Plant Molecular Biology Reporter (2018) 36:800–811
https://doi.org/10.1007/s11105-018-1120-0

ORIGINAL PAPER

Development of Simple Sequence Repeat Markers in Hazelnut (*Corylus avellana* L.) by Next-Generation Sequencing and Discrimination of Turkish Hazelnut Cultivars

Süleyman Can Öztürk[1] · Mehmet Göktay[1] · Jens Allmer[1,2] · Sami Doğanlar[1] · Anne Frary[1]

Published online: 13 November 2018

Abstract

European hazelnut (*Corylus avellana*) is a diploid tree species and is widely used in confections. Hazelnuts are, to a large part, produced in Turkey with the cultivar "Tombul" widely grown in the Black Sea region. In this work, the "Tombul" genome was partially sequenced by next-generation sequencing technology yielding 29.2% (111.85 Mb) of the ~ 385 Mb (1C). This sequence information was used to develop genetic markers in order to enable differentiation of material before the long maturation process and to facilitate future breeding strategies. A total of 90,142 simple sequence repeats (SSRs) were identified in the contigs giving a frequency of 1 SSR per 1240 nt in the assembly. Mononucleotides were the most abundant SSR marker type (60.9%) followed by di- and trinucleotides. Primer pairs were designed for 75,139 (83.3%) of the SSRs. Fifty SSR primers were applied to 47 hazelnut accessions from nine countries to test their effectiveness and polymorphism. The markers amplified an average of 3.2 fragments. The highest polymorphism information content value was for cavSSR11062 (0.97) and the lowest (0.04) was for cavSSR13386. Two markers were monomorphic: cavSSR12855 and cavSSR13267. Single-copy SSR primers were also assessed for their ability to discriminate 19 Turkish cultivars, and it was found that seven primer pairs (Cav4217, Cav14875, Cav14418, Cav2704, Cav12862, Cav3909, Cav1361) were sufficient for this task. Thus, this study developed new SSR markers for use in hazelnut breeding and genetic studies and also provide a method to distinguish and identify true-type Turkish cultivars.

Keywords Illumina sequencing · Filbert genomic SSRs · Fingerprinting · Genetic diversity · Microsatellites · Population structure · Barcode

Introduction

Hazelnut is one of the most important edible nut species in the world. European hazelnut (*Corylus avellana* L.) belongs to the Betulaceae family and is grown all over Europe, Anatolia, and the USA (Boccacci et al. 2006). Turkey is the world's main hazelnut producer and accounts for 63% of world production (FAO 2017a, b). Turkish hazelnut cultivars are classified depending on their nut shape and kernel quality, and cultivar names refer to a group of tree which have the same agro-morphological traits (Kafkas et al. 2009). Because hazelnut does not reach maturity for 5 to 10 years, nut and kernel traits

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11105-018-1120-0) contains supplementary material, which is available to authorized users.

✉ Anne Frary
annefrary@iyte.edu.tr

Süleyman Can Öztürk
scanozturk48@gmail.com

Mehmet Göktay
mehmetgoktay1989@gmail.com

Jens Allmer
jens@allmer.de

Sami Doğanlar
samidoganlar@iyte.edu.tr

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

[2] Present address: Applied Bioinformatics, Wageningen University and Research, Wageningen, The Netherlands

cannot be used to distinguish and verify cultivars when an orchard is established. In addition, as with other cultivars such as "Tonda Gentile delle Langhe" (Valentini et al. 2014), "Longue d'Espagne," "Daviana," and "Merveille de Bollwiller" (Ghanbari et al. 2005), Turkish hazelnut cultivars can have many variants at the molecular level which results in problems with certification (Islam and Ozgüven 2000; Islam 2003; Balik and Beyhan 2014). This problem can be solved using genetic discrimination analysis.

