CrossMark

ORIGINAL ARTICLE

Molecular genetic diversity of the Turkish national hazelnut collection and selection of a core set

Süleyman Can Öztürk[1] · Hüseyin İrfan Balık[2] · Selda Kayalak Balık[2] ·
Gökhan Kızılcı[2,3] · Ömür Duyar[2,4] · Sami Doğanlar[1] · Anne Frary[1]

Abstract European hazelnut (*Corylus avellana* L.) is an economically and nutritionally important nut crop with wild and cultivated populations found throughout Europe and in parts of Asia. This study examined the molecular genetic diversity and population structure of 402 genotypes including 143 wild individuals, 239 landraces, and 20 cultivars from the Turkish national hazelnut collection using simple sequence repeat (SSR) markers. A total of 30 SSR markers yielded 407 polymorphic fragments. Diversity analysis of the Turkish hazelnut genotypes indicated that they fell into three subpopulations according to ad hoc statistics and neighbor-joining algorithm. Although all cultivars clustered together, they overlapped with the wild accessions and landraces. Thus, the dendrogram, principal coordinate, and population structure analyses suggest that they share the same gene pool. A total of 78 accessions were selected as a core set to encompass the molecular genetic and morphological diversity present in the national collection. This core set should have priority in preservation efforts and in trait characterization.

Communicated by Z. Kaya

✉ Anne Frary
annefrary@iyte.edu.tr

[1] Department of Molecular Biology and Genetics, Izmir Institute of Technology, Izmir, Turkey

[2] Hazelnut Research Institute, Giresun, Turkey

[3] General Directorate of Agricultural Research and Policies, Ankara, Turkey

[4] Directorate of Provincial Food Agriculture and Livestock, Ordu, Turkey

Introduction

European hazelnut (*Corylus avellana*) is one of the most economically important nut species and is cultivated in parts of Europe, Anatolia, and the USA (Boccacci et al. 2006). Turkey is the world's main hazelnut producer with 450,000 tons grown on 701,141 ha, accounting for 61% of world production (FAO 2014). Approximately 163,000 tons of hazelnuts is exported from Turkey each year (FAO 2013). The most suitable climatic conditions for hazelnut production in Turkey are in the Black Sea region where Turkish cultivars such as 'Tombul,' 'Palaz,' 'Mincane,' 'Cakıldak,' and 'Sivri' are grown (Koksal 2002). In addition, the area contains many wild hazelnut trees and landraces.

Turkish hazelnut germplasm has been systematically collected and grown at the Hazelnut Research Institute in Giresun since its establishment in 1936 with substantial additions made to the collection from 1969 to 1972 (Caliskan and Cetiner 1997, H.İ. Balik personal communication). The collection currently contains 430 accessions grown at the institute's orchard and includes both selected and bred cultivars, landraces, and wild accessions that were found near commercial orchards. Wild accessions and landraces were established in the research institute's orchard by transfer of side shoots from naturally occurring trees. Germplasm collections are valuable reservoirs of genetic diversity. Diversity prevents catastrophic losses due to biotic and abiotic stresses and is also necessary for improvement of hazelnut to meet future climate, stress, grower, and consumer demands. In addition to preserving germplasm, the institute has characterized the material, with special emphasis on the cultivars, for morphological

and phenological traits (Caliskan and Cetiner 1997; Koksal 2002). However, it has not yet examined all of the accessions for their molecular genetic diversity. This is necessary to understand the genetic relationships among individuals, information which is especially valuable when selecting parents for hybrid breeding, a relatively recent area of interest to the institute (Balik et al. 2015). Both molecular and morphological data are also useful in selecting a core set of germplasm. A core set is a subset of germplasm that encompasses the maximum genetic diversity in a minimum number of accessions from the entire collection (Frankel 1984; Brown 1995). Core set selection can help to prioritize preservation and propagation of the collection as well as provide a reasonable number of diverse samples for the measurement of characters and properties that are expensive, time-consuming, and/or laborious. Moreover, core sets provide ideal material for association mapping of traits in tree species like hazelnut.

Simple sequence repeat (SSR) markers are the most commonly used marker for assessing molecular diversity in hazelnut because they are multi-allelic and have a high level of reproducibility. A total of 300 genomic SSR markers were developed for hazelnut (Boccacci et al. 2005; Bassil et al. 2005, 2013; Mehlenbacher et al. 2006; Gurcan and Mehlenbacher, 2010; Gurcan et al. 2010a). In addition, Boccacci et al. (2015) investigated 1282 non-redundant SSRs in ESTs from the family Betulaceae, but only 20 polymorphic SSR markers were developed. More recently, Colburn et al. (2017) developed 111 polymorphic SSRs from the transcriptome sequences of 'Jefferson' hazelnut. The SAFENUT European Commission Action analyzed the genetic diversity of European hazelnuts using SSR markers (Bacchetta et al. 2014). As part of this project, SSR markers were used to study the genetic diversity of hazelnut germplasm in Spain (Boccacci et al. 2005, 2006, 2008) and Southern Europe (Boccacci et al. 2013). In addition, SSR markers were used to detect high levels of diversity in Turkish, Georgian, and Azeri hazelnuts (Gurcan et al. 2010a). These markers were also useful to study the geographic origin and distribution of hazelnut trees (Boccacci et al. 2006; Gökirmak et al. 2009; Gurcan et al. 2010a) and to identify synonymous trees (Gökirmak et al. 2009; Gurcan et al. 2010a; Valentini et al. 2014). In addition, SSR markers were used to construct linkage maps (Gurcan and Mehlenbacher 2010; Gurcan et al. 2010b; Colburn et al. 2017) and to map genes for eastern filbert blight resistance (Sathuvalli et al. 2011), self-incompatibility, and style color (Ives et al. 2014). In another study, 275 F1 hybrids of 'Tonda Gentile delle Langhe' × 'Merveille de Bollwiller' hazelnut trees were used for quantitative trait locus (QTL) identification for traits such as vigor, sucker habit, and time of bud burst (Beltramo et al. 2016). In the most recent study, kernel and nut shape traits were examined in Slovenian hazelnuts and the QTL for these traits were detected using SSR markers and association mapping analysis (Ozturk et al. 2017).

The goal of this research was to analyze the molecular genetic diversity and population structure of 402 hazelnut accessions (143 wild accessions, 239 landraces, and 20 cultivars) in the Turkish national collection using SSR markers. We also selected a core set of the most diverse material for further morphological and biochemical profiling and association mapping analyses. The core collection will be an efficient and economical resource for future hazelnut preservation, characterization, and improvement.

