39° 54' | 1970 | M | 33 |
| **101** | Eastern | Erzincan | 39° 37' | 39° 43' | 1160 | M | 34 |
| **276** | Eastern | Elazig | 38° 44' | 39° 58' | 1000 | M | 35 |
| **587** | Eastern | Erzincan | 39° 38' | 40° 26' | 1350 | M | 1 |
| **821** | Eastern | Erzurum | 41° 30' | 40° 24' | 1500 | M | 36 |
| **957** | Eastern | Erzincan | 39° 37' | 39° 43' | 1160 | M | 37 |
| **1232** | Eastern | Erzurum | 41° 29' | 40° 47' | 540 | M | 38 |
| **1246** | Eastern | Sivas | 36° 29' | 39° 23' | 1268 | M | 1 |
| **1401** | Eastern | Malatya | 37° 53' | 37° 55' | 1245 | M | 1 |
| **1476** | Aegean | Konya | 31° 23' | 38° 24' | 1000 | F | 11 |
| **1371** | Aegean | Konya | 31° 20' | 38° 22' | 1200 | F | 9 |
| **181** | Aegean | Konya | 31° 24' | 38° 21' | 950 | F | 3 |
| **236** | Aegean | Eskisehir | 30° 33' | 39° 45' | 810 | F | 1 |
| **1162** | Aegean | Isparta | 30° 33' | 37° 45' | 1000 | F | 38 |
| **377** | Aegean | Isparta | 31° 10' | 38° 18' | 1100 | F | 56 |
| **453** | Aegean | Isparta | 31° 10' | 38° 18' | 1100 | F | 1 |
| **489** | Aegean | Isparta | 31° 05' | 38° 09' | 880 | F | 57 |
| **557** | Aegean | Isparta | 31° 10' | 38° 18' | 1100 | F | 60 |
| **211** | Aegean | Afyon | 30° 19' | 38° 39' | 1010 | F | 39 |
| **426** | Aegean | Denizli | 29° 30' | 37° 19' | 800 | F | 40 |
| **491** | Aegean | Denizli | 29° 40' | 37° 12' | 900 | F | 41 |
| **686** | Aegean | Afyon | 29° 51' | 37° 53' | 880 | F | 42 |
| **886** | Aegean | Kutahya | 30° 03' | 39° 29' | 969 | F | 43 |
| **942** | Aegean | Kutahya | 29° 56' | 39° 25' | 1208 | F | 44 |
| **1071** | Aegean | Afyon | 30° 27' | 38° 48' | 1013 | F | 1 |

**Table B (Cont'd)**

| Sample | Region | City | Longitude | Latitude | Altitude | Gender | Genotype |
|---|---|---|---|---|---|---|---|
| **1112** | Aegean | Kutahya | 30° 29' | 38° 46' | 1009 | F | 45 |
| **1146** | Aegean | Afyon | 31° 02' | 38° 43' | 995 | F | 46 |
| **1406** | Aegean | Usak | 29° 36' | 38° 27' | 900 | F | 1 |
| **1482** | Aegean | Denizli | 29° 16' | 37° 45' | 550 | F | 1 |
| **676** | Aegean | Eskisehir | 30° 30' | 39° 45' | 820 | M | 16 |
| **636** | Aegean | Konya | 31° 23' | 38° 24' | 1000 | M | 15 |
| **597** | Aegean | Usak | 29° 37' | 38° 33' | 920 | M | 1 |
| **1241** | Aegean | Afyon | 30° 19' | 38° 39' | 1010 | M | 38 |
| **416** | Blacksea | Karabk | 32° 35' | 40° 59' | 620 | F | 48 |
| **196** | Blacksea | Kastamonu | 34° 43' | 41° 27' | 456 | F | 47 |
| **406** | Blacksea | Amasya | 35° 41' | 40° 27' | 683 | F | 1 |
| **421** | Blacksea | Amasya | 36° 01' | 40° 53' | 1002 | F | 48 |
| **591** | Blacksea | Tokat | 36° 31' | 40° 21' | 709 | F | 49 |
| **906** | Blacksea | Kastamonu | 33° 36' | 40° 55' | 1013 | F | 1 |
| **929** | Blacksea | Tokat | 37° 49' | 40° 18' | 981 | F | 1 |
| **1173** | Blacksea | Samsun | 35° 38' | 40° 54' | 586 | F | 38 |
| **1197** | Blacksea | Tokat | 36° 34' | 40° 38' | 401 | F | 38 |
| **1216** | Blacksea | Kastamonu | 33° 54' | 41° 29' | 671 | F | 50 |
| **1316** | Blacksea | Samsun | 35° 37' | 40° 47' | 470 | F | 51 |
| **1351** | Blacksea | Sinop | 35° 31' | 41° 02' | 622 | F | 1 |
| **1426** | Blacksea | Tokat | 36° 31' | 40° 41' | 218 | F | 2 |
| **381** | Blacksea | ? | 38° 25' | 40° 17' | 1400 | M | 14 |
| **246** | Blacksea | Tokat | 36° 03' | 40° 23' | 565 | M | 52 |
| **717** | Blacksea | Amasya | 36° 45' | 40° 32' | 555 | M | 53 |
| **1086** | Blacksea | Kastamonu | 34° 11' | 41° 31' | 574 | M | 1 |
| **1126** | Blacksea | Kastamonu | 34° 11' | 41° 31' | 584 | M | 54 |
| **1311** | Blacksea | Bartin | 32° 20' | 41° 30' | 50 | M | 1 |
| **1361** | Blacksea | Amasya | 34° 28' | 40° 44' | 801 | M | 1 |
| **1471** | Blacksea | Amasya | 35° 49' | 40° 38' | 413 | M | 55 |
| **527** | Marmara | Sakarya | 30° 22' | 40° 42' | 39 | F | 58 |
| **547** | Marmara | Yalova | 29° 53' | 40° 27' | 665 | F | 59 |
| **1101** | Marmara | Sakarya | 30° 17' | 40° 39' | 489 | F | 61 |
| **1251** | Marmara | Sakarya | 30° 22' | 40° 42' | 39 | F | 1 |
| **1422** | Marmara | Bilecik | 30° 00' | 40° 21' | 106 | F | 1 |