Starting in the 1990s, molecular analyses were done to discriminate cultivars and find true-type (clonal) accessions. For example, Solar et al. (1996) showed isozyme polymorphism for three enzyme systems in 15 hazelnut cultivars. Later research used DNA-level polymorphism. In early work, five randomly amplified polymorphic DNA (RAPD) markers were used to discriminate six cultivars and their variants from the Campania region of Italy (Galderisi et al. 1999). In another study, 10 of 18 Turkish cultivars were distinguished using five RAPD, four inter-simple sequence repeat (ISSR), and eight amplified fragment length polymorphism (AFLP) primers which yielded 34 cultivar-specific markers (Kafkas et al. 2009). Chloroplast DNA was also used to find the origins of 75 cultivars from Spain, Italy, Turkey (10 cultivars), and Iran using four polymorphic simple sequence repeat (SSR) loci (Boccacci and Botta 2009). In another study, 78 hazelnuts were analyzed with 16 SSR markers for fingerprinting hazelnuts from germplasm repositories (Boccacci et al. 2006). This was the first time that SSR markers were used for discrimination in hazelnut despite the fact that they have been previously shown to be convenient for fingerprinting in many other tree species such as apple, apricot, peach, pear, and olive (Sosinski et al. 2000; Hokanson et al. 2001; Yamamoto et al. 2001; Hormaza 2002; Rallo et al. 2000). In a more recent study, 14 SSRs were developed for fingerprinting 102 worldwide cultivars (Akin et al. 2016). Most recently, Lucchetti et al. (2018) reported that they could distinguish five cultivars with two SSRs using DNA samples which were isolated from hazelnut oil.

SSRs are short nucleotide repeats (1-6) that occur throughout the coding and non-coding regions of the genome (Gupta et al. 1996; Haq et al. 2014; Squirrell et al. 2003). SSR markers are effective because they are multiallelic, easy to score, and reproducible. As a result, they are commonly used in plant genetic diversity and breeding analysis. Genomic SSRs and genic SSRs are derived from DNA (genomic libraries) and RNA (expressed sequence tags, transcriptomic libraries) sequences, respectively. Length polymorphism in these coding and non-coding sequences can be easily detected by polymerase chain reaction. To date, approximately 650 genomic SSRs (Boccacci et al. 2005; Bassil et al. 2005, 2013; Mehlenbacher et al. 2006; Gurcan and Mehlenbacher 2010; Gurcan et al. 2010a; Bhattarai 2015; Bhattarai and Mehlenbacher 2017), 20 polymorphic EST-SSRs from the Betulaceae family (Boccacci et al. 2015), and 230 polymorphic SSR loci from transcriptome analysis were developed and used in hazelnut genome analyses (Peterschmidt 2013; Colburn et al. 2017). These analyses included determination of genetic diversity (Bacchetta et al. 2014; Boccacci et al. 2005; 2006, 2008, 2013; Ozturk et al. 2017b), geographic origin (Boccacci et al. 2006; Gökirmak et al. 2009; Gurcan et al. 2010a), identification of synonymous trees (Gökirmak et al. 2009; Gurcan et al. 2010a; Valentini et al. 2014), and construction of linkage maps (Gurcan and Mehlenbacher 2010; Gurcan et al. 2010b; Colburn et al. 2017; Bhattarai et al. 2017; Sathuvalli et al. 2017). All of these studies show that SSR markers are effective for hazelnut genomic research and suggest that the development of even more SSR markers will be useful for more comprehensive analyses. Moreover, additional single-copy SSR markers are needed as most of the previously identified markers were multicopy in Turkish germplasm (Ozturk et al. 2017b), thus complicating subsequent mapping. The present study aimed to identify hazelnut specific, single/low-copy SSR markers using next-generation sequencing technology. To achieve this goal, genomic DNA of a popular Turkish hazelnut cultivar (*C. avellana* cv. "Tombul") was sequenced by Illumina Next Generation Sequencing (NGS) technology for identification of SSRs. Finally, a set of 50 SSR markers was validated in 47 hazelnut accessions to demonstrate their usefulness for examination of genetic diversity and population structure. Seven of the 50 SSR markers were sufficient to allow discrimination among 19 Turkish cultivars.

Materials and Methods

Plant Material, DNA Isolation, and Sequencing

The Turkish cultivar "Tombul" obtained from the Hazelnut Research Institute (Giresun, Turkey) was used for sequencing. Total genomic DNA of "Tombul" was extracted using the Wizard Magnetic 96 Plant System (Promega Crop., Madison, WI, USA) and the Beckman Coulter Biomek NX Workstation. IIlumina Mi-Seq sequencing was performed at the Biotechnology Center at the University of Wisconsin-Madison, USA (https://www.biotech.wisc.edu/).