Materials and methods

DNA isolation

A total of 402 hazelnut accessions which represent the diversity of material present in Turkey's Black Sea region was used from the Hazelnut Research Institute, Giresun. This collection contains all 20 Turkish cultivars (Balik et al. 2016) as well as landraces and wild accessions collected by the research institute from Giresun (240 accessions), Ordu (49 accessions), Trabzon (49 accessions), Samsun (4 accessions), Rize (3 accessions), Sinop (2 accessions), and Artvin, Duzce, Kastamonu, and Erzurum (1 accession each) (Suppl. Fig. 1, Suppl. Table 1) (Caliskan and Cetiner, 1997). The remaining 31 accessions were of unknown origin but collected from the Black Sea region. Total genomic DNA was isolated from leaves sampled from individual trees according to Fulton et al. (1995).

SSR amplification

Thirty SSR markers with high levels of polymorphism as reported by Gurcan et al. (2010a) were used for genetic diversity determination (Table 1). For all primer combinations, PCR amplification was performed with 20 ng DNA in a 20-μl reaction containing 10 pmol each primer pair, 200 μm dNTPs, 2 μl 10× Taq polymerase buffer, and 0.6 Unit Taq polymerase. A GeneAmp PCR system 9700 (Perkin Elmer Applied Biosystems) machine was used for PCR amplification. Reaction conditions were denaturation at 94 °C for 30 s; 30 cycles of denaturation at 94 °C, annealing at 55 °C for 30 s, and extension at 72 °C for 30 s; and final extension at 72 °C for 5 min. PCR fragments were separated by capillary electrophoresis using a Fragment Analyzer™ (Applied Biosystems) with the DNF-900 dsDNA Reagent Kit (Advanced Analytical) according to the manufacturer's instructions. Because many of the primer pairs yielded more than two fragments and allelism could not be determined, the individual fragments were scored binomially (presence 1, absence 0).

Table 1 SSR marker sequence and polymorphism information for the Turkish hazelnut accessions

| Primer name	Forward primer (5′ to 3′)	Reverse primer (5′ to 3′)	No. of polymorphic fragments	Ho[a]	PIC[b]
A601	TTACATGGTTCGGCAATGTG	AGATGGGAGCAGAGTGAACTG	11	0.25	0.97
A602	AAGAGTGGGGGTGCACTATG	GGATTCATGCCTGCGATACT	21	0.22	0.99
A604	GCTCCCGAGGACTTCCAG	CCACGACATTTCCCTCTCAG	15	0.37	0.99
A605	CACCCTCAAAACTGTGACGA	TGGGTCGCATTCAATAACAC	9	0.37	0.97
A606	CACCTAGCTTGTTGGTGAAGC	TGACAATAATTAACCCTACA CACTTTG	9	0.24	0.95
A611	CACTAGCCAGCCCCTTTACA	CTGATGCCACAAACACAAGG	10	0.39	0.98
A613	CACACGCCTTGTCACTCTTT	CCCCTTTCACATGTTTGCTT	17	0.34	0.99
A616	CACTCATACCGCAAACTCCA	ATGGCTTTTGCTTCGTTTTG	13	0.34	0.98
A635	GGATCTGTGGTTGGCTTTTTGGTACTAT	TTACCCAATGGATGATGGAC TAGCATT	12	0.33	0.98
A640	TGCCTCTGCAGTTAGTCATCAAATGT AGG	CGCCATATAATTGGGATGCTTGTTG	10	0.39	0.98
B602	AAGAGTGGGGGTGCACTATG	GGATTCATGCCTGCGATACT	14	0.30	0.97
B603	TGGTGGTGATAGGGAAGGAG	TCTTTTCTTCTTCAATCAGACGA	17	0.30	0.97
B606	TCTTGTGGTTTAGCATACTTCTCG	GAAGAAAGCAAGAAGAGAGGAGA	10	0.26	0.95
B612	GCACCTCAAACTCCTTGGAC	CCCAAACACACCCTTAGTGC	20	0.37	0.99
B613	CGCGTTTTGAGTCCCTTTAG	CTACCCGCCTGCGAGAAC	14	0.42	0.99
B625	CGCAAGTCATTGCACATTTT	GTGTGCTGTGCTCCTTTGAA	17	0.36	0.99
B628	AATCCCCTCTAGCCCCATTA	CACAGAATATTTGTAATTAC CACCACA	5	0.17	0.66
B631	TGAAGCAGACAAGCGAATAGC	TTGTGTCTCTTTGTCTTGTAAATCG	13	0.27	0.95
B635	GCATCGCCAAATTATCGTCT	CTTCAACAAATCCAGGATGC	11	0.36	0.99
B640	CTGCATTGATGGATTGGTTG	TTAAGAAAGGTACAAGGGCTCTC	18	0.25	0.99
B641a	CTCCCATGAAATGATTATTCTTAG	CAAGCCATCTGTTTTGCTGA	9	0.30	0.94
B641b	ATATATATAGGCTGTGTGTGTGTGTG	ACAAGCCATCTGTTTTGCTG	18	0.31	0.99
B648	TGAAAGCGCCCAAAACTTAT	CTTGCGTCTTTTTGGAGAGC	17	0.41	0.99
B651	TTTTCTGGAATGTCGCACAG	TCTCCTCCTTCCAACAGTGG	26	0.18	0.98
B652	AGGATGCGTGGTTGTGATTT	TGGAGTAGGGTGATGAGAATGA	22	0.29	0.99
B660	TGTTGTAGCACAACCCTTTCA	TGCTAGCAGCAAATGGCTTA	8	0.39	0.96
B662	CGAAAGATGGACTTCCATGAC	CAAGTTGAGATTCTTCCTGCAA	12	0.35	0.98
B788	TCCCTTTCTCCGTCATCAAC	TCGTCACCGTCACCAGATAA	9	0.44	0.98
B789	GCCACGTCCAGAATCAAAAT	CCTCAGGGCTGAGAAGTTGA	6	0.18	0.77
CAC-B753	AAGGGTTGTTACCCATGCAC	GGTGCATTTAGTGCTTCTGG	13	0.32	0.97

All primers are from Gurcan et al. (2010a)