# APPENDIX C

# R SIMULATION PROTOTYPE SCRIPT

```
this.dir <- dirname(parent.frame(2)$ofile)
setwd(this.dir)

require("adegenet")
require("mmod")
require("poppr")
require("hierfstat")



print("PopSim V.0.5")

{
fileName <- readline(prompt = "File name:\n")
allCode <- as.integer(readline(prompt = "Number of characters used to
code an allele:\n"))
cloneCorrect <- readline(prompt = "Execute clone correction to remove
duplicate genotypes in each population: 'YES' or 'NO'\n")
genderBool <- readline(prompt = "Do you have gender file? 'YES' or
'NO'\n")

if (genderBool == 'YES'){
        gender <- 'YES'
        genderFile <- readline(prompt = "Enter gender file name:\n")
}

if (genderBool == 'NO') {
        gender <- 'NO'
}

excLoci <- readline(prompt = "Choose loci to include, put ',' between
entries. To analyze all loci, enter 'ALL':\n")
Ngen <- as.numeric(readline(prompt = "How many generations to run (1-
400):\n"))
Mratio <- as.numeric(readline(prompt = "Male/population ratio (0-
1):\n"))
popMulti <- as.numeric(readline(prompt = "Population size multiplier
(1-10):\n"))
isolation <- as.numeric(readline(prompt = "Introduce isolation at
generation (1-generation), enter '0' for no isolation:\n"))
genExp <- as.numeric(readline(prompt = "Generation to export stats (1-
generation), enter '0' if you don't want to export (WARNING:
Generation export can cause memory overload!):\n"))
}

if (gender == 'YES'){
  genderFile <- read.table(file = paste0(getwd(), "/", genderFile))
}

popS <- read.genepop(paste0(getwd(), "/", fileName), ncode = allCode)
# Import GENEPOP data into genind object

if(gender == 'YES'){
  popS@strata <- data.frame(Population = popS@pop, Gender =
```

```
unlist(genderFile, use.names = FALSE)) # Append gender information
}else {
  popS@strata <- data.frame(Population = popS@pop)
}

# Clone correction
if(cloneCorrect == 'YES'){
        popS <- clonecorrect(popS, ~Population)
}

# Include loci
excLoci <- unlist(strsplit(excLoci,","))
if(excLoci != 'ALL'){
        popS <- popS[loc = excLoci]
}

# Extra strata, index to combine all populations
popS@strata <- cbind(popS@strata, all = sapply(1:nInd(popS),
function(x) x = 1))


# Migration pop to pop
mList <- c()
popNames <- levels(popS@pop)
combL <- length(combn(popNames,2))
migNames <- vector(mode = "character", length = length(combL))

i3 <- 1
for(i1 in 1:length(popNames)){
        for (i2 in 1:length(popNames)){
                if(i1 == i2){
                        next
                }
                migNames[i3] <- paste(popNames[i1], "->",popNames[i2])
                i3 <- i3+1
        }
}

{
        print("Provide migration info:")
        for (i in 1:length(migNames)){
                mList[i] <- as.numeric(readline(prompt =
paste(migNames[i], "\n")))
        }

}
names(mList) <- migNames
mList <- round(popMulti*mList)


#mList <- round(popMulti*c(12.0*0.098, 12.6*0.098, 14.4*0.098,
12.0*0.098, 12.6*0.098, 15.5*0.098, 11.4*0.098, 11.0*0.098,
19.3*0.097, 10.1*0.097,
#                               16.6*0.097, 7.7*0.097, 10.6*0.097,
21.9*0.097, 17.9*0.097, 4.7*0.097, 7.6*0.096, 11.5*0.096, 13.7*0.096,
```

```
7.5*0.096))


# mList <- (sapply(1:length(mList), function(x) mList[x] <-
2.38))*popMulti



## Functions


setMethod("tab", signature(x="genpop"), function(x, freq=FALSE,
NA.method=c("asis","mean","zero"), ...){
 ## handle arguments
    NA.method <- match.arg(NA.method)
    # outdim <- dim(x@tab)
    ## get matrix of data
    if(!freq) {
        out <- x@tab
    } else {
