

Detection of genetic diversity and population structure in Iranian Hazelnut (*Corylus avellana* L.) using SCoT markers and Phytochemical traits

Mehran Ochi-Ardabili

Islamic Azad University Miyaneh Branch

Hassan Nourafcan

Islamic Azad University of Miyaneh: Islamic Azad University Miyaneh Branch

Hassanali Naghdi Badi (✉ Naghdibadi@yahoo.com)

Shahed University <https://orcid.org/0000-0001-7070-8467>

Nasser Mohebalipour

Islamic Azad University Miyaneh Branch

Ardeshtir Qaderi

ACECR: Academic Center for Education Culture and Research

Research Article

Keywords: *Corylus avellana*, Population Genetics, Molecular marker, SCoT, Taxol.

Posted Date: July 13th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-3135434/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published at Genetic Resources and Crop Evolution on November 2nd, 2023. See the published version at <https://doi.org/10.1007/s10722-023-01790-7>.

Abstract

Corylus avellana L. is one of the valuable nuts of the world, which grows wild in the natural forests of Fandoguloo in Ardabil province of Iran and has a high diversity in this region. To evaluate the genetic diversity within and between populations, 78 genotypes from different regions of Iran, including Ardabil, Arasbaran, and Miyaneh, were analyzed using targeted start codon polymorphism (SCoT) method. Also, phytochemical traits of *C. avellana*, such as Ash, Taxol, Total phenol, Palmitic acid, Fibre, Palmitoleic acid, carbohydrate, Stearic acid, flavonoid, Oleic acid, Linoleic acid, Fat, Antioxidant compounds, and ethanolic phenol were evaluated. The result indicated that, from the 15 SCoT primers used, the genotypes contained 165 polymorphic bands with the highest and lowest Polymorphic Information Content related to SC3 and SC21, respectively. The Marker index (MI) for the primer sets varied from 2.6 to 6.1. The highest allele frequency was related to the Ardabil population. Also, 100% polymorphism rate was obtained from all primers studied. The highest polymorphic rates corresponded to the primers SC21, SC20, SC5 (14 bands), with SC9 showing the most uniformity (six bands) among the populations. The gene flow between and within populations was 0.4 and 0.7, respectively, and the stabilization index was 0.4. The current research showed the effectiveness of the SCoT marker to estimate the genetic diversity of *C. avellana*, which can be considered as a starting point for further study on the population and evolutionary genetics of this economically important species.

1. Introduction

The common hazel (*Corylus avellana*) is a long-lived, monoecious, and multistemmed shrub. This wind-pollinated species is distributed across the North-Africa (Mediterranean region), all over the Europe, and western Asia (Leinemann et al. 2013). It's not clear that European hazel (*Corylus avellana* L.) domesticated in which place and what time although it was already cultivated by the Romans. Based on chloroplast microsatellite investigation, hazelnut appears to have been domesticated independently in three areas including the Mediterranean, Turkey, and Iran. (Boccacci and Botta 2009).

Furthermore, it is regarded as an essential nut due to its rising production and consumption (Fideghelli and De Salvador 2008). Turkey has been the leading producer and exporter of this nut for several years, with 75% annual production, followed by Italy (13%), the United States (4%), Azerbaijan, Spain, Georgia, and Iran (20% total). Due to favorable climatic conditions, South Caucasus can be potentially turn to one of the main growers, hence the genetic diversity of this area's native hazelnut populations should be evaluated (Król and Gantner 2020).

For crop and tree breeding, genetic diversity has been recognized as the most significant tool. Molecular techniques with improved replicability are typically considered in marker research (Govindaraj et al. 2015). There are a variety of molecular markers used to examine genetic variation in different plant species (Kumar et al. 2009). DNA markers, including restriction fragment length polymorphism (RFLP) and randomly amplified polymorphic DNA (RAPD), are potent tools for studying plant genetics (Garrido-Cardenas et al. 2018) that could be traditionally employed in genetic mapping. In addition, DNA markers could help describe the contribution of various main genetic loci to complicated growth processes (Sanghvi and Dave 2017). Representation of marker polymorphism is the main source of genotypic diversity, both at the morphological and biochemical/ molecular levels (Bellemou et al. 2020).

A new procedure for producing DNA markers in plants is based on the short protected length near the ATG codon. In this procedure, generally, single 18 bp primers are employed for polymerase chain reactions (PCR) at 50°C annealing temperature (Barman and Kundu 2019) and PCR amplicons are visualized by standard agarose gel electrophoresis. This procedure has been confirmed in rice using a diverse collection of genotypes and cross populations (Bertani et al. 2021). The SCoT markers can be used to assess the replicability of research using repeated samples and to perform PCR across multiple days. SCoT markers are generally repeatable, although certain exceptions have revealed that

primer length and annealing temperatures are not the only parameters that determine reproducibility (Gupta et al. 2019). It has been proposed that this method could be used for genetic analysis, separate bulk analysis, and quantitative trait loci mapping (Jiang 2015; Wang et al. 2019).

The current breeding strategy based on the selection from wild species does not meet the increasing demand for hazelnut crops (Kavas et al. 2020). Molecular markers are considered useful tools in the effort for the vast production and proliferation of this crop (Gedil and Menkir 2019). According to (Bhattacharyya et al. 2013), EST-SSR markers can define polymorphism and heterozygosity rates among *Corylus* species and also separate *C. coluna* from *C. avellana* more properly. Accompanied by EST-SSR, SCoT markers system, which can be produced from the transcript sequences, increases the number of markers for fingerprinting experiments and may be useful for identifying the marker-trait associations and analyzing the genetic diversity (Bhattacharyya et al. 2013). SCoT markers is considered as powerful marker to study genetic diversity and defining edible and toxic materials in various wild taxa and crops (Sorkheh et al. 2016). This marker has been successfully uncovered population structure in rose, orchids, *Jatropha*, grape, peanut etc. (Agarwal et al. 2019; Mulpuri et al. 2013; Bhattacharyya et al., 2013; Guo et al. 2012; Gorji et al. 2011). Also, SCoT benefits from the advantage of the combination of RAPD and ISSR, including the simplicity, inexpensiveness, effectiveness, and high rate of replicability (Bhattacharyya et al. 2013).

Considering horticulture, the selection of hazelnut masses for such essential traits as fruit weight, nut to fruit ratio, and fruit quality is of a great performance due to the fact that the fruiting of the hazelnut tree entails many years. The study of molecular markers associated with hazelnut garden traits boosted by reproducibility and accessibility can play a significant role in hazelnut breeding programs.