[a] Observed heterozygosity

[b] Polymorphism information content

Data analysis

PowerMarker software (Liu and Muse 2005) was used to calculate polymorphism information content (PIC) and observed heterozygosity (Ho) values. DARwin 5 software was used to analyze the data with the Dice coefficient (Dice 1945) and unweighted neighbor-joining algorithm (Perrier and Jacquemoud-Collet 2006). This program was also used for principal coordinate analysis (PCoA). Structure 2.3.4 (Pritchard et al. 2000) software was used to determine population structure. Ad hoc statistics were used to determine the

best number of subpopulations (Evanno et al. 2005). The data were evaluated for 2 to 20 subpopulations ($K = 2$ to 20) with 50,000 cycles. Each subpopulation model was tested 10 times with 300,000 iterations per K. The probability change of each group (ΔK) was calculated using the program Structure Harvester (Earl and Von Holdt 2012). The best number of subpopulations was determined from the highest ΔK. Hazelnuts were clustered using a threshold of inferred ancestry ≥ 0.70. Accessions that did not meet this threshold were considered as admixed. A second population structure program, InStruct (Gao et al. 2007), was used to confirm the
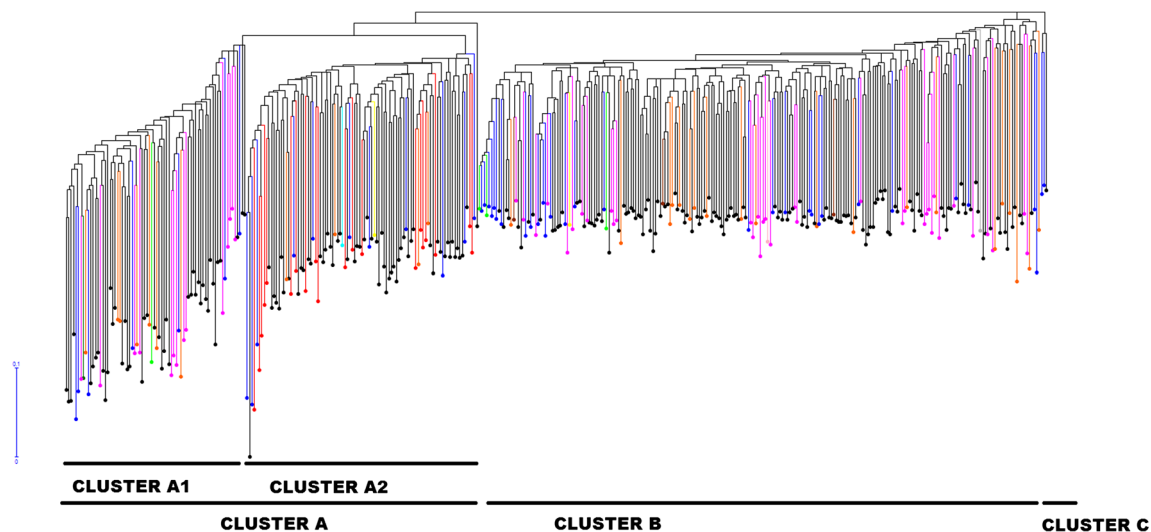
Fig. 1 Unweighted neighbor-joining dendrogram of the 402 hazelnut accessions based on SSR data. Accessions are color coded by origin: cultivar: red, Giresun: black, Ordu: blue, Duzce: light blue, Artvin: gray, Erzurum: light pink, Kastamonu: dark green, Rize: brown, Samsun: green, Sinop: yellow, Trabzon: fuchsia, and unknown: orange

results of Structure and to test $K = 1$. To select core set accessions, the SSR dataset for the hazelnut accessions was analyzed with PowerCore 1.0 software which uses the M (maximization) strategy and a modified heuristic algorithm (Kim et al. 2007). PowerCore software develops a core set by maximizing the number of alleles represented in a minimum number of individuals, thus reducing redundancy.

Results

A total of 402 hazelnut accessions from the Hazelnut Research Institute in Giresun, Turkey, were studied (Suppl. Table 1). The accessions include 20 standard cultivars, 143 wild accessions, and 239 landraces from the Black Sea provinces of Giresun, Ordu, Rize, Trabzon, Samsun, Artvin, Duzce, Sinop, and Erzurum (Caliskan and Cetiner 1997) (Suppl. Fig. 1, Suppl Table 1). Thus, the material included all known Turkish cultivars and a wide selection of naturally occurring accessions.

SSR marker polymorphism

A total of 30 SSR marker primer pairs were used and yielded 407 fragments, 406 (99.8%) of which were polymorphic (Table 1). The average allele number for each SSR marker was 13.6. Marker B651 had the most polymorphic alleles (26) while B628 had the fewest (5). Observed heterozygosity for the markers varied from 0.17 to 0.42. PIC values ranged between 0.66 and 0.99. B628 and B789 were the least polymorphic markers (PIC = 0.66 and 0.77 respectively).

Genetic diversity

The SSR data were used to construct a distance matrix based on the Dice coefficient and to construct a dendrogram of the accessions using the unweighted neighbor-joining algorithm (Fig. 1). The Dice dissimilarity coefficient ranged from 0.10 to 0.84 with a mean of 0.49 for the pairwise comparisons between accessions. Landraces had the highest average dissimilarity coefficient (0.50) while cultivars and wild accessions had lower values (0.47). Materials from Giresun, Trabzon, and Ordu (the provinces with the most trees in the collection) were also compared to cultivars. Accessions from these locations (0.48–0.49) were found to be only slightly more diverse than cultivars (0.47) (Table 2).

A Mantel test showed a high correlation between the distance matrix and dendrogram ($r = 0.97$). The dendrogram consisted of three clusters: A, B, and C. Wild accessions and landraces were found in all three clusters while cultivars

Table 2 Average Dice coefficient dissimilarity values for cultivars, landraces, and wild hazelnut accessions as determined with SSR markers. Wild accessions and landraces were also combined and classified by origin for those collected in Giresun, Trabzon, and Ordu (the most common locations)