        out <- x@tab
        f1 <- function(vec) return(vec/sum(vec,na.rm=TRUE))
        ## compute frequencies
        fac <- x@loc.fac
        if (is.null(fac)) fac <- rep(1, nLoc(x))
        out <- apply(x@tab, 1, tapply, fac, f1, simplify = FALSE)
        if (ncol(x@tab) > 1){
          ## reshape into matrix
          col.names <- do.call(c,lapply(out[[1]],names))
          row.names <- names(out)
          out <- matrix(unlist(out), byrow=TRUE, nrow=nrow(x@tab),
                        dimnames=list(row.names, col.names))
          ## reorder columns
          out <- out[, colnames(x@tab), drop = FALSE]
        } else {
          out <- matrix(out, nrow = length(out), ncol = 1,
                        dimnames = list(rownames(x@tab),
colnames(x@tab)))
        }

    }

    ## replace NAs if needed
    if(NA.method=="mean"){
        f1 <- function(vec){
            m <- mean(vec, na.rm=TRUE)
            vec[is.na(vec)] <- m
            return(vec)
        }

        out <- apply(out, 2, f1)
    }
    if(NA.method=="zero"){
        out[is.na(out)] <- ifelse(freq, 0, 0L)
    }
```

```
    # dim(out) <- outdim
    ## return output
    return(out)
})

# Function to divide the generation into two subgroups based on sex
ratio for next hybridization, returns a list containing male and
female subgroups
sexDivider <- function(pop, Mratio){
        popM <- pop[sample(1:nInd(pop), round(Mratio*nInd(pop)))]
        temp <- indNames(pop)
        temp <- setdiff(temp, indNames(popM))
        popF <- pop[temp]
        popMF <- list(popM = popM, popF = popF)
        return(popMF)
}


# Function to implement migration between populations, returns a list
containing updated new populations. Fixes individual numbers in each
population if migration rates differ between populations
migrator <- function(popList, mList){

        popListN <- popList
        n <- length(popList)
        drawList <- list()
        i3 <- 1
        for (i1 in 1:length(popListN)){
                for (i2 in setdiff(1:length(popListN), i1)){
                        temp <- popListN[[i2]]
[sample(1:nInd(popListN[[i2]]), round(mList[[i3]]))]
                        drawList[[i3]] <- indNames(temp)
                        if (mList[[i3]] == 0){}
                        else {names(drawList[[i3]]) <- i2}
                        popListN[[i1]] <- repool(popListN[[i1]], temp)
                        i3 <- i3 + 1
                }

        }
        for (i1 in 1:length(popListN)){
                for (i2 in 1:length(drawList)){
                        if (!is.null(names(drawList[[i2]][1])) &&
names(drawList[[i2]][1]) == i1){
                                namesP <- indNames(popListN[[i1]])
                                popListN[[i1]] <- popListN[[i1]]
[setdiff(namesP, drawList[[i2]])]
                        }
                }
                if (nInd(popListN[[i1]]) > nInd(popList[[i1]])){
                        popListN[[i1]] <- popListN[[i1]]
[sample(1:nInd(popListN[[i1]]), nInd(popList[[i1]]))]
                }
                else if (nInd(popListN[[i1]]) < nInd(popList[[i1]])){
                        sample <- popListN[[i1]]
[sample(1:nInd(popListN[[i1]]), nInd(popList[[i1]])-
```

```
nInd(popListN[[i1]]))]
                        popListN[[i1]] <- repool(popListN[[i1]],
sample)
                }
                pop(popListN[[i1]]) <- rep(levels(pop(popList[[i1]])),
nInd(popListN[[i1]]))
        }
        return(popListN)
}


locCount <- length(names(popS@loc.n.all))