To date, the genetic diversity and polymorphism of the sequence near the start codon in the South Caucasus native hazelnuts have not been explored, hence the aim of this study was to investigate the genetic diversity of the native hazelnut in South Caucasus to develop programs for conservation and breeding.

2. Materials and methods

2.1. Plant material and DNA extraction

Seventy-eight different hazelnut genotypes were collected from the forests of Fandoghloo located in Ardabil province, Arasbaran and Miyaneh region in Iran and were evaluated along with Fertile and Daviana genotypes (Fig. 1). These genotypes were selected from 5 discrete areas, which were called subpopulations.

Geographical data of the Hazelnut habitats have been given in detail in Supplementary Table 1. Seeds of each genotype were seeded in Murashige and Skoog (MS) medium with macro and micronutrients ratio of 4:1. The plants were grown in optimum conditions in phytotron (16/8 hours light/darkness, 23 ± 2 degrees Celsius). Fresh leaves of each genotype were sampled for DNA extraction. Total genomic DNA from 0.1g leaf with trivial changes was isolated by CTAB detergent (Baek et al., 2008). DNA quality was evaluated by gel agarose electrophoresis 0.8%.

2.2. Evaluation of Genotypes with the SCoT markers

15 SCoT primers (Table 1) were designed according to Collard and Mackill method (Collard and Mackill 2009). All primers amplified remarkable polymorph bands. PCR reaction was performed in 30 μ l as described earlier. PCR reactions were amplified under the following conditions: primary denaturation at 95°C for 10 minutes followed by 38 periods of denaturation at 95°C for 1 minute, annealing at 45°C (for all primers) for 45 seconds and extension to 72°C for 2 minutes followed by final extension at 72°C for 7 minutes PCR Program performed.

2.3. Phytochemical

Total carbohydrates were evaluated by the Schlegel method (Sheligl 1986). To determine the concentration of taxol, an HPLC device (Knauer, Germany) equipped with a c-18 column (μm , 250 x 4.6mm Teknokrome, Germany Perfectsil Target ODS3,5) was used with the help of taxol standard (SIGMA). To determine the profile of fatty acids, the oil was extracted using Soxhlet and the profile of fatty acids was determined by gas chromatography. The amount of total phenol was measured by the Folin Sieve Caltio method (Grillo et al. 2019) and the antioxidant capacity was measured by the DPPH method (Sirivibulkovit et al. 2018). After extracting phosphorus by calorimetric method, calcium, potassium, magnesium, iron and zinc elements were measured using an atomic absorption device (AS. A 2000-CTA) and the amount of hazelnut kernel elements was expressed in mg/kg dry matter (Sujka et al. 2017).

Table 1
SCoT primers sequence used in this research.

Name	Sequence	Tm (c)	GC content%
SC3	CAACAATGGCTACCACCG	55.2	55.60%
SC4	CAACAATGGCTACCACCT	53.5	50%
SC5	CAACAATGGCTACCACGA	54.0	50%
SC6	CAACAATGGCTACCACGC	55.9	55.60%
SC7	CAACAATGGCTACCACGG	55.2	55.60%
SC8	CAACAATGGCTACCACGT	54.3	50%
SC9	CAACAATGGCTACCAGCA	54.3	50%
SC11	AAGCAATGGCTACCACCA	55.4	50%
SC12	ACGACATGGCGACCAACG	59.8	61.10%
SC17	CATGGCTACCACCGGCCC	61.9	72.20%
SC18	ACCATGGCTACCACCGCG	61.5	66.70%
SC19	GCAACAATGGCTACCACC	55.4	55.60%
SC20	AACCATGGCTACCACCGC	59.1	61.10%
SC21	CACCATGGCTACCACCAT	55.6	55.60%
SC23	ACCATGGCTACCACGGGC	61.1	66.70%

2.4. Data collection and processing

All transparent polymorph bands in the profile (1) and SCoT (0) were calculated based on the existence and absence of the bands. The differentiation potential of the used primers was evaluated via the data parameters. In order to distinguish genetic diversity within and among the genotypes, molecular variance analysis (AMOVA) was carried out to calculate the number of polymorphic bands (NPB), polymorphic phase boundary (PPB), polymorphism information content (PIC), resolving power (Rp) and marker index (MI) (Angel 2002; Nayak and Prakash 2018). Also, genotypes distributions were evaluated by principal coordinates analysis (PCoA) (Wang et al. 2022). This analysis was carried out by GenAlix version 6.5 (Caniglia et al., 2016). Seven indices of genetic variations including the number of observed (Na) and effective (Ne) allele, genetic diversity Nei (H), Shannon data index (I), percentage of polymorph locus (PPL), genetic differentiation among genotypes (Gst), and gene flow (Nm) were estimated by POPGENE version 1.31 (Warghat et al. 2012) (Tang and Ma 2020) (Liang et al., 2008). Genetic dissimilarity was estimated based on the Jaccard similarity coefficient by using the DARwin software version 6 (Odongo et al. 2015). UPGMA cluster analyses was

performed on the basis of coefficients obtained from PCoA, and dendrogram were constructed using Wilks-Lambda coefficients. Nearest neighbour dendrograms were presented to demonstrate relations among the different genotypes (Ronikier and Zalewska-Gałosz 2014). Step-by-step multivariate regression method was used to investigate the alleles affecting the diversity and changes of the studied traits (Mooney and Wilmot 2015).

3. Results

In this research, 78 genotypes were evaluated using the SCoT primers. All used primers showed clear and scalable band patterns (Fig. 2). In total, 165 bands were detected, all of which were polymorphic at the genotypes level. The number of bands varied from six bands for SC9 to 14 bands for SC21, SC12, and SC5.

The PIC index for each primer was determined across the genotypes investigated. The results shown in Table 2 showed that SC3 and SC21 had the greatest and lowest PIC values, indicating that SC3 could better specify the genetic distance of the examined genotypes than the others, and so might be deemed an acceptable primer for evaluating their genetic diversity. The MI index (Table 2) ranged between 2.6 and 6.1. The SC12 primer had the greatest MI value, indicating its high efficiency in detecting polymorphism in the studied genotypes, while SC3 had the lowest rate. Higher RP values suggest that they are more effective in differentiating genotypes. SC21, SC12, and SC5 primers had the greatest RP (28), indicating that they have a greater ability to detect genotypes differences compared to other markers. Based on the 15 SCoT primers used on 78 hazelnut genotypes for estimating the SCoT efficiency, the PIC and RP values were calculated to be 0.417 and 28, respectively.