Type/origin	No. of accessions	Min.	Max.	Mean
Cultivar	20	0.26	0.65	0.47
Wild	143	0.12	0.77	0.47
Landrace	239	0.10	0.84	0.50
Giresun	240	0.12	0.82	0.49
Trabzon	49	0.10	0.76	0.48
Ordu	49	0.15	0.72	0.49
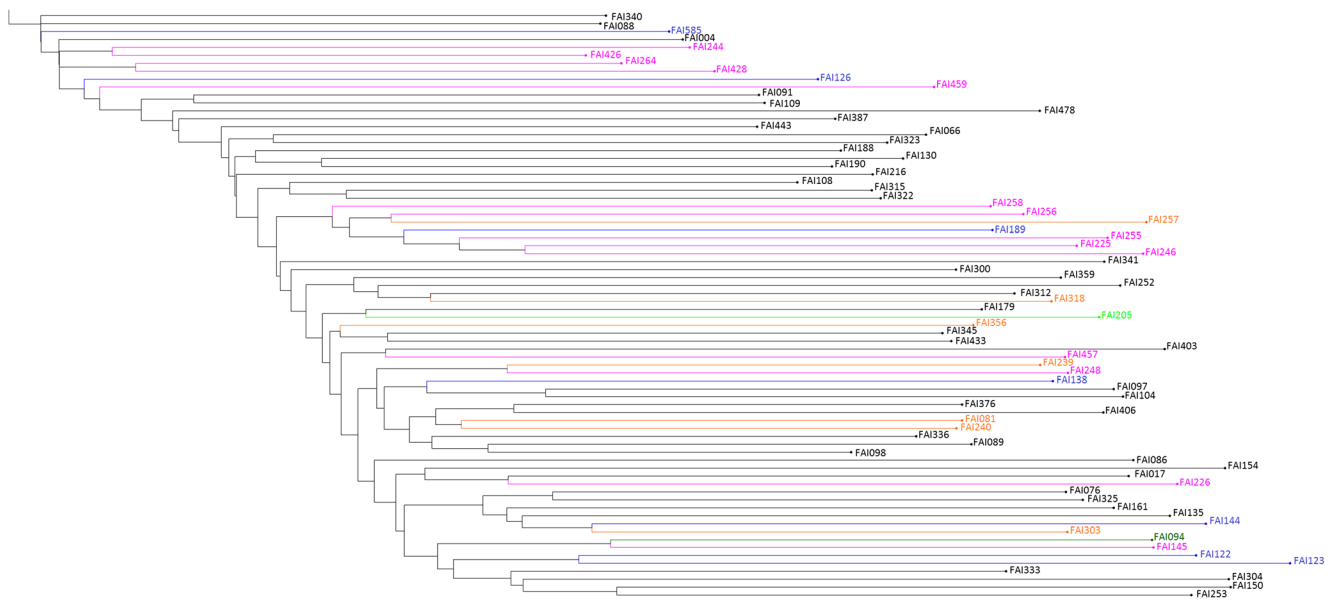
Fig. 2 Cluster A1 of the dendrogram. Accessions are color coded by origin: Giresun: black, Ordu: blue, Kastamonu: dark green, Rize: brown, Samsun: green, Trabzon: fuchsia, and unknown: orange

were limited to cluster A. Cluster A contained 169 accessions in subclusters A1 and A2 which had 74 and 95 accessions, respectively (Figs. 2 and 3). Cluster A1 contained accessions from Giresun (44 accessions), Trabzon (14), Ordu (7), Samsun (1), Kastamonu (1), and unknown places (7). Cluster A2 contained all of the cultivars (20 accessions) which were distributed among the wild accessions and

landraces from Giresun (58), Ordu (11), Sinop (1), Duzce (1), and unknown places (4). Cluster B was the largest with 230 accessions. It contained accessions from Giresun (137 accessions), Trabzon (35), Ordu (29), Rize (3), Samsun (3), Sinop (1), Artvin (1), Erzurum (1), and unknown places (20) (Fig. 4). Cluster C had only three accessions: one from Giresun and two from Ordu.
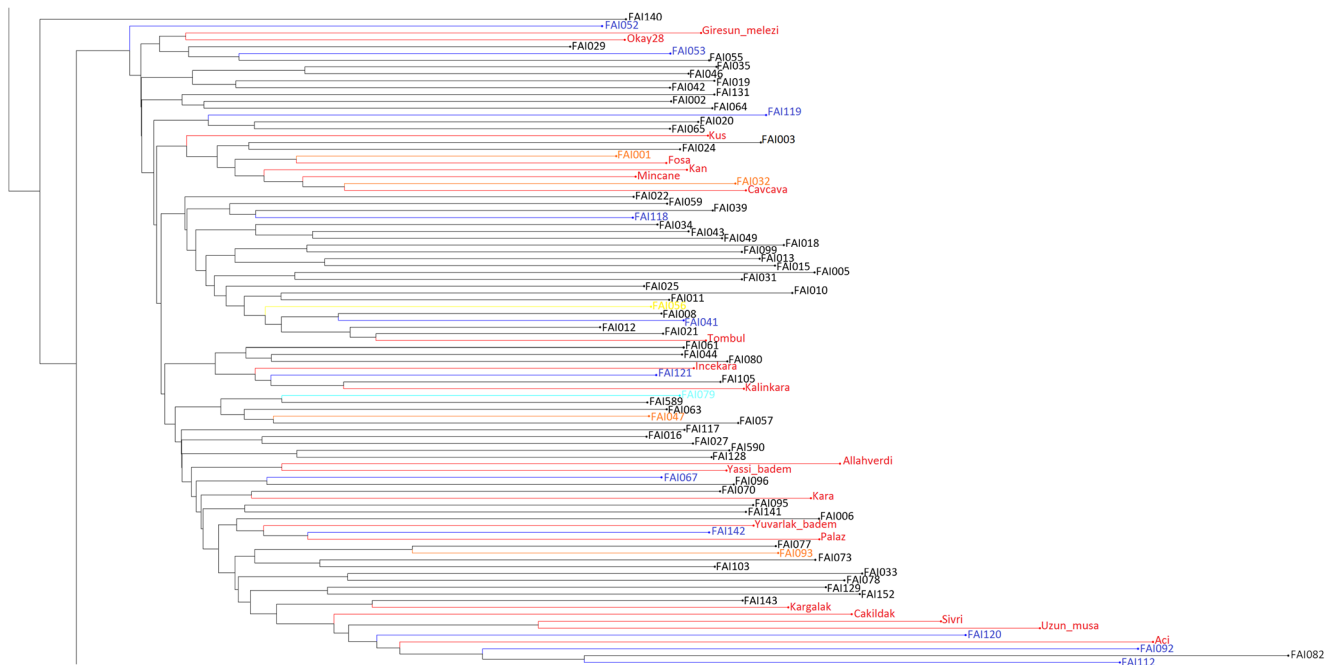
Fig. 3 Cluster A2 of the dendrogram. Accessions are color coded by origin: cultivar: red, Giresun: black, Ordu: blue, Duzce: light blue, Sinop: yellow, and unknown: orange