sumPop <- summary(popS)
popSizes <- paste(sumPop$n.by.pop, collapse = " ")
allNum <- paste(sumPop$loc.n.all, collapse = " ")
allPop <- paste(sumPop$pop.n.all, collapse = " ")
missData <- sumPop$NA.perc
sumPop$Hobs <- sapply(1:locCount, function(x) round(sumPop$Hobs[x],
digits = 4))
hobsLoc <- paste(sumPop$Hobs, collapse = " ")
sumPop$Hexp <- sapply(1:locCount, function(x) round(sumPop$Hexp[x],
digits = 4))
hexpLoc <- paste(sumPop$Hobs, collapse = " ")

# Summary Text

cat(paste("Number of individuals", "\n", sumPop$n, "\n\n"), file =
"summary.txt")
cat(paste("Group sizes", "\n", popSizes, "\n\n"), file =
"summary.txt", append = TRUE)
cat(paste("Number of alleles per locus", "\n", allNum, "\n\n"), file =
"summary.txt", append = TRUE)
cat(paste("Number of alleles per group", "\n", allPop, "\n\n"), file =
"summary.txt", append = TRUE)
cat(paste("Percentage of missing data", "\n", round(missData, digits =
4), "\n\n"), file = "summary.txt", append = TRUE)
cat(paste("Observed heterozygosity", "\n", hobsLoc, "\n\n"), file =
"summary.txt", append = TRUE)
cat(paste("Expected heterozygosity", "\n", hexpLoc, "\n\n"), file =
"summary.txt", append = TRUE)

diff_popS <- diff_stats(popS)
diff_popS$global <- diff_popS$global[-6]
diff_popS$global[6] <- wc(popS)$FST
diff_popS$global[7] <- wc(popS)$FIS
names(diff_popS$global) <- c("Hs", "Ht", "Gst", "G'st", "D", "Fst",
"Fis")
total <- diff_popS$global

diff_popS$per.locus <- cbind(diff_popS$per.locus, Fst =
wc(popS)$per.loc$FST, Fis = wc(popS)$per.loc$FIS)
diff_popS$per.locus <- as.data.frame(diff_popS$per.locus)
diff_popS$per.locus[locCount+1,] <- total
row.names(diff_popS$per.locus)[locCount+1] <- "Overall"
```

```
write.csv2(round(diff_popS$per.locus, digits = 4), file = "F-
stats.csv") # F-stats
```

```
popNum <- length(levels(popS@pop))

popTot <- list() # List containing each generation
gstatList <- c() # List holding Gst of each generation
hobsList <- c() # List holding Hobs of each generation
hexpNeiList <- c() # List holding Hexp of each generation
popList <- list() # List holding seperated populations
fisList <- c()
fstList <- c()
alleleCount <- c()
alleleFreq <- list()
hexPopList <- list()
hexPop <- c()

# Initialization with gender information, first generation is produced
if(gender == "YES"){

        setPop(popS) <- ~Population/Gender
        popSep <- seppop(popS)


        i2 <- 1
        for (i in seq(1, length(popSep)-1, 2)){
                popG <- popSep[[i]]@strata[[1]][1]
                popG <- as.character(popG)
                popList[[i2]] <- hybridize(popSep[[i]], popSep[[i+1]],
(nInd(popSep[[i]])+nInd(popSep[[i+1]]))*popMulti, hyb.label = popG)
                #temp <- hybridize(popSep[[i]], popSep[[i+1]],
(nInd(popSep[[i]])+nInd(popSep[[i+1]])))
                #popSep[[i]] <- repool(temp, popSep[[i]],
popSep[[i+1]])
                #popList[[i2]] <- popSep[[i]]
                pop(popList[[i2]]) <- rep(popG, nInd(popList[[i2]]))
                hexPop[i2] <- poppr(popList[[i2]], quiet = TRUE)[1,10]
                i2 <- i2+1
        }
        hexPopList[[1]] <- hexPop
        popTot[[1]] <- repool(popList)
        gstatList[1] <- Gst_Hedrick(popTot[[1]])$global
        hobsList[1] <- mean(summary(popTot[[1]])$Hobs)
        hexpNeiList[1] <- poppr(popTot[[1]], quiet = TRUE)
[(popNum+1),10]
        fisList[1] <- wc(popTot[[1]])$FIS
        fstList[1] <- wc(popTot[[1]])$FST
        alleleCount[1] <- sum(nAll(popTot[[1]]))
        pop(popTot[[1]]) <- rep("all", nInd(popTot[[1]]))
```

```