Data Pre-Processing

IIlumina sequencing technology depends on adapters (synthetic short DNA sequences) to sequence DNA fragments. These adapter sequences may decrease assembly quality and must be removed. Thus, adapter sequences were removed from reads using the Cutadapt version

1.8.3 software using default settings (Martin 2011). At the end of this step, any reads smaller than 20 nucleotides were discarded. To detect human contaminants in the dataset, cleaned reads were mapped against the human genome using Bowtie version 2.1.0 (Langmead and Salzberg 2012) and possible contaminants were removed.

Sequence Assembly

ABySS version 1.3.6 (Simpson et al. 2009), a de novo, parallel, paired-end sequence assembler, was used to perform genomic DNA sequence assembly. To produce the best possible assembly, more than 100 runs were performed using a parameter sweep for example changing the kmer parameter (all possible substrings of length k contained in reads) and the required number of reads to make a contig. In de novo genome assembly, there is not just one measurement or parameter to determine the best assembly; instead, a combination of different measurements or parameters influences the quality of the final assembly. For this purpose, N50 value (weighted median of contig length), assembly nucleotide length (closeness to estimated size of the *C. avellana* genome), and length of the largest contig were used to identify the best assembly. The settings that were finally chosen to create contigs were: (kmer = 45) with otherwise default settings.

SSR Detection, Annotation, and Primer Design

Contigs shorter than 1000 nucleotides were removed from the assembly. Thus, we only analyzed contigs larger than 1000 nucleotides for SSR detection using our in-house tool SiSeeR (http://bioinformatics.iyte.edu.tr/index.php?n=Softwares.SiSeeR). The minimum number of repeats required to identify perfect SSRs was ten for mononucleotides, four for dinucleotides, and three for motifs comprised of three or more nucleotides. To annotate these identified SSRs, SSR sequences were extracted with their genomic context (padded with 100 nucleotides) and were converted to FASTA formatted sequences. These sequences were treated as query sequences and searched against the Uniprot non-redundant plant protein database (taxonomy = Viridiplantae) with BLASTX version 2.2.30 (Altschul et al. 1990). The Primer 3 program (primer_core) version 2.3.6 (Koressaar and Remm 2007) was used to design primer pairs for the SSRs with the default settings and primer task = generic, primer optimum size = 20, primer maximum size = 24, primer minimum size = 18, primer product size = 100–300, primer minimum Tm = 50, primer maximum Tm = 60, and primer optimum Tm = 55.

Sequencing of SSR Loci

To ensure that the expected SSRs were amplified by the primers, "Tombul" DNA was used as a template and the dye-terminator sequencing method was performed to validate SSR motifs. Eight primer pairs were randomly selected and PCR fragments were purified with the DNA Clean & Concentrator–5 Kit (Zymo Research) and used as templates for sequencing using GenomeLab DTCS Quick Start Kit (Beckman Coulter). Thermal cycling conditions of the sequencing reactions were 30 cycles of 96 °C for 20 s, 50 °C for 20 s, and 60 °C for 4 min. The reaction mixture for each SSR amplicon was then purified using ZR DNA Sequencing Clean-up Kit (Zymo Research) and DNA was resuspended in 30 μL of sample loading solution (Beckman Coulter) and run on a Beckman CEQ8800 capillary electrophoresis device using the LFR-c method (injection voltage 2.0 kV for 10–15 s, separation temperature 60 °C, separation voltage 7.4 kV, separation time 45 min).

Validation of Genomic SSR Markers in Hazelnut Accessions

SSR marker validation was done using 19 Turkish cultivars; 27 cultivars from nine other countries: Italy, the USA, France, UK, Croatia, Germany, Romania, Spain, and Hungary (samples provided by Dr. Anita Solar, University of Ljubjana, Slovenia); and one wild genotype from the Hazelnut Research Institute (Table 1). Total genomic DNA was isolated from leaves sampled from individual trees following the protocol of Fulton et al. (1995). Amplification of the hazelnut DNA with genomic SSR primers 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. Thermal cycling conditions consisted of one cycle of initial denaturation for 10 min at 94 °C, followed by 30 cycles of 94 °C for 30 s, 55 °C for 30 s, 72 °C for 30 s, with a final extension step of 10 min at 72 °C. 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 and were scored binomially (presence 1, absence 0).