Table 2
Allele frequency and genetic diversity indices relating the SCoT markers.

Marker	TB	P	P%	PIC	GD	MI	Rp
Sc3	12	12	100	0.458251	0.352194	5.499	24
Sc4	10	10	100	0.381131	0.306215	3.811	20
Sc5	14	14	100	0.390509	0.308457	5.467	28
Sc6	9	9	100	0.407298	0.322158	3.665	18
Sc7	8	8	100	0.400764	0.317750	3.206	16
Sc8	9	9	100	0.415297	0.319346	3.737	18
Sc9	6	6	100	0.435404	0.338674	2.612	12
Sc11	11	11	100	0.409300	0.323852	4.502	22
Sc12	14	14	100	0.435733	0.334761	6.100	28
Sc17	11	11	100	0.430817	0.336974	4.733	22
Sc18	13	13	100	0.441056	0.342464	5.733	26
Sc19	11	11	100	0.410645	0.323300	4.517	22
Sc20	13	13	100	0.446847	0.344534	5.809	26
Sc21	14	14	100	0.355217	0.284998	4.973	28
Sc23	10	10	100	0.446746	0.345946	4.467	20
Mean	11	11	100	0.417668	0.326775	4.589	22
TB: Total band, P: Polymorphism band, P%: polymorphism percentage, PIC: Polymorphic information content, GD: Gene Diversity, MI: Marker Index, Rp: Resolving power							

3.1. Genetic diversity indices and frequency of alleles in hazelnut Genotypes

The number of genotypes in each subpopulation and the mean and standard deviation of the diversity coefficients for each of them are calculated and shown in Table 3. Nei' genetic diversity index ranged from 1.26 ± 0.02 to 1.69 ± 0.02 . The highest (1.69 ± 0.02) and lowest (0.141 ± 0.02) number of effective alleles belonged to the Essygran (B) and Fandoghloo (A) genotypes, respectively. The highest N_i (0.388 ± 0.01) and Shannon (0.564 ± 0.02) indices were related to the Essygran genotype indicating a favorable allelic frequency and the highest genetic diversity among the genotypes. As for the estimated indices for the primers, the SC3 primer had the greatest PIC (0.45) and RP (0.35), showing that it is superior to the others in terms of evaluating genetic diversity.

Table 3
Allelic frequency and genetic diversity indices relating the hazelnut populations.

Sub Population	h	I	N	Ne	PB	PPL
Fandoghloo- Ardabil (A)	0.190 ± 0.01	0.324 ± 0.02	14	1.26 ± 0.02	0	100%
Essygran- Ardabil (B)	0.388 ± 0.01	0.564 ± 0.02	14	1.69 ± 0.02	0	100%
Abibighloo-Ardabil (C)	0.370 ± 0.01	0.544 ± 0.02	14	1.64 ± 0.02	0	100%
Miyaneh-Esat Azarbaijan (D)	0.366 ± 0.01	0.534 ± 0.02	14	1.64 ± 0.02	1	100%
Arasbaran- East Azarbaijan (E)	0.347 ± 0.01	0.511 ± 0.02	20	1.60 ± 0.02	5	100%
Average	0.332 ± 0.01	0.171 ± 0.01	-	1.57 ± 0.01	-	28.45%
h: Nei index, I: Shannon information index, N: sample number per population, Ne: effective allele number, PB: specific band number, PPL: Polymorphism percentage						

3.2. Polymorphism within the genotypes using the SCoT primers

The result of Polymorphism within the genotypes using the SCoT primers was showed that, the polymorphism percentage at the genotypes level was 100%, and the highest polymorphic bands corresponded to SC12, SC5, and SC21 with 14 bands which indicate the lowest uniformity among the primers. The lowest band number was related to the SC9 primer with six polymorphic bands indicating the most genotype-level uniformity (Table 4).

Table 4
Gene flow among the hazelnut populations.

Population	Subpop. A	Subpop. B	Subpop. C	Subpop. D	Subpop. E
Fandoghloo- Ardabil (A)	****	0.869	0.932	0.825	0.816
Essygran- Ardabil (B)	0.141	****	0.964	0.923	0.930
Abibighloo-Ardabil (C)	0.070	0.037	****	0.922	0.912
Miyaneh-Esat Azarbaijan (D)	0.192	0.080	0.081	****	0.970
Arasbaran- East Azarbaijan (E)	0.204	0.073	0.092	0.030	****

A- Fandoghloo (Ardabil); B- Essygran (Ardabil); C- Abibighloo (Ardabil); D- Miyaneh (Esatern Azarbayjan); E- Arasbaran (Esatern Azarbayjan); F- Astara (Gilan); G- Astara (Gilan).

3.3. Gene flow among the genotypes

Relations between the five genotypes were evaluated separately (Table 5). For this purpose, indices related to the genetic structure of inter/intra genotypes were calculated. The maximum Nm and lowest fixation index (Fst) values were 0.7 and 0.4, respectively. These revealed that, despite relatively high Nm values, there was some trait segregation, which might be attributable to geographical and altitude factors.

Total heterozygosity (Ht) and heterozygosity within the studied subpopulations (Hs) were estimated to be 0.395 and 0.237, respectively (Table 5). The rate of heterozygosity among genotypes (Dst) was calculated to be 0.158, showing that there is less diversity between genotypes than there is within them. The Hs index showed higher (0.237) diversity than the Dst index (0.158).

Table 5

Similarity indices (above the diagonal) and genetic distance (below the diagonal) among the hazelnut populations based on the Nei index.

Genotypes	Genotype Size	Total heterozygosity Ht	Heterozygosity Within Genotypes HS	Heterozygosity Between Genotypes Dst	Coefficient of Genetic Differentiation (Gst)	Gene Flow Nm	Fixation Index Fst Index
Total Genotypes	78	0.395	0.237	0.158	0.402	0.7	0.554

3.4. Molecular variance analysis within and among the population

The variance within and among genotypes was estimated to be 29.5 and 6.43, respectively (Table 6). Also, the molecular variance within and among genotypes was calculated to be 82 and 18%, respectively (Fig. 3).