        alleleFreq[[1]] <- tab(genind2genpop(popTot[[1]], quiet =
TRUE), freq = TRUE)
}



# Initialization without gender information, first generation is
produced
if(gender == "NO"){

        setPop(popS) <- ~Population
        popSep <- seppop(popS)


        for (i in 1:length(popSep)){
                popG <- levels(pop(popSep[[i]]))
                popMF <- sexDivider(popSep[[i]], Mratio)
                popM <- popMF$popM
                popF <- popMF$popF
                popList[[i]] <- hybridize(popM, popF,
nInd(popSep[[i]])*popMulti, hyb.label = popG)
                #temp <- hybridize(popSep[[i]], popSep[[i]],
nInd(popSep[[i]])*2)
                #popSep[[i]] <- repool(temp, popSep[[i]])
                #popList[[i]] <- popSep[[i]]
                pop(popList[[i]]) <- rep(popG, nInd(popList[[i]]))
                hexPop[i] <- poppr(popList[[i]], quiet = TRUE)[1,10]

        }
        hexPopList[[1]] <- hexPop
        popTot[[1]] <- repool(popList)
        gstatList[1] <- Gst_Hedrick(popTot[[1]])$global
        hobsList[1] <- mean(summary(popTot[[1]])$Hobs)
        hexpNeiList[1] <- poppr(popTot[[1]], quiet = TRUE)
[(popNum+1),10]
        fisList[1] <- wc(popTot[[1]])$FIS
        fstList[1] <- wc(popTot[[1]])$FST
        alleleCount[1] <- sum(nAll(popTot[[1]]))
        pop(popTot[[1]]) <- rep("all", nInd(popTot[[1]]))
        alleleFreq[[1]] <- tab(genind2genpop(popTot[[1]], quiet =
TRUE), freq = TRUE)
}



start <- Sys.time()
for (i0 in 2:Ngen){
        for (i in 1:length(popList)){
                popG <- levels(pop(popList[[i]]))
                iNames <- paste0(i0, popG)
                popMF <- sexDivider(popList[[i]], Mratio)
                popM <- popMF$popM
                popF <- popMF$popF
                popList[[i]] <- hybridize(popM, popF,
nInd(popList[[i]]), hyb.label = iNames)
```

```
                pop(popList[[i]]) <- rep(popG, nInd(popList[[i]]))
                hexPop[i] <- poppr(popList[[i]], quiet = TRUE)[1,10]
        }
        if ((isolation != 0) & (i0 == isolation)){
                isolation <- isolation + 1
        }
        else {
        popList <- migrator(popList, mList)
        }
        hexPopList[[i0]] <- hexPop
        popTot[[i0]] <- repool(popList)
        gstatList[i0] <- Gst_Hedrick(popTot[[i0]])$global
        hobsList[i0] <- mean(summary(popTot[[i0]])$Hobs)
        hexpNeiList[i0] <- poppr(popTot[[i0]], quiet = TRUE)
[(popNum+1),10]
        fisList[i0] <- wc(popTot[[i0]])$FIS
        fstList[i0] <- wc(popTot[[i0]])$FST
        alleleCount[i0] <- sum(nAll(popTot[[i0]]))
        pop(popTot[[i0]]) <- rep("all", nInd(popTot[[i0]]))
        alleleFreq[[i0]] <- tab(genind2genpop(popTot[[i0]], quiet =
TRUE), freq = TRUE)
        if (genExp == 0){
                popTot[[i0]] <- NA
        }

        print(paste("Generation:", i0))
}
end <- Sys.time()
end - start



if (genExp != 0){

  sumPop <- summary(popTot[[genExp]])
  popSizes <- paste(sumPop$n.by.pop, collapse = " ")
  allNum <- paste(sumPop$loc.n.all, collapse = " ")
  allPop <- paste(sumPop$pop.n.all, collapse = " ")
  missData <- sumPop$NA.perc
  sumPop$Hobs <- sapply(1:locCount, function(x) round(sumPop$Hobs[x],
digits = 4))
  hobsLoc <- paste(sumPop$Hobs, collapse = " ")
  sumPop$Hexp <- sapply(1:locCount, function(x) round(sumPop$Hexp[x],
digits = 4))
  hexpLoc <- paste(sumPop$Hobs, collapse = " ")

  cat(paste("Number of individuals", "\n", sumPop$n, "\n\n"), file =
"summaryGen.txt")
  cat(paste("Group sizes", "\n", popSizes, "\n\n"), file =