Data Analysis

The PowerMarker software (Liu and Muse 2005) was used to calculate polymorphism information content (PIC) and observed heterozygosity (Ho) values. Polymorphic loci were used to analyze molecular genetic diversity and determine population structure. A dissimilarity matrix was created using the Dice coefficient (Dice 1945) and then used to construct a dendrogram with the unweighted neighbor-joining algorithm using the DARwin 5 software (Perrier and Jacquemoud-Collet 2006). The program

was also used for principal coordinate analysis (PCoA). Population structure was determined using the Structure 2.3.4 computer program (Pritchard et al. 2000). Ad hoc statistics were used to find the best reflected subpopulation number for the hazelnut genotypes (Evanno et al. 2005). For this analysis, the data were evaluated for 2 to 10 subpopulations ($K = 2$ to 10) with a burn-in time of 50,000 cycles. Each model was tested 10 times with 300,000 iterations per K. The Structure Harvester program (Earl and von Holdt 2012) was used to calculate the probability change of each group (ΔK). The model with the highest ΔK value was chosen as the best number of subpopulations. Accessions were clustered using a threshold of inferred ancestry ≥ 0.70, and if the greatest probability of belonging to a subpopulation was lower than 0.70, the accession was assigned to the admixed group. To confirm the results of the Structure software and to test $K = 1$, the InStruct software (Gao et al. 2007) was used. Since the outcome was the same, these results are not shown in the paper.

Discrimination Analysis for Turkish Cultivars

The binomial data set was analyzed to discriminate Turkish cultivars from each other with a minimum number of SSR markers. The SSRs which gave two alleles after PCR amplification were chosen to ensure that the SSR was single copy in the hazelnut genome and to simplify scoring. Combinations of SSRs were tested until all standard Turkish cultivars which are grown at the Hazelnut Research Institute orchard were discriminated from each other by the minimum set of SSRs.

Results

Sequence Assembly, Simple Sequence Repeat Identification, and Primer Design

Sequencing of the hazelnut cultivar "Tombul" produced 15,319,058 sequence reads comprising more than 4595 Mb. Removal of adapter sequences and quality clipping from the raw reads resulted in 4535 Mb sequence with an average size of 296.1 nucleotides (nt). This sequence was used for contig assembly and only contigs larger than 1000 nucleotides were further analyzed for SSR identification. As a result, 56,665 contigs were assembled which encompassed 111.85 Mb, representing 29.2% of the ~385 Mb (1C) hazelnut genome (Table S1).

Overall, 90,142 SSRs were identified in the contigs with 1 SSR every 1240 nt in the assembly. SSR length ranged from 6 to 49 nt with an average of 15.4 nt. Among all identified SSRs, the most abundant type was mononucleotide repeats (60.9%). Dinucleotides and trinucleotides were the second and third most common type representing 26.5 and 4.4% of the SSRs, respectively (Fig. 1). The most common motifs were A/T repeats (99.2%) for mononucleotides and AT repeats (25.4%) for dinucleotides (Table 2). Among trinucleotides, the most frequent repeats were ATT/AAT repeats which accounted for 32.2% of trinucleotides (Table 3). A total of 75,139 primer pairs were successfully designed for the 90,142 identified SSRs.