Fig. 4 Cluster B of the dendrogram. Accessions are color coded by origin: Giresun: black, Ordu: blue, Artvin: gray, Erzurum: light pink, Rize: brown, Samsun: green, Sinop: yellow, Trabzon: fuchsia, and unknown: orange
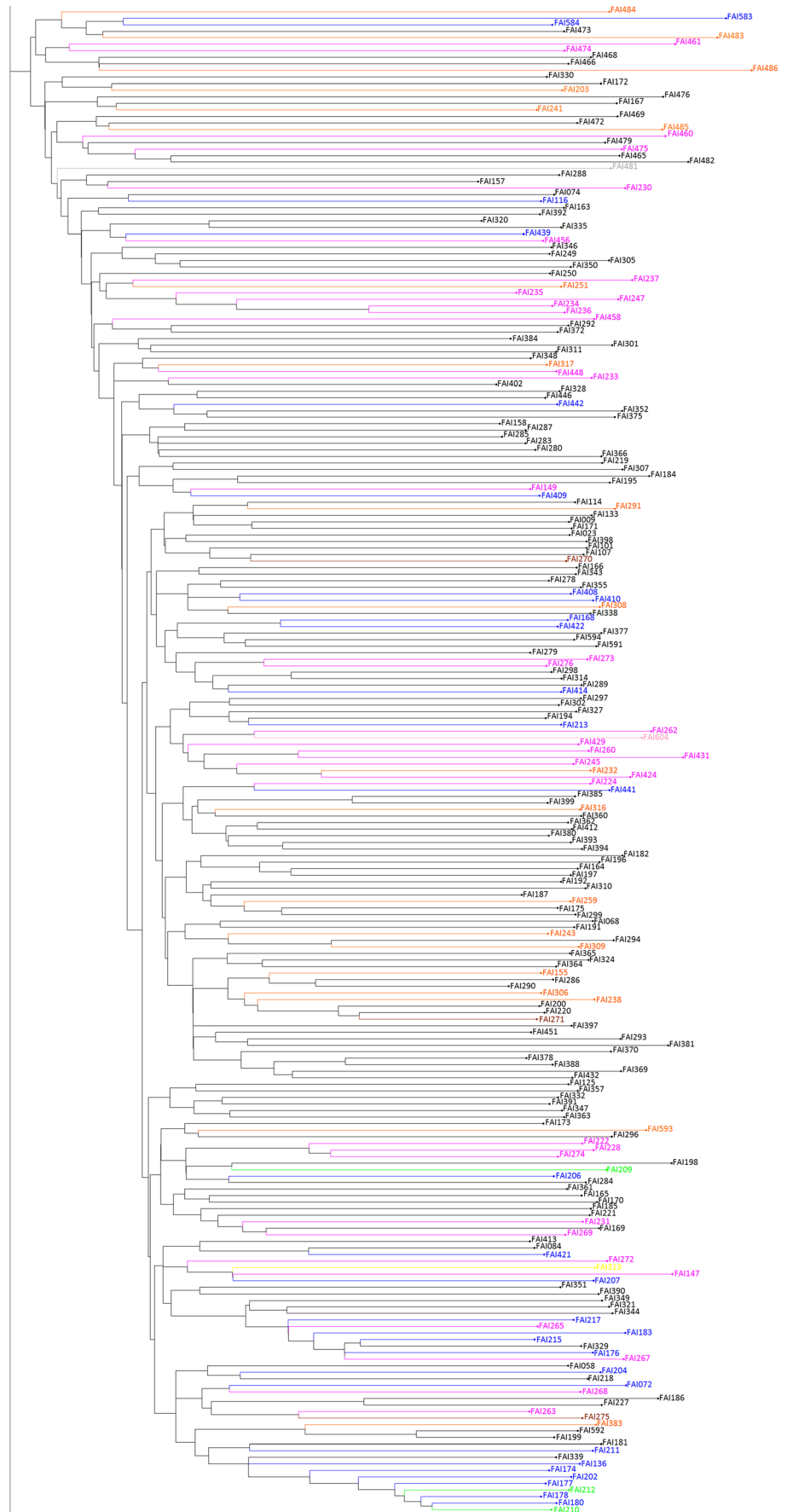
Geographical clustering was limited; however, accessions from Trabzon were found only in clusters A1 and B while those from Giresun and Ordu were distributed throughout clusters A and B. Minor clustering of accessions from the same or neighboring locations was also observed. For example, seven accessions from Ordu and two from Samsun formed a small but distinct group in cluster B.

Principal coordinate analysis of the SSR dataset did not show clear separation of the wild accessions and landraces from the cultivars, but all cultivars were clustered in the lower-left quadrant of the two-dimensional PCoA plot (Fig. 5). The clusters in the PCoA analysis correspond to the clusters in the dendrogram analysis.

Population structure

Population structure analysis indicated that the SSR data were best described by a model containing two subpopulations ($K = 2$). Thus, a membership threshold of $P \geq 0.7$ was used (Suppl. Fig. 2). In this way, 139 (35%) accessions were assigned to subpopulation 1, 185 (46%) accessions were assigned to subpopulation 2, while 78 (19%) were admixed (Suppl. Fig. 3). All but five Turkish hazelnut cultivars belonged to subpopulation 1. The exceptions were 'Fosa,' 'Giresun Melezi,' 'Incekara,' 'Kan,' and 'Okay28,' all of which had admixed ancestry (Suppl. Table 1). Subpopulation 1 included accessions from Giresun (100), Ordu (15), Trabzon (12), Samsun (1), Kastamonu (1), Duzce (1), and unknown places (9). Similarly, subpopulation 2 had accessions from Giresun (113), Ordu (26), Trabzon (24), Rize (3), Samsun (3), Sinop (1), Erzurum (1), and unknown places (14). The 78 individuals which were admixed included accessions from Giresun (47), Trabzon (13), Ordu (8), Sinop (1), Artvin (1), and unknown places (8). When the population structure results were compared with the dendrogram and PCoA plot, cluster A corresponded to subpopulation 1 plus 30 admixed accessions and cluster B corresponded to subpopulation 2 plus 45 admixed accessions. Cluster C had only admixed accessions.
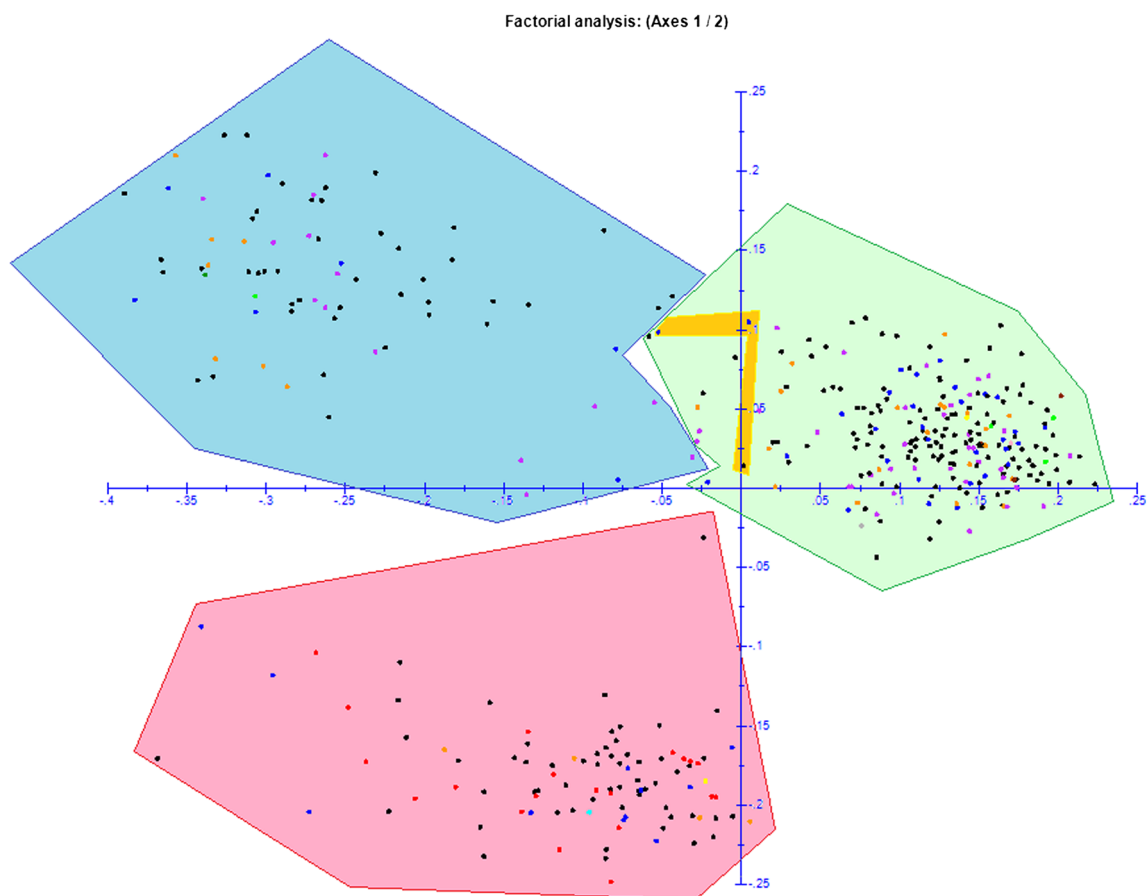
Fig. 5 Principal coordinate analysis of hazelnut accessions according to the first two eigen vectors which explained 12.8 and 6.0% of the variance, respectively. Cultivars and A1 accessions are clustered in the red area, A2 accessions are clustered in the blue area, B accessions are clustered in the green area, and C accessions are found in the yellow area. Accessions are color coded by origin: cultivar: red, Giresun: black, Ordu: blue, Duzce: light blue, Artvin: gray, Erzurum: light pink, Kastamonu: dark green, Rize: brown, Samsun: green, Sinop: yellow, Trabzon: fuchsia, and unknown: orange