"summaryGen.txt", append = TRUE)
  cat(paste("Number of alleles per locus", "\n", allNum, "\n\n"), file
= "summaryGen.txt", append = TRUE)
  cat(paste("Number of alleles per group", "\n", allPop, "\n\n"), file
= "summaryGen.txt", append = TRUE)
  cat(paste("Percentage of missing data", "\n", round(missData, digits
```

```
= 4), "\n\n"), file = "summaryGen.txt", append = TRUE)
  cat(paste("Observed heterozygosity", "\n", hobsLoc, "\n\n"), file =
"summaryGen.txt", append = TRUE)
  cat(paste("Expected heterozygosity", "\n", hexpLoc, "\n\n"), file =
"summaryGen.txt", append = TRUE)

  diff_popS <- diff_stats(popTot[[genExp]])
  diff_popS$global <- diff_popS$global[-6]
  diff_popS$global[6] <- wc(popS)$FST
  diff_popS$global[7] <- wc(popS)$FIS
  names(diff_popS$global) <- c("Hs", "Ht", "Gst", "G'st", "D", "Fst",
"Fis")
  total <- diff_popS$global

  diff_popS$per.locus <- cbind(diff_popS$per.locus, Fst =
wc(popS)$per.loc$FST, Fis = wc(popS)$per.loc$FIS)
  diff_popS$per.locus <- as.data.frame(diff_popS$per.locus)
  diff_popS$per.locus[locCount+1,] <- total
  row.names(diff_popS$per.locus)[locCount+1] <- "Overall"

  write.csv2(round(diff_popS$per.locus, digits = 4), file = "F-
statsGen.csv") # F-stats
}



#colPal <- colorRampPalette(c("red", "blue"))


png(filename = "Gstat.png")
plot(c(1:Ngen), gstatList, xlab = "Generation", ylab = "Gst", ylim =
c(0,1))
abline(lm(unlist(gstatList)~c(1:Ngen)), col = "red")
dev.off()

png(filename = "Hobs.png")
plot(c(1:Ngen), hobsList, xlab = "Generation", ylab = "Hobs", type =
"l", ylim = c(0,1))
abline(lm(unlist(hobsList)~c(1:Ngen)), col = "red")
dev.off()

png(filename = "Hexp.png")
plot(c(1:Ngen), hexpNeiList, xlab = "Generation", ylab = "Hexp", type
= "l", ylim = c(0,1))
abline(lm(unlist(hexpNeiList)~c(1:Ngen)), col = "red")
dev.off()

png(filename = "Fis.png")
plot(c(1:Ngen), fisList, xlab = "Generation", ylab = "Fis", ylim =
c(-0.1,0.1))
abline(lm(unlist(fisList)~c(1:Ngen)), col = "red")
dev.off()


png(filename = "Fst.png")
```

```
plot(c(1:Ngen), fstList, xlab = "Generation", ylab = "Fst", ylim =
c(0,1))
abline(lm(unlist(fstList)~c(1:Ngen)), col = "red")
dev.off()

png(filename = "AlleleCount.png")
plot(c(1:Ngen), alleleCount, xlab = "Generation", ylab = "Allele
Count", type = "l", lwd = 1)
abline(lm(unlist(alleleCount)~c(1:Ngen)), col = "red")
dev.off()


alleleFreq <- lapply(1:length(alleleFreq), function(x) alleleFreq[[x]]
<- as.data.frame(alleleFreq[[x]]))
alleleFreq <- lapply(1:length(alleleFreq), function(x) alleleFreq[[x]]
<- cbind(unlist(alleleFreq[[x]])))

initFreq <- tab(genind2genpop(popS, ~all, quiet = TRUE), freq = TRUE)
initFreq <- as.data.frame(initFreq)
names(initFreq)


temp <- c()
png(filename = "AlleleFreq.png")
plot(c(1:Ngen), temp, ylim = c(0,1), xlim = c(0,Ngen), xlab =
"Generation", ylab = "Allele Frequency", lty = 1, lwd = 0.8, type =
"l")
for (i in names(initFreq)){
        if(initFreq[i] > 0.0){
        temp <- lapply(1:length(alleleFreq), function(x) temp[x] <-
alleleFreq[[x]][,1][i])
        lines(c(1:Ngen), temp, type = "l", lwd = 0.8, col =
sample(colors(), 100))
        }
}
dev.off()


temp <- c()
png(filename = "HexpPops.png")
plot(c(1:Ngen), temp, ylim = c(0.0, 0.8), xlim = c(0,Ngen), xlab =
"Generation", ylab = "Hexp", lty = 1, lwd = 1, type = "l")
for (i in 1:length(hexPopList[[1]])){
        temp <- sapply(1:length(hexPopList), function(x) temp[x] <-
hexPopList[[x]][i])
        lines(c(1:Ngen), temp, type = "l", lwd = 1, col =
sample(colors(), 100))
}
dev.off()
```