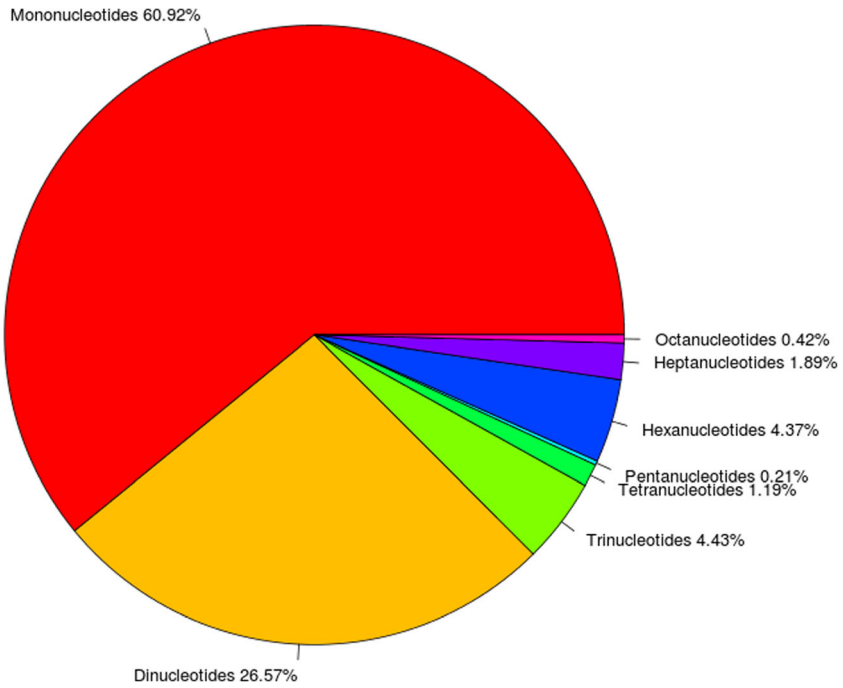

Fig. 1 Simple sequence repeat types in *C. avellana*

SSR Validation

To confirm that the designed primers amplified the expected SSRs, PCR products from eight primer pairs amplified on "Tombul" DNA were sequenced with the dye-terminator method (data not shown). All eight sequences contained the expected SSR motifs, proving that the primers amplified regions containing SSRs.

Fifty of the newly developed SSR markers were validated by amplification using 47 accessions representing the hazelnuts of ten countries (Table 1). In all, 45 of the primers (90%) produced polymorphic bands and generated 163 loci, 104 of which were polymorphic (64%; Table 3). Average allele number for each SSR marker was 3.2. The PIC value was highest for cavSSR11062 (0.97). The lowest value was 0 for two monomorphic markers: cavSSR12855 and cavSSR13267.

The SSR data were used to construct a dendrogram using the Dice coefficient and unweighted neighbor-joining algorithm. The Mantel test showed a high correlation between the distance matrix and dendrogram ($r = 0.95$). The average diversity of accessions was 0.17 with the highest value (0.30) between "Tombul" and "FAI604" together with "Allahverdi" and "Fosa" and the lowest value (0.05) between Corabel (N-473) and Ferwiller. The hazelnut accessions grouped into three clusters (A, B, and C) in the dendrogram (Fig. 2). Cluster A contained 14 accessions in two subclusters (A1 and A2). Cluster A1 contained 12 accessions and A2 contained two accessions. Genetic diversity in cluster A ranged from 0.09 to 0.29 with an average diversity value of 0.17 (data not shown). Most of the Turkish hazelnut cultivars were found in cluster A. Two non-Turkish accessions were found among the 12 Turkish accessions in cluster A including "UNITO119" and "Negret." Cluster B contained 21 accessions in three subclusters (B1, B2, and B3) with genetic diversity ranging from 0.09 to 0.29 with an average of 0.18. Six Turkish accessions were found in cluster B1 with "Brixnut." Cluster B2 contained seven European hazelnut cultivars from Croatia ("Istrska dolgoplodna leska"), France ("Bearn"), Germany ("Gunslebert," "Landsberg"), Hungary ("Romoi"), Romania ("Arutela"), and the UK ("Badnuss"). On the other hand, six European cultivars from France ("Corabel," "Ferwiller"), Italy ("Tonda di Giffoni," "UNITO101") and the Turkish cultivars "Giresun Melezi" and "Okay28" ("Kargalak" × "Tombul" hybrids) were found in cluster B3 with the US cultivar "Lansing." Cluster C had 12 accessions and genetic diversity ranged from 0.06 to 0.18 with an average diversity value of 0.11. Hazelnut cultivars from Croatia ("Istrska okrogloplodna leska"), France ("Feriale"), Italy ("Riccia di Talanico," "104F," "Daria"), Spain ("Pauetet"), the UK ("Cosford," "Mogul"), and the USA ("Lewis," "Ennis," "Willamette") were found in cluster C.