Core set selection

The SSR data were analyzed to select a core set of Turkish hazelnut accessions representing the diversity of the entire collection. Five cultivars and 29 hazelnuts were chosen for the core set based on their high levels of molecular genetic diversity. The remaining 15 cultivars were also added to the core set because of their economic importance and/or distinct features which merited giving them a name (Koksal 2002; Balik et al. 2016). An additional 29 hazelnuts with interesting phenotypic traits such as unusual shape, color, and size of the kernel and fruit were also included (Caliskan and Cetiner 1997, H.I. Balik personal communication). Thus, 78 individuals (19% of the collection) were chosen to represent the molecular genetic and morphological diversity of the entire collection (Table 3). The average genetic dissimilarity of the core set based on SSR markers was 0.53. The core set contained accessions from Giresun (45; 25 accessions and 20 cultivars), Trabzon (12), Ordu (9), Sinop (1), Artvin (1), Duzce (1), and unknown places (9) (Table 3). In terms of population structure, the core collection contained 38 accessions from subpopulation 1, 19 accessions from subpopulation 2, and 21 admixed accessions.

Discussion

Marker polymorphism

The 30 SSR markers provided sufficient polymorphism in the 402 hazelnut accessions with 13.6 fragments per marker. This value is slightly higher than those obtained by others using SSR markers in hazelnut which varied from 3 to 10.6 fragments per primer pair (Bassil et al. 2005; Boccacci et al. 2005, 2008; Gökirmak et al. 2009; Boccacci and Botta 2010; Gurcan and Mehlenbacher, 2010; Gurcan et al. 2010a, b; Campa et al. 2011; Bassil et al. 2013; Ozturk et al., 2017). In previous work, a higher annealing temperature was used for these SSR primers (Gurcan et al., 2010a) (60 rather than 55 °C). In addition, some studies used only single- or low-copy SSRs (Gurcan et al. 2010b; Sathuvalli et al. 2011; Bassil et al. 2013; Martins et al. 2013), thereby limiting marker polymorphism. We chose a more permissive temperature to allow amplification of additional fragments. This reduced the total number of markers needed to be analyzed, increased the efficiency, and decreased the cost of the work. Such practical measures are often required for characterization of large germplasm collections. The greater number of polymorphic fragments could also be partially due to the large number and breadth of genetic material used in this study as most other studies limited themselves to cultivated material.

Table 3 Accessions in the core set of Turkish hazelnuts selected based on SSR data and morphology. Subsets were selected based on molecular data (A), morphological features (B), and identity as a named cultivar (C). Subpopulation assignment for each accession is given in parentheses

Subset A	Subset B	Subset C
Aci (1)	FAI056 (admixed)	Allahverdi (1)
Giresun Melezi (admixed)	FAI126 (1)	Cavcava (1)
Kan (admixed)	FAI137 (admixed)	Cakıldak (1)
Kargalak (1)	FAI174 (2)	Fosa (admixed)
Mincane (1)	FAI177 (2)	Incekara (admixed)
FAI005 (1)	FAI225 (1)	Kalınkara (1)
FAI018 (1)	FAI241 (admixed)	Karafındık (1)
FAI032 (1)	FAI248 (1)	Kus (1)
FAI065 (admixed)	FAI265 (2)	Okay28 (admixed)
FAI079 (1)	FAI318 (1)	Palaz (1)
FAI081 (1)	FAI324 (2)	Sivri (1)
FAI093 (1)	FAI333 (1)	Tombul (1)
FAI096 (1)	FAI351 (2)	Uzunmusa (1)
FAI112 (1)	FAI388 (2)	Yassıbadem (1)
FAI144 (1)	FAI406 (1)	Yuvarlakbadem (1)
FAI145 (1)	FAI422 (2)	
FAI150 (1)	FAI457 (1)	
FAI161 (1)	FAI458 (admixed)	
FAI172 (admixed)	FAI459 (1)	
FAI228 (2)	FAI461 (admixed)	
FAI279 (2)	FAI469 (admixed)	
FAI289 (2)	FAI472 (admixed)	
FAI302 (2)	FAI474 (admixed)	
FAI314 (2)	FAI478 (1)	
FAI316 (2)	FAI479 (admixed)	
FAI349 (2)	FAI481 (admixed)	
FAI408 (2)	FAI485 (admixed)	
FAI409 (2)		
FAI429 (2)		
FAI451 (2)		
FAI460 (admixed)		
FAI466 (admixed)		
FAI484 (admixed)		
FAI590 (1)		