Table 6
Analysis of molecular variance of hazelnut populations by SCoT marker

Sources of variation	Degree of Freedom	sum of squares	Average of squares	Estimated variance	Percentage of variance	PhiPT	P
Between Populations	4	506.854	126.713	6.430	18	179.0	0.01
Within Population	71	2100.463	29.854	29.584	82	-	-
Total	75	2607.289	-	36.014	100	-	-

3.5. Genetic distance and Grouping the Subpopulations

The genetic distance and similarity among the Subpopulations based on Nei (1978) have been depicted in Table 7. The highest genetic distance (0.204) and the lowest similarity (0.816) corresponded to the A and E subpopulations. In order to group hazelnut populations, various methods were employed. As for the highest cophenetic coefficient, the UPGMA algorithm for the individual cluster was selected based on the Jaccard similarity coefficient (Table 7).

Table 7
Calculated cophenetic index for UPGMA algorithm and similarity indices
Simple matching, Dice, and Jaccard

Cophenetic Correlation Coefficient			
Jaccard	Dice	Simple matching	Algorithm Similarity Coefficient
0.81	0.443	0.412	UPGMA

3.6. Analysis of the principle components

Analysis into principal components was done to group the studied genotypes, then it was grouped using independent z-coefficients (cluster method). The results are listed in Fig. 4. According to the principal component analysis plot, and the cluster analysis output the results of both analyses are in agreement. The evaluated genotypes were classified into four different groups. This classification was somewhat consistent with the geographical distribution.

3.7. Analysis of alleles accounting the studied traits

In order to analyze the role of each allele in phytochemical traits multiple regression (step by step) was run and the results appeared to be as the following (Table 8). The trait of fruit ash 0.94 significantly accounts for alleles 83a, 26a, 19a, 146a, and 135a. Taxol inside the leaves with six alleles including 165a, 101a, 22a, 96a, 148a, 55a was explained to be 94 percent significant. Regarding total phenol, only the 117a allele was significantly represented by 31 percent. Palmitic acid along with two alleles, namely 153a and 152a was significantly supplied by 59 percent and fiber content was shown to be 77 percent significant with alleles 80a and 116a.

Palmitic acid was shown to be 94 percent significant through alleles 57a, 66a and 19a. 24 percent of carbohydrate was significantly explained by allele 46a. 26 percent of total flavonoids and 30 percent of stearic acid were accounted for by 110a and 46a alleles, respectively. Following step-by-step multiple regression in linoleic acid, 68 percent was represented by alleles 83a and 158a. However, considering the trait of linoleic acid, five dominant alleles including 83a, 46a, 66a, 90a and 61a were significantly accounted for.

The phenolic ethanol was shown to be 86 percent significant with alleles 93a, 117a and 132a. When 72 percent of the fruit content was fat, alleles 112a, 34a and 79a were outstanding. Alleles 40a, 92a, 116a, 112a, 34a, and 79a significantly represented 98 percent of antioxidant compounds. From among the alleles, 83 alleles for three different traits containing linoleic acid, oleic acid and ash; allele 19 for palmitic acid and ash traits; allele 66 for palmitic acid and linoleic acid; allele 116 for two different traits of antioxidant compounds and fiber, and finally allele 146 accounted for the two attributes of linoleic acid and ash. The remaining alleles were only characterized as describing one trait. Allele 24a achieved greatest role of the allelic unit in the carbohydrate traits represented by 34 percent.

Table 8
The role of SCoT alleles accounting traits related to phytochemistry and medicinal substances of *C. avellana*

Trait	Rf	R2	Sig	Number of effective Alleles	Accounting Alleles
Ash	0.973	0.94	0.00	5	135a, 146a, 19a, 83a, 26a
Taxol	0.975	0.94	0.00	6	55a, 148a, 96a, 22a, 101a, 165a
Total phenol	0.56	0.31	0.018	1	117a
Palmitic acid	0.772	0.59	0.002	2	152a, 153a
Fibre	0.82	0.77	0.001	2	80a, 116a
Palmitoleic acid	0.975	0.94	0.00	3	19a, 66a, 57a
Carbohydrate	0.62	0.34	0.008	1	24a
Stearic acid	0.55	0.3	0.02	1	46a
Flavonoid	0.51	0.26	0.00	1	110a
Oleic acid	0.83	0.68	0.003	2	83a, 158a
Linoleic acid	0.83	0.68	0.00	5	64a, 66a, 83a, 90a, 146a
Fat	0.85	0.72	0.08	2	109a, 115a
Antioxidant compounds	0.99	0.98	0.00	6	34a, 40a, 79a, 92a, 113a, 116a
ethanolic phenol%	0.93	0.86	0.019	2	109a, 115a

4. Discussion

The SCoT markers used to analyze the genetic diversity of 78 hazelnut genotypes from in the natural forests of Fandoguloo in Ardabil province of Iran exhibited substantial polymorphism, with six polymorphic bands belonging to SC9 and 14 to SC21, SC12, and SC5 markers out of 165 polymorphic bands at the population level. Due to a large number of genes in different loci of the genome, for each primer, various numbers of bands were obtained per genotype (Collard and Mackill 2009). This has the advantage over other primers, which either act randomly or amplify nongenic loci. The marker index (MI) varied from 2.6 to 6.1, indicating its high efficiency to detect the amount of polymorphism across the examined populations. The SC21, SC12, and SC5 primers possessed the highest MI values (28). The resolving power index (RP) associated with the SCoT primers showed their efficiency in detecting the differentiation of the populations. It has been mentioned by Bhattacharyya et al. (2013), that evaluated the genetic diversity of wild orchid and supported the efficiency of 15 SCoT markers. Mulpuri and colleagues (2013) in their survey using SCoT marker, revealed high polymorphism in *Jatropha curcas* and confirmed application of this marker in distinguishing the toxic and edible genotypes (Mulpuri et al. 2013).

The highest N_e (0.02–1.69), N_{ei} (0.01–0.388), and I (0.02–0.564) corresponded to the Essygran (Ardabil) subpopulation, and the lowest one (0.02–1.141) was estimated for the subpopulation from Fandoghloo (Ardabil), indicating the high allelic frequency and genetic diversity among the subpopulations. According to the calculated indices for the primers, SC3 had the highest PIC (0.45) and RP (0.35) values. These findings also suggest that SC3 outperforms other primers in terms of assessing genetic diversity. The analysis of SCoT markers revealed 100% polymorphism at the population level, and the SC12, SC5, and SC21 primers had the highest polymorphism with 14 bands, but the lowest uniformity when compared to other primers.

SC9 was allocated the lowest band number, with six polymorph bands reflecting the best population-level consistency. Further, the greatest N_m (0.7) and lowest F_{st} (0.554) values across the groups corresponded to intra populations. This illustrates that, despite relatively large N_m values, dispersion among populations can be observed to some extent that possibly influenced by glacial-interglacial climatic changes in the area. Similar pattern of climatic-affected differentiation in the Italian hazel populations was also reported by Palme' and Vendramin (2002).