Table 1 Hazelnut accessions used in this study

Name	Origin	Cultivar/ wild	Inferred ancestry	Dendogram cluster
UNITO101	Italy	Cultivar	4	B3
UNITO119	Italy	Cultivar	3	A2
Aci	Turkey	Cultivar	1	A1
Allahverdi	Turkey	Cultivar	2	B1
Arutela	Romania	Cultivar	3	B2
Badnuss	UK	Cultivar	3	B2
Bearn	France	Cultivar	3	B2
Brixnut	USA	Cultivar	Admixed	B1
Cavcava	Turkey	Cultivar	2	B1
Corabel	France	Cultivar	Admixed	B3
Cosford	UK	Cultivar	Admixed	C
Cakıldak	Turkey	Cultivar	1	A1
Daria	Italy	Cultivar	3	C
Ennis	USA	Cultivar	Admixed	C
104F	Italy	Cultivar	3	C
FAI604	Turkey	Wild	2	B1
Feriale	France	Cultivar	3	C
Ferwiller	France	Cultivar	Admixed	B3
Fosa	Turkey	Cultivar	1	A1
Giresun Melezi	Turkey	Cultivar	4	B3
Gunslebert	Germany	Cultivar	3	B2
Istrska dolgoplodna leska	Croatia	Cultivar	3	B2
Istrska okrogloplodna leska	Croatia	Cultivar	3	C
Incekara	Turkey	Cultivar	1	A1
Kalinkara	Turkey	Cultivar	1	A1
Kan	Turkey	Cultivar	2	B1
Kara	Turkey	Cultivar	1	A1
Kargalak	Turkey	Cultivar	1	A1
Kuş	Turkey	Cultivar	1	A1
Landsberg	Germany	Cultivar	3	B2
Lansing	USA	Cultivar	4	B3
Lewis	USA	Cultivar	3	C
Mogul	UK	Cultivar	3	C
Negret	Spain	Cultivar	4	A2
Okay28	Turkey	Cultivar	4	B3
Palaz	Turkey	Cultivar	1	A1
Pauetet	Spain	Cultivar	3	C
Riccia di Talanico	Italy	Cultivar	Admixed	C
Romoi	Hungary	Cultivar	3	B2
Sivri	Turkey	Cultivar	Admixed	A1
Tombul	Turkey	Cultivar	1	A1
Tonda di Giffoni	Italy	Cultivar	4	B3
Uzun Musa	Turkey	Cultivar	1	A1
Valcea	Romania	Cultivar	3	C
Willamette	USA	Cultivar	3	C
Yassi Badem	Turkey	Cultivar	2	B1
Yuvarlak Badem	Turkey	Cultivar	Admixed	B1

Cluster assignments of the 47 accessions are according to population structure (inferred ancestry column) and genetic diversity (dendrogram cluster column) analyses

Population structure was also determined using the SSR data. The model with four subpopulations ($K = 4$) was determined as the best model for population structure (Suppl. Fig. 1, Suppl. Fig. 2). Subpopulations 1 and 2 included 11 and five accessions from Turkey, respectively. A total of 17 accessions were found in subpopulation 3 and six accessions were found in subpopulation 4. The remaining eight

Table 2 Most abundant simple sequence repeat (SSR) motifs in the *C. avellana* genomic sequence

SSR motif[a]	Number of SSRs	Motif frequency (%)
A/T	54,489	99.2
GA/TC	4452	18.5
AT	6092	25.4
AG/CT	4603	19.2
TA	5836	24.4
ATA/TAT	434	11.0
ATT/AAT	1287	32.2
TTA/TAA	734	19.0
ATTT/AAAT	284	26.5
AGAAA/TTTCT	21	11.0
AAAAAT/ATTTTT	440	11.1
AAAAATA/TATTTTT	192	11.2
TTTTTAT/ATAAAAA	189	11.1

[a] Motif frequencies are relative to SSR type. Only motif with a frequency ≥ 10% are listed

accessions were admixed: "Brixnut," "Corabel," "Cosford," "Ennis," "Ferwiller," "Riccia di Talanico," "Sivri," and "Yuvarlak Badem" (Table 1). Turkish accessions were distributed throughout all subpopulations (1, 2, 4, and admixed) except subpopulation 3. When the population structure results were compared with the dendrogram analysis, subcluster A1 corresponded to subpopulation 1 with the addition of one admixed accession ("Sivri"). Subcluster B1 corresponded to subpopulation 2 with two admixed accessions ("Yuvarlak Badem," "Brixnut"). In addition, subcluster B2 and cluster C corresponded to subpopulation 3 with three admixed accessions from cluster C ("Cosford," "Ennis," "Riccia di Talanico"). Subcluster B3 corresponded to subpopulation 4 with two admixed accessions ("Ferwiller," "Corabel"). Principal coordinate analysis (PCoA) of the SSR dataset did not show a clear separation between subpopulations (Fig. 3). In the PCoA plot, the dendrogram subclusters that were predominantly Turkish material (A1 and B1) were clearly separated from the remaining accessions.