Diversity and population structure of wild and cultivated hazelnuts

Hazelnut is a wind-pollinated species and has a self-incompatible mating system; thus, genetic diversity is expected to be high in naturally occurring plants. In Turkey, hazelnuts are clonally propagated using rooted suckers. In this way, trees which have desirable allele combinations are preserved and kept in the heterozygous condition. Average diversity of the hazelnut cultivars was similar to that of the entire collection, 0.47 and 0.49, respectively. Similarly high levels of

genetic diversity were observed in cultivars and wild accessions from Spain (Campa et al. 2011), Portugal (Martins et al. 2015), and Slovenia (Ozturk et al. 2017). Most of the accessions were collected in Giresun which is an area with extensive hazelnut cultivation and production. The high level of diversity in this region may be related to its high density of trees which allowed cross-pollination of cultivars with nearby landraces and wild trees thereby resulting in "new" material which was collected by the Hazelnut Research Institute. The importance of Giresun to the hazelnut industry in Turkey is reflected in the fact that the quality of Turkish hazelnuts is classified as Giresun (premium) or Levant (secondary). 'Tombul' is the most well-known Turkish cultivar of Giresun quality with both national and international reputations (Alasalvar et al. 2003). 'Tombul' and other "Giresun-quality" hazelnuts were clonally propagated and distributed to other areas along the Black Sea coast.

Dendrogram and PCoA analyses of the hazelnut collection indicated that cultivars were loosely clustered but not genetically distinct from landraces and wild material. In contrast, a clear separation between wild and cultivated accessions was observed in materials from Spain, Portugal, and Slovenia (Campa et al. 2011; Martins et al. 2014, 2015; Ozturk et al., 2017). The difference between these studies and ours may lie in the fact that reference cultivars were used in the other studies. Most of these references cultivars were not of local origin with the exception of some Spanish-Italian cultivars examined in the work of Campa et al. (2011) which focused on wild and local materials from northern Spain. Thus, the gene pools of the cultivars and wild accessions would not be expected to overlap. In contrast, the current work examined only Turkish cultivars, all but three ('Allahverdi,' 'Giresun Melezi,' and 'Okay28') of which originated from selection and cultivation of formerly wild individuals. 'Giresun Melezi' and 'Okay28' are new cultivars developed from 'Kargalak' and 'Tombul' hybrids (Balik et al. 2015). Thus, most of the materials have a common gene pool. Because hazelnut trees from Giresun were the source of most of the cultivars and other genetic resources growing in the region, clear separation of accessions by location was not observed. However, there was minor clustering of hazelnut trees from geographically close regions such as those from Samsun and Ordu. In addition, accessions which were collected from same valleys tend to be in the same cluster. For example, accessions from the western part of Ordu were collected from the same valley and clustered together. In the same way, accessions from eastern valleys of Trabzon province clustered together.

Population structure analysis was consistent with the dendrogram and PCoA results. For example, 100% of the subpopulation 1 genotypes were located in dendrogram cluster A and 100% of subpopulation 2 genotypes were in dendrogram cluster B. Wild material and landraces fell into subpopulation 1 (33%), subpopulation 2 (49%), and the admixed group (19%). The majority of cultivars fell into subpopulation 1 (75%) with the exception of the five admixed cultivars. Two of these five admixed cultivars were developed by hybridization; therefore, it is not surprising that they have admixed ancestry. Admixed accessions are also the result of natural hybrids due to cross-pollination.

Core set selection

A core set of hazelnut accessions was selected using the SSR data, and it was found that the molecular genetic diversity of the entire collection (including all alleles) was encompassed by just 8.4% of the accessions: 29 accessions and five cultivars. Similarly, in the SAFENUT project, 6.5% of 306 accessions were chosen as a core set to cover the genetic diversity in different characters (Bacchetta et al. 2014). Of course, molecular genetic diversity is not the sole parameter by which core sets should be selected. Morphological diversity is also an important criterion, and Turkey has phenotypically diverse hazelnut resources. In addition, tree yield and quality are traits that must be preserved in a core set. For these reasons, accessions with unique phenotypes and the remaining 15 cultivars were included in the core set, thus maintaining important characters and allele combinations. The core set had representation from different geographical locations and each of the subpopulations (1, 2, and admixed). Such core sets are important in prioritizing germplasm conservation and maintenance. This is especially crucial in long-lived tree crops like hazelnut for which wild populations are under threat from abiotic and biotic stresses and deforestation (Bacchetta et al. 2014). Thus, our work represents an important step toward the selection of genetically diverse wild hazelnut material that should be preserved for future generations. The molecular genetic data presented here can be combined with morphological data to ensure the future diversity and improvement of hazelnut in Turkey and other regions.

Compliance with ethical standards

Data archiving statement   Data will be available at http://plantmolgen. iyte.edu.tr/data/ upon publication.

References

Alasalvar C, Shahidi F, Liyanapathirana CM, Ohshima T (2003) Turkish tombul hazelnut (*Corylus avellana* L.). 1. Compositional characteristics. J Agric Food Chem 51:3790–3796

Bacchetta L, Avanzato D, Di Giovanni B, Botta R, Boccacci P, Drogoudi P, Metzidakis I, Rovira M, Sarraquigne JP, Silva AP, Solar A, Spera

D (2014) The reorganisation of European hazelnut genetic resources in the SAFENUT (AGRI GEN RES) project. Acta Hort 1052:67–74

Balik HI, Balik SK, Okay AN (2015) Yeni Fındık Çeşitleri (Okay 28 ve Giresun Melezi). Harran Tarım ve Gıda Bilimleri Dergisi 19:104–109

Balik HI, Balik SK, Beyhan N, Erdogan V (2016) Hazelnut cultivars, ISBN 978-605-137-559-5

Bassil NV, Botta R, Mehlenbacher SA (2005) Additional microsatellite markers of the European hazelnut. Acta Hort 686:105

Bassil N, Boccacci P, Botta R, Postman J, Mehlenbacher SA (2013) Nuclear and chloroplast microsatellite markers to assess genetic diversity and evolution in hazelnut species, hybrids and cultivars. Gen Res Crop Evol 60:543–568

Beltramo C, Valentini N, Portis E, Marinoni DT, Boccacci P, Prando MAS, Botta R (2016) Genetic mapping and QTL analysis in European hazelnut (Corylus avellana L.) Mol Breed 36:1–17