The evaluation of genetic diversity among geographic locations due to factors such as linkage, inbreeding, migration, and differences associated with samples is of great importance (Mohammadi and Prasanna 2003). In other words, the validity of such analyses highly depends on the number of individuals examined, the number of allelic loci, the allelic and genotypic characteristics of the specimens, sample size, and the type of the intercross (Strauss et al. 1992). The H_t and H_s parameters were calculated to be 0.395 and 0.237, respectively. According to the population analysis, the amount of heterozygosity within subpopulations ($H_s = 0.237$) was higher than between subpopulations ($D_s = 0.158$). Genetic differentiation may result from differences in genetic features caused by environmental changes and natural selection (Bossdorf et al. 2005), as well as random events such as mutation, migration, and the degree of population divergence, which could be evaluated using several parameters (Slatkin 1995). In the case of this study, our results suggest that the isolation of hazel populations might be due to past climatic fluctuations during glacial periods that led to fragmented populations in the studied area. This pattern has been also reported by several authors for different entities in this area (Rajaei Sh et al. 2013; Ahmadzadeh et al. 2013; Rezazadeh et al. 2020).

Based on the combination of SCoT markers used, the variance within and among the subpopulations was estimated to be 29.5 and 6.43, respectively, implying considerable demographic differentiation in the studied area. The highest genetic distance (0.204) and the lowest similarity (0.816) among the subpopulations corresponded to the Fandoghloo and Essygran subpopulations. Due to the non-random feature of the Jaccard indices, they have no definite statistical distribution, which, in turn, makes it difficult to calculate sampling variance and, consequently, confidence intervals (Lombard et al. 2000). However, bootstrapping techniques could tackle the problem (Tivang et al. 1994).

Simple matching, Dice, and Jaccard were used to categorize the genotypes, with similar cophenetic indices indicating the concordance between the grouping outcomes. The population grouping was found to be in harmony with geographic locations, and this result was in accordance with the grouping based on the PCoA analysis; the examination of the genetic structure of the examined populations deviated to three main groups with distinct genetic structures; 1- Ardabil (Fandoghloo, Essygran and Abibigloo subpopulations with less geographic distances), 2- Miyaneh and 3- Arasbaran.

The rate of total heterozygosity and heterozygosity within the subpopulations was estimated to be 0.395 and 0.237, respectively. However, the inter-subpopulation heterozygosity was estimated to be 0.157, showing that there is less genetic variation among subpopulations than within subpopulations. This was determined by estimating the inter and intra-subpopulation variances using the SCoT data from the Fandoghloo (Ardabil) and Arasbaran subpopulations. The Nei distance between the subpopulations of Fandoghloo and Arasbaran was 0.204. The current geographic isolation of these two subpopulations related to probable difficulty in gene flow, because these forests distributed in mountainous areas, and southern Caucasus range and Sabalan mountains may act as natural barriers for pollination.

This work is first survey on the genetic diversity of *C. avellana* in Iran by applying SCoT marker as a powerful and sensitive technique to identify low levels of genetic variations. The present study revealed higher genetic diversity within-populations but moderate genetic diversity among-populations. It can be concluded that identification, conservation and breeding programs of such genetically diverse wild germplasm from in the natural forests of Fandoghloo in Ardabil province of Iran would be necessary for this economically important taxon. Our results revealed a general trend for isolation by distance plus genetic differentiation between hazelnut populations of different geographic areas. It may relate to existing of natural geographic barriers between isolated populations as well as habitat degradation due to antropogenic activities, like livestock overgrazing and unsustainable utilization of forests. The remarkable diversity was found within the populations in this study (specially Fandoghloo forest) provides essential insights into maintenance and breeding programs of this valuable crop.

Declarations

Acknowledgements

The authors thank Dr. Mohammad Zakizadeh for his scientific advice regarding molecular analysis.

Competing interests

The authors declare no competing interests.

Authors' contributions

I/we declare that the contribution of particular Authors to the publication is as follows: Developing the concept, the methods and the assumptions: Hassanali Naghdi Badi, Hassan Nourafcan, Ardeshir Qaderi and Mehran Ochi-Ardabili; conducting the studies: Mehran Ochi-Ardabili, Hassanali Naghdi Badi and Hassan Nourafcan; statistical analyzes of data: Mehran Ochi-Ardabili, Ardeshir Qaderi and Nasser Mohebalipour; All authors contributed to the drafts and gave final approval for publication.

Funding

The authors did not receive support from any organization for the submitted work.