# APPENDIX D

## FREQUENCIES OF GENDER SPECIFIC ALLELES

**Table D**

| Male | Central A. (N = 20) | Eastern A. (N = 27) | Aegean (N = 24) | Blacksea (N = 21) | Marmara (N = 11) |
|---|---|---|---|---|---|
| **W14 - 234** | 0 | 0 | 0 | 0.02 | 0 |
| **W14 -260** | 0 | 0 | 0 | 0.02 | 0 |
| **W10 - 246** | 0 | 0 | 0 | 0.02 | 0 |
| **W10 - 256** | 0 | 0 | 0 | 0.02 | 0 |
| **W03 - 270** | 0.03 | 0 | 0 | 0 | 0 |
| **W03 - 280** | 0 | 0.02 | 0 | 0 | 0.05 |
| **P14 - 195** | 0 | 0 | 0 | 0 | 0.05 |
| **P14 - 211** | 0 | 0.02 | 0 | 0 | 0.14 |
| **P21 - 238** | 0 | 0 | 0 | 0 | 0.09 |
| **P21 - 244** | 0 | 0 | 0 | 0.02 | 0 |
| **P21 - 254** | 0 | 0 | 0 | 0.05 | 0 |
| **Female** | **Central A. (N = 20)** | **Eastern A. (N = 27)** | **Aegean (N = 24)** | **Blacksea (N = 21)** | **Marmara (N = 11)** |
| **W15 - 195** | 0 | 0.04 | 0.02 | 0 | 0 |
| **W14 - 237** | 0 | 0.02 | 0 | 0 | 0 |
| **W14 - 252** | 0.03 | 0.02 | 0 | 0 | 0.05 |
| **W14 - 261** | 0 | 0 | 0.02 | 0 | 0 |
| **W10 - 231** | 0 | 0.02 | 0 | 0 | 0 |
| **W10 - 240** | 0 | 0.02 | 0 | 0 | 0.05 |
| **W10 - 250** | 0 | 0.04 | 0 | 0 | 0 |
| **W05 - 278** | 0 | 0 | 0.08 | 0.09 | 0 |
| **W05 - 284** | 0 | 0.02 | 0 | 0 | 0 |
| **W20 - 237** | 0 | 0.04 | 0 | 0 | 0 |
| **W03 - 266** | 0 | 0 | 0.02 | 0.05 | 0 |
| **W03 - 272** | 0 | 0.04 | 0 | 0 | 0.09 |
| **W12 - 152** | 0 | 0.04 | 0.02 | 0 | 0 |
| **W12 - 169** | 0 | 0.02 | 0 | 0 | 0.05 |
| **W12 - 177** | 0.03 | 0 | 0 | 0 | 0 |
| **W12 - 189** | 0 | 0 | 0.02 | 0 | 0 |
| **P21 - 220** | 0 | 0 | 0 | 0.02 | 0 |
| **P21 - 246** | 0.03 | 0.04 | 0.05 | 0 | 0 |
| **P21 - 258** | 0.03 | 0.02 | 0 | 0 | 0 |
| **P21 - 263** | 0 | 0.02 | 0 | 0 | 0 |
| **P21 - 270** | 0 | 0 | 0.02 | 0 | 0 |