Boccacci P, Botta R (2010) Microsatellite variability and genetic structure in hazelnut (Corylus avellana L.) cultivars from different growing regions. Sci Hortic 124:128–133

Boccacci P, Akkak A, Bassil NV, Mehlenbacher SA, Botta R (2005) Characterization and evaluation of microsatellite loci in European hazelnut (Corylus avellana L.) and their transferability to other Corylus species. Mol Ecol Notes 5:934–937

Boccacci P, Akkak A, Botta R (2006) DNA typing and genetic relationship among European hazelnut (Corylus avellana L.) cultivars using microsatellite markers. Genome 49:598–611

Boccacci P, Botta R, Rovira M (2008) Genetic diversity of hazelnut (Corylus avellana L.) germplasm in northeastern Spain. Hort Sci 43:667–672

Boccacci P, Aramini M, Valentini N, Bacchetta L, Rovira M, Drogoudi P, Silva AP, Solar A, Calizzano F, Erdogan V, Cristoferi V, Ciarmiello LF, Contessa C, Ferreira JJ, Marra FP, Botta R (2013) Molecular and morphological diversity of on-farm hazelnut (Corylus avellana L.) landraces from southern Europe and their role in the origin and diffusion of cultivated germplasm. Tree Genet. Genomes 9:1465–1480

Boccacci P, Beltramo C, Prando MS, Lembo A, Sartor C, Mehlenbacher SA, Botta R, Marinoni DT (2015) In silico mining, characterization and cross-species transferability of EST-SSR markers for European hazelnut (Corylus avellana L.) Mol Breed 35:1–14

Brown AHD (1995) The core collection at the crossroads. In: Hodgkin T, Brown AHD, van Hintum TJL, Morales EAV (eds) Core collections of plant genetic resources. Wiley, London, pp 3–19

Caliskan T, Cetiner E (1997) Characterization studies on some hazelnut cultivars and types. In IV International Symposium on Hazelnut 445:2–12

Campa A, Trabanco N, Pérez-Vega E, Rovira M, Ferreira JJ (2011) Genetic relationship between cultivated and wild hazelnuts (Corylus avellana L.) collected in northern Spain. Plant Breed 130:360–366

Colburn BC, Mehlenbacher SA, Sathuvalli VR (2017) Development and mapping of microsatellite markers from transcriptome sequences of European hazelnut (Corylus avellana L.) and use for germplasm characterization. Mol Breed 37:16

Dice LR (1945) Measures of the amount of ecologic association between species. Ecology 26:297–302

Earl DA, von Holdt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Food and Agriculture Organization of the United Nations, FAOSTAT (2013) http://www.fao.org/faostat/en/#data/TP Accessed 10.01. 2017

Food and Agriculture Organization of the United Nations, FAOSTAT (2014) http://www.fao.org/faostat/en/#data/QC Accessed 10.01. 2017

Frankel OH (1984) Genetic perspectives of germplasm conservation. In: Arber W, Llimensee K, Peacock WJ, Starlinger P (eds) Genetic manipulation: impact on man and society. Cambridge University Press, Cambridge, pp 161–170

Fulton TM, Chunwongse J, Tanksley SD (1995) Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Mol Biol Rep 13:207–207

Gao H, Williamson S, Bustamante CD (2007) A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. Genetics 176:1635–1651

Gökirmak T, Mehlenbacher SA, Bassil NV (2009) Characterization of European hazelnut (Corylus avellana L.) cultivars using SSR markers. Genet Resour Crop Evol 56:147–172

Gurcan K, Mehlenbacher SA (2010) Development of microsatellite marker loci for European hazelnut (Corylus avellana L.) from ISSR fragments. Mol Breed 26:551–559

Gurcan K, Mehlenbacher SA, Botta R, Boccacci P (2010a) Development, characterization, segregation, and mapping of microsatellite markers for European hazelnut (Corylus avellana L.) from enriched genomic libraries and usefulness in genetic diversity studies. Tree Genet Genomes 6:513–531

Gurcan K, Mehlenbacher SA, Erdogan V (2010b) Genetic diversity in hazelnut (Corylus avellana L.) cultivars from Black Sea countries assessed using SSR markers. Plant Breed 129:422–434

Ives C, Sathuvalli VR, Colburn BC, Mehlenbacher SA (2014) Mapping the incompatibility and style color loci in two hazelnut progenies. Hortscience 49:250–253

Kim KW, Chung HK, Cho GT, Ma KH, Chandrabalan D, Gwag JG, Kim TS, Cho EG, Park YJ (2007) PowerCore: a program applying the advanced M strategy with a heuristic search for establishing allele mining sets. Bioinformatics 23:2155–2162

Koksal, AI (2002) Turkish hazelnut cultivars. Hazelnut Promotion Group, Ankara ISBN, 975–92886

Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129

Martins S, Simões F, Mendonça D, Matos J, Silva AP, Carnide V (2013) Chloroplast SSR genetic diversity indicates a refuge for Corylus avellana in northern Portugal. Genet Resour Crop Evol 60:1289–1295

Martins S, Simões F, Matos J, Silva AP, Carnide V (2014) Genetic relationship among wild, landraces and cultivars of hazelnut (Corylus avellana L.) from Portugal revealed through ISSR and AFLP markers. Plant Syst Evol 300:1035–1046

Martins S, Simões F, Mendonça D, Matos J, Silva AP, Carnide V (2015) Western European wild and landraces hazelnuts evaluated by SSR markers. Plant Mol Biol Rep: 1–9

Mehlenbacher SA, Brown RN, Nouhra ER, Gokirmak T, Bassil NV, Kubisiak TL (2006) A genetic linkage map for hazelnut (Corylus avellana L.) based on RAPD and SSR markers. Genome 49:122–133

Ozturk SC, Ozturk SE, Celik I, Stampar F, Veberic R, Doganlar S, Solar A, Frary A (2017) Molecular genetic diversity and association mapping of nut and kernel traits in Slovenian hazelnut (Corylus avellana) germplasm. Tree Genet Genomes 13:16

Perrier X, Jacquemoud-Collet JP (2006) DARwin software. Available: http://darwin.cirad.fr/darwin

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000) Association mapping in structured populations. Am J Hum Genet 67:170–181

Sathuvalli VR, Chen H, Mehlenbacher SA, Smith DC (2011) DNA markers linked to eastern filbert blight resistance in "Ratoli" hazelnut (Corylus avellana L.). Tree Genet. & Genomes 7:337-345

Valentini N, Calizzano F, Boccacci P, Botta R (2014) Investigation on clonal variants within the hazelnut (Corylus avellana L.) cultivar 'Tonda Gentile delle Langhe'. Sci Hortic 165:303–310