Availability of data and materials

References

1. Agarwal A, Gupta V, Haq SU, Jatav PK, Kothari S, Kachhwaha S (2019) Assessment of genetic diversity in 29 rose germplasms using SCoT marker. *Journal of King Saud University-Science* 31: 780-788. <https://doi.org/10.1016/j.jksus.2018.04.022>
2. Ahmadzadeh F, Flecks M, Rödder D, Böhme W, Ilgaz Ç, Harris DJ, Engler JO, Üzümlü N, Carretero MA (2013) Multiple dispersal out of Anatolia: biogeography and evolution of oriental green lizards. *Biological Journal of the Linnean Society* 110(2):398-408. <https://doi.org/10.1111/bij.12129>
3. Angel R (2002) Genetic diversity of *Halodule wrightii* using random amplified polymorphic DNA. *Aquatic Botany*. 74(2):165-74. [https://doi.org/10.1016/S0304-3770\(02\)00079-7](https://doi.org/10.1016/S0304-3770(02)00079-7)
4. Baek MH, Chung BY, Kim JH, Kim JS, Lee SS, An BC, Lee IJ, Kim TH (2008) cDNA cloning and expression pattern of cinnamate-4-hydroxylase in the Korean black raspberry. *BMB reports* 41(7):529-36. <https://doi.org/10.5483/BMBRep.2008.41.7.529>
5. Barman M, Kundu S (2019) Molecular markers and a new vista in plant breeding: A review. *Int J Curr Microbiol App Sci* 8(12):1921-9. <https://doi.org/10.20546/ijcmas.2019.812.230>
6. Bellemou D, Millàn T, Gil J, Abdelguerfi A, Laouar M (2020) Genetic diversity and population structure of Algerian chickpea (*Cicer arietinum*) genotypes: use of agro-morphological traits and molecular markers linked or not linked to the gene or QTL of interest. *Crop and Pasture Science* 71(2):155-70. <https://doi.org/10.1071/CP19255>
7. Bertani RP, Perera MF, Joya CM, Henriquez DD, Funes C, Chaves S, González V, Welin B, Cuenya MI, Castagnaro AP (2021) Genetic diversity and population structure of *Acidovorax avenae* subsp. *avenae* isolated from sugarcane in Argentina. *Plant Pathology* 70(7):1719-32. <https://doi.org/10.1111/ppa.13413>
8. Bhattacharyya P, Kumaria S, Kumar S, Tandon P (2013) Start Codon Targeted (SCoT) marker reveals genetic diversity of *Dendrobium nobile* Lindl., an endangered medicinal orchid species. *Gene* 529(1):21-6. <https://doi.org/10.1016/j.gene.2013.07.096>
9. Boccacci P, Botta R (2009) Investigating the origin of hazelnut (*Corylus avellana* L.) cultivars using chloroplast microsatellites. *Genet Resour Crop Evol* 56: 851–859. <https://doi.org/10.1007/s10722-009-9406-6>
10. Bossdorf O, Auge H, Lafuma L, Rogers WE, Siemann E, Prati D (2005) Phenotypic and genetic differentiation between native and introduced plant populations. *Oecologia* 144:1-11. <https://doi.org/10.1007/s00442-005-0070-z>
11. Caniglia R, Galaverni M, Delogu M, Fabbri E, Musto C, Randi E (2016) Big bad wolf or man's best friend? Unmasking a false wolf aggression on humans. *Forensic science international: genetics* 24:e4-6. <https://doi.org/10.1016/j.fsigen.2016.06.009>
12. Collard BCY, Mackill DJ (2009) Start codon targeted (SCoT) polymorphism: a simple, novel DNA marker technique for generating gene-targeted markers in plants. *Plant molecular biology reporter*. 27:86-93. <https://doi.org/10.1007/s11105-008-0060-5>
13. Fideghelli C, De Salvador FR (2008) World hazelnut situation and perspectives. In VII International Congress on Hazelnut 845: 39-52. <https://doi.org/10.17660/ActaHortic.2009.845.2>
14. Garrido-Cardenas JA, Mesa-Valle C, Manzano-Agugliaro F (2018) Trends in plant research using molecular markers. *Planta* 247:543-57. <https://doi.org/10.1007/s00425-017-2829-y>
15. Gedil M, Menkir A (2019) An integrated molecular and conventional breeding scheme for enhancing genetic gain in maize in Africa. *Frontiers in Plant Science* 10:1430. <https://doi.org/10.3389/fpls.2019.01430>

16. Germain E (1992) The reproduction of hazelnut (*Corylus avellana* L.): a review. In III International Congress on Hazelnut 351:195-210. <http://doi.org/10.17660/ActaHortic.1994.351.19>
17. Gorji AM, Poczai P, Polgar Z, Teller J (2011) Efficiency of arbitrarily amplified dominant markers (SCoT, ISSR and RAPD) for diagnostic fingerprinting in tetraploid potato. American journal of potato research 88:226-37. <https://doi.org/10.1007/s12230-011-9187-2>
18. Govindaraj M, Vetriventhan M, Srinivasan M (2015) Importance of genetic diversity assessment in crop plants and its recent advances: an overview of its analytical perspectives. Genetics research international. ID 431487. <http://dx.doi.org/10.1155/2015/431487>
19. Grillo G, Boffa L, Binello A, Mantegna S, Cravotto G, Chemat F, Dizhbite T, Lauberte L, Telysheva G (2019) Cocoa bean shell waste valorisation; extraction from lab to pilot-scale cavitation reactors. Food research international 115:200-8. <https://doi.org/10.1016/j.foodres.2018.08.057>
20. Guo DL, Zhang JY, Liu CH (2012) Genetic diversity in some grape varieties revealed by SCoT analyses. Molecular biology reports 39:5307-13. <https://doi.org/10.1007/s11033-011-1329-6>
21. Jennison C, Arnott A, Tessier N, Tavul L, Koepfli C, Felger I, Siba PM, Reeder JC, Bahlo M, Mueller I, Barry AE (2015) Plasmodium vivax populations are more genetically diverse and less structured than sympatric Plasmodium falciparum populations. PLoS neglected tropical diseases. 9(4):e0003634. <https://doi.org/10.1371/journal.pntd.0003634>
22. Jiang GL (2015) Molecular marker-assisted breeding: a plant breeder's review. Advances in plant breeding strategies: breeding, biotechnology and molecular tools. 431-72. https://doi.org/10.1007/978-3-319-22521-0_15
23. Kavas M, Yıldırım K, Seçgin Z, Gökdemir G (2020) Discovery of simple sequence repeat (SSR) markers in hazelnut (*Corylus avellana* L.) by transcriptome sequencing and SSR-based characterization of hazelnut cultivars. Scandinavian Journal of Forest Research 35:227-37. <https://doi.org/10.1080/02827581.2020.1797155>
24. Król K, Gantner M (2020) Morphological traits and chemical composition of hazelnut from different geographical origins: A review. Agriculture. 10(9):375. <https://doi.org/10.3390/agriculture10090375>
25. Kumar P, Gupta VK, Misra AK, Modi DR, Pandey BK (2009) Potential of molecular markers in plant biotechnology. Plant omics. 2(4):141-62.
26. Leinemann L, Steiner W, Hosius B, Kuchma O, Arenhövel W, Fussi B, Haase B, Kätzel R, Rogge M, Finkeldey R (2013) Genetic variation of chloroplast and nuclear markers in natural populations of hazelnut (*Corylus avellana* L.) in Germany. Plant Systematics and Evolution 299:369-78. <https://doi.org/10.1007/s00606-012-0727-0>
27. Liang HH, Cheng Z, Yang XL, Li S, Ding ZQ, Zhou TS, Zhang WJ, Chen JK (2008) Genetic diversity and structure of *Cordyceps sinensis* populations from extensive geographical regions in China as revealed by inter-simple sequence repeat markers. The Journal of Microbiology 46:549-56. <https://doi.org/10.1007/s12275-008-0107-1>
28. Lombard V, Baril CP, Dubreuil P, Blouet F, Zhang D. Genetic relationships and fingerprinting of rapeseed cultivars by AFLP: consequences for varietal registration. Crop Science 40(5):1417-25. <https://doi.org/10.2135/cropsci2000.4051417x>
29. Mohammadi SA, Prasanna BM (2003) Analysis of genetic diversity in crop plants—salient statistical tools and considerations. Crop science 43(4):1235-48. <https://doi.org/10.2135/cropsci2003.1235>
30. Mooney MA, Wilmot B (2015) Gene set analysis: a step-by-step guide. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics 168(7):517-27. <https://doi.org/10.1002/ajmg.b.32328>
31. Mulpuri S, Muddanuru T, Francis G (2013) Start codon targeted (SCoT) polymorphism in toxic and non-toxic accessions of *Jatropha curcas* L. and development of a codominant SCAR marker. Plant science 207:117-27. <https://doi.org/10.1016/j.plantsci.2013.02.013>