## APPENDIX E

## MIGRATE SOFTWARE OUTPUT (INITIAL PAGE)

MIGRATION RATE AND POPULATION SIZE ESTIMATION
using the coalescent and maximum likelihood or Bayesian inference
Migrate-n version 3.6.11 [June-18-15]
Program started at Sat Jun 10 02:35:59 2017
Program finished at Sat Jun 10 03:11:16 2017

### *Options*

Datatype: Microsatellite data [Brownian motion]
Missing data: not included

Inheritance scalers in use for Thetas:
All loci use an inheritance scaler of 1.0
[The locus with a scaler of 1.0 used as reference]
Random number seed: (with internal timer) 2887913400
Start parameters:

Theta values were generated from the FST-calculation

M values were generated from the FST-calculation

Connection type matrix:
where m = average (average over a group of Thetas or M,
s = symmetric M, S = symmetric 4Nm, 0 = zero, and not estimated,
* = free to vary, Thetas are on diagonal

| Population | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 CentralAnatolia | * | * | * | * | * |
| 2 EasternAnatolia | * | * | * | * | * |
| 3 Aegean | * | * | * | * | * |
| 4 Blacksea | * | * | * | * | * |
| 5 Marmara | * | * | * | * | * |

Order of parameters:

| | | |
|---|---|---|
| 1 | $\Theta_1$ | <displayed> |
| 2 | $\Theta_2$ | <displayed> |

## APPENDIX F

## STRUCTURE RESULTS SHOWING ASSIGNMENT OF 5 GEOGRAPHIC POPULATIONS TO 3 ESTIMATED CLUSTERS

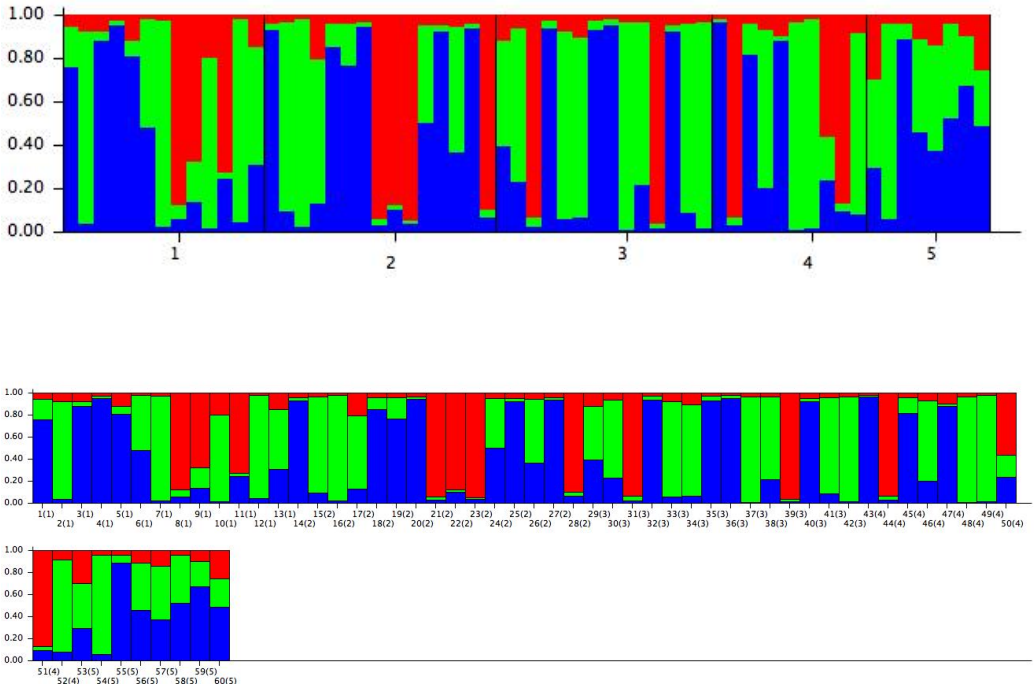



**Figure F.** Individuals 1-13 belong to Central Anatolia, 14-28 belong to Eastern Anatolia, 29-42 belong to Aegean, 43-52 belong to Blacksea, 53-60 belong to Marmara populations.