32. Nayak PP, Prakash J (2018) Molecular characterization of the indigenous stingless bees (*Tetragonula* spp. Complex) using ISSR marker from Southern Peninsular India. *Neotropical entomology*. 47: 106-17. <https://doi.org/10.1007/s13744-017-0538-7>
33. Odongo FO, Oyoo ME, Wasike V, Owuochi JO, Karanja L, Korir P (2015) Genetic diversity of Bambara groundnut (*Vigna subterranea* (L.) verdc.) landraces in Kenya using microsatellite markers. *African Journal of Biotechnology* 14(4):283-91. <http://doi.org/10.5897/AJB2014.14082>
34. Rezazadeh E, Aliabadian M, Darvish J, Ahmadzadeh F (2020) Diversification and evolutionary history of brush-tailed mice, Calomyscidae (Rodentia), in southwestern Asia. *Organisms Diversity & Evolution* 20:155-70. <https://doi.org/10.1007/s13127-019-00426-y>
35. Ronikier M, Zalewska-Gałosz J (2014) Independent evolutionary history between the Balkan ranges and more northerly mountains in *Campanula alpina* sl (Campanulaceae): genetic divergence and morphological segregation of taxa. *Taxon*. 63(1):116-31. <https://doi.org/10.12705/631.4>
36. Sanghvi GV, Dave GS (2017) Molecular markers in plant biotechnology. In "Plant Biotechnology" pp. 415-453. Apple Academic Press
37. Rajaei Sh H, Rödder D, Weigand AM, Dambach J, Raupach MJ, Wägele JW (2013) Quaternary refugia in southwestern Iran: insights from two sympatric moth species (Insecta, Lepidoptera). *Organisms Diversity & Evolution* 13:409-23. <https://doi.org/10.1007/s13127-013-0126-6>
38. Sheligl HQ (1986) The utilization of organic acids by chlorella light. *Planta Journal* 47(51):51-3
39. Sirivibulkovit K, Nouanthavong S, Sameenoi Y (2018) Based DPPH assay for antioxidant activity analysis. *Analytical sciences* 34(7):795-800. <https://doi.org/10.2116/analsci.18P014>
40. Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139(1):457-62. <https://doi.org/10.1093/genetics/139.1.457>
41. Sorkheh K, Amirbakhtiar N, Ercisli S (2016) RETRACTED ARTICLE: Potential Start Codon Targeted (SCoT) and Inter-retrotransposon Amplified Polymorphism (IRAP) Markers for Evaluation of Genetic Diversity and Conservation of Wild Pistacia Species Population. *Biochemical genetics* 54:368-87. <https://doi.org/10.1007/s10528-016-9725-1>
42. Strauss SH, Lande R, Namkoong G (1992) Limitations of molecular-marker-aided selection in forest tree breeding. *Canadian Journal of Forest Research* 22(7):1050-61. <https://doi.org/10.1139/x92-140>
43. Sujka K, Koczoń P, Ceglińska A, Reder M, Ciemniowska-Żytewicz H (2017) The application of FT-IR spectroscopy for quality control of flours obtained from polish producers. *Journal of analytical methods in chemistry*. 2017: 4315678. <https://doi.org/10.1155/2017/4315678>
44. Tang JS, Ma M (2020) Genetic diversity and genetic differentiation of invasive weed *Xanthium italicum* in China. *Comptes Rendus. Biologies* 343(1):63-72. <https://doi.org/10.5802/crbio.7>
45. Tivang JG, Nienhuis J, Smith OS (1994) Estimation of sampling variance of molecular marker data using the bootstrap procedure. *Theoretical and Applied Genetics* 89:259-64. <https://doi.org/10.1007/BF00225151>
46. Wang C, Qiao A, Fang X, Sun L, Gao P, Davis AR, Liu S, Luan F (2019) Fine mapping of lycopene content and flesh color related gene and development of molecular marker-assisted selection for flesh color in watermelon (*Citrullus lanatus*). *Frontiers in Plant Science*. 10:1240. <http://doi.org/10.3389/fpls.2019.01240>
47. Wang Y, Sun F, Lin W, Zhang S (2022) AC-PCoA: Adjustment for confounding factors using principal coordinate analysis. *PLoS Computational Biology* 18(7):e1010184. <https://doi.org/10.1371/journal.pcbi.1010184>
48. Warghat AR, Bajpai PK, Murkute AA, Sood H, Chaurasia OP, Srivastava RB (2012) Genetic diversity and population structure of *Dactylorhiza hatagirea* (Orchidaceae) in cold desert Ladakh region of India. *Journal of Medicinal Plants Research* 6(12):2388-95. <https://doi.org/10.5897/JMPR11.1007>

Figures



Figure 1

Natural distribution area of *Corylus avellana* in Lesser Caucasus according to Germain1992; Leinmann et al. 2013. The sampling locations in this study are related to the Iranian habitats of Hazelnut as follows: A- Fandaghlo (Ardebil). B- Essygran (Ardebil); C- Abibigloo (Ardebil); D- Miyaneh (East Azerbaijan); E- Arsbaran (East Azerbaijan); F- Astara (Gilan); and G- Astara (Gilan).

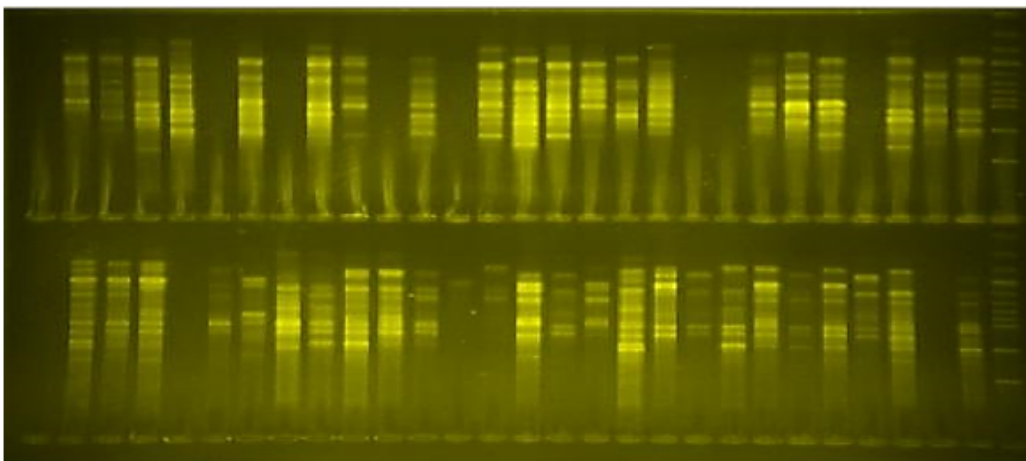


Figure 2

Banding profile resulted by SCoT3 primers in genotypes of *C. avellana*.

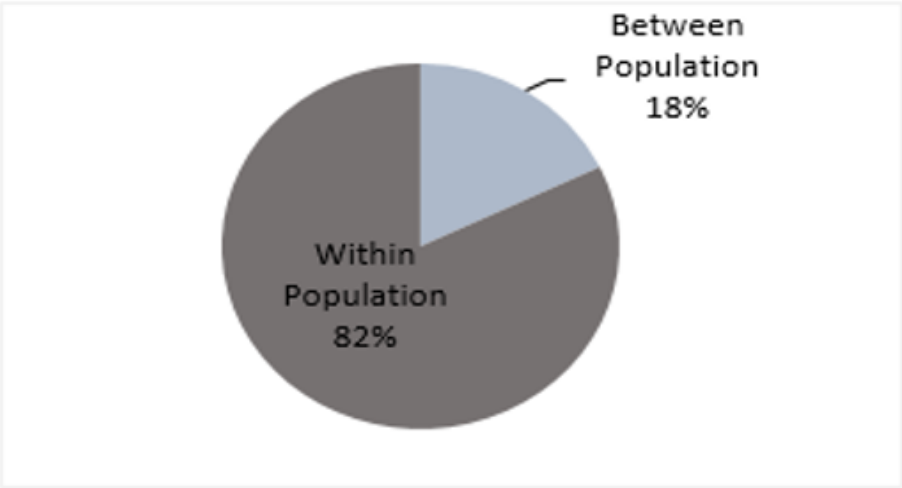


Figure 3

Allelic frequency and genetic diversity indices relating the hazelnut populations.

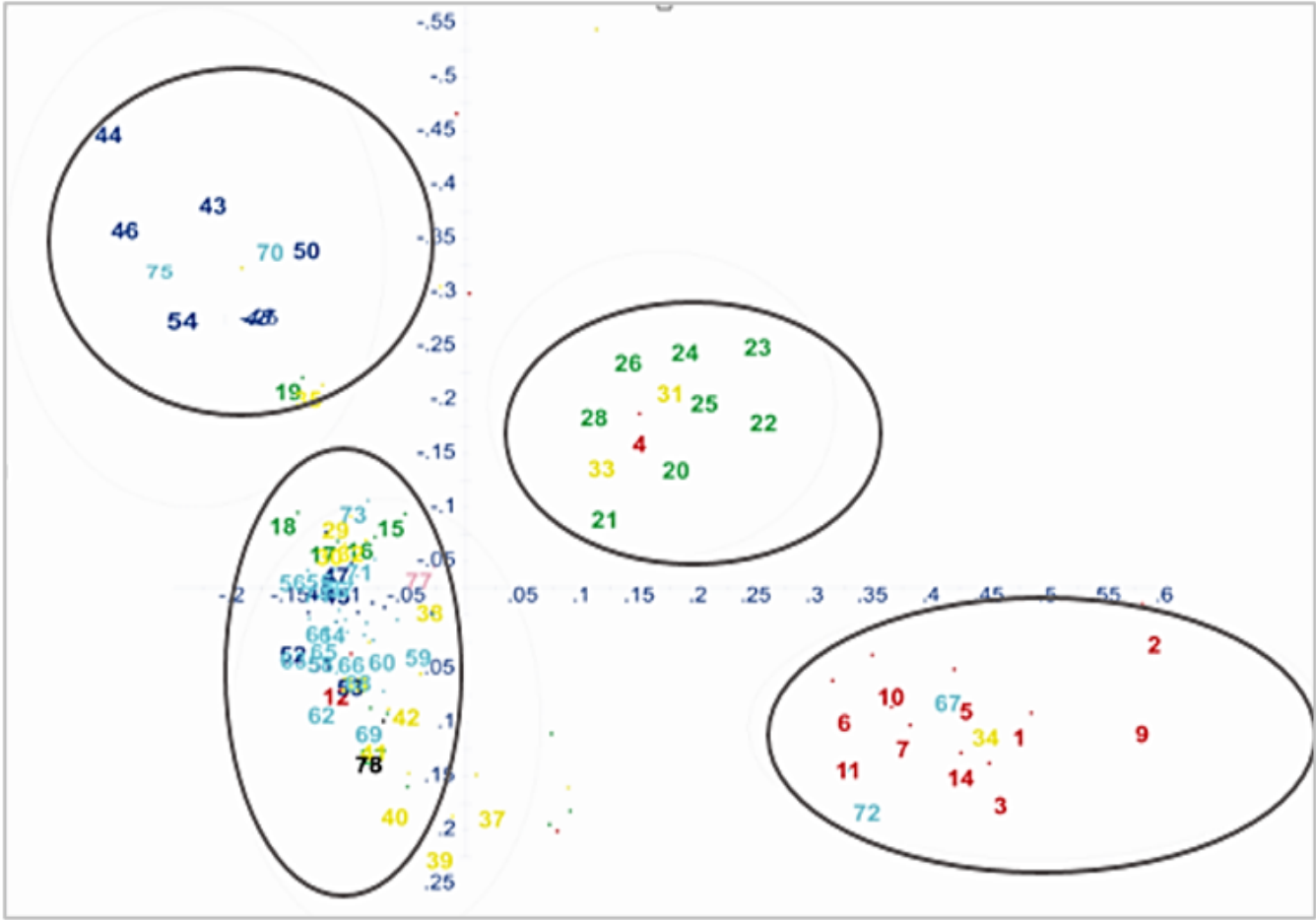


Figure 4

PCoA two-dimensional diagram in 78 hazelnut genotypes

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryTable1.docx](#)