Seven (Cav4217 Cav14875, Cav14418, Cav2704, Cav12862, Cav3909, Cav1361) of the 50 SSR markers were chosen as the minimum set of primers needed to discriminate 19 Turkish hazelnut accessions from each other (Fig. 4). Combinations of two, three, four, and five SSR markers were sufficient to discriminate the Turkish accessions. Cav4217 and Cav14875 were able to separate "Palaz" and "Cakıldak" cultivars from each other in two-step PCR. "Kan" and "Giresun Melezi," "Uzun Musa," and "Kargalak" could be discriminated from each other with all primers. Cav4217, Cav14875, and Cav14418 assays were common for all cultivars, and determination of the heterozygosity and homozygosity of these markers was enough to discriminate several of the cultivars. However, different combinations of additional markers (Cav2704, Cav12862, Cav3909, and Cav1361) were needed for complete discrimination of the 19 Turkish cultivars.

Discussion

Validation of Genomic SSR Markers

Sequencing of "Tombul" yielded 56,665 contigs which were assembled into 111.85 Mb, representing 29.2% of the hazelnut

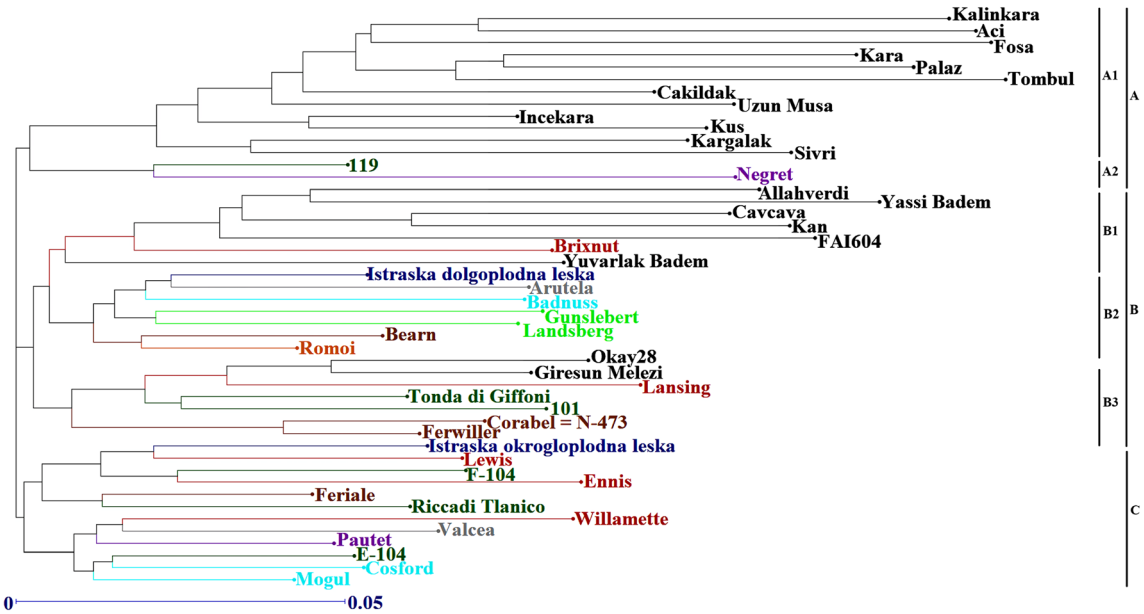

Fig. 2 Unweighted neighbor-joining dendrogram of the 47 hazelnut accessions based on SSR data. Accessions are color coded by origin: Croatia: dark blue, France: brown, Germany: light green, Hungary: orange, Italy: dark green, Romania: gray, Spain: purple, Turkey: black, UK: light blue, USA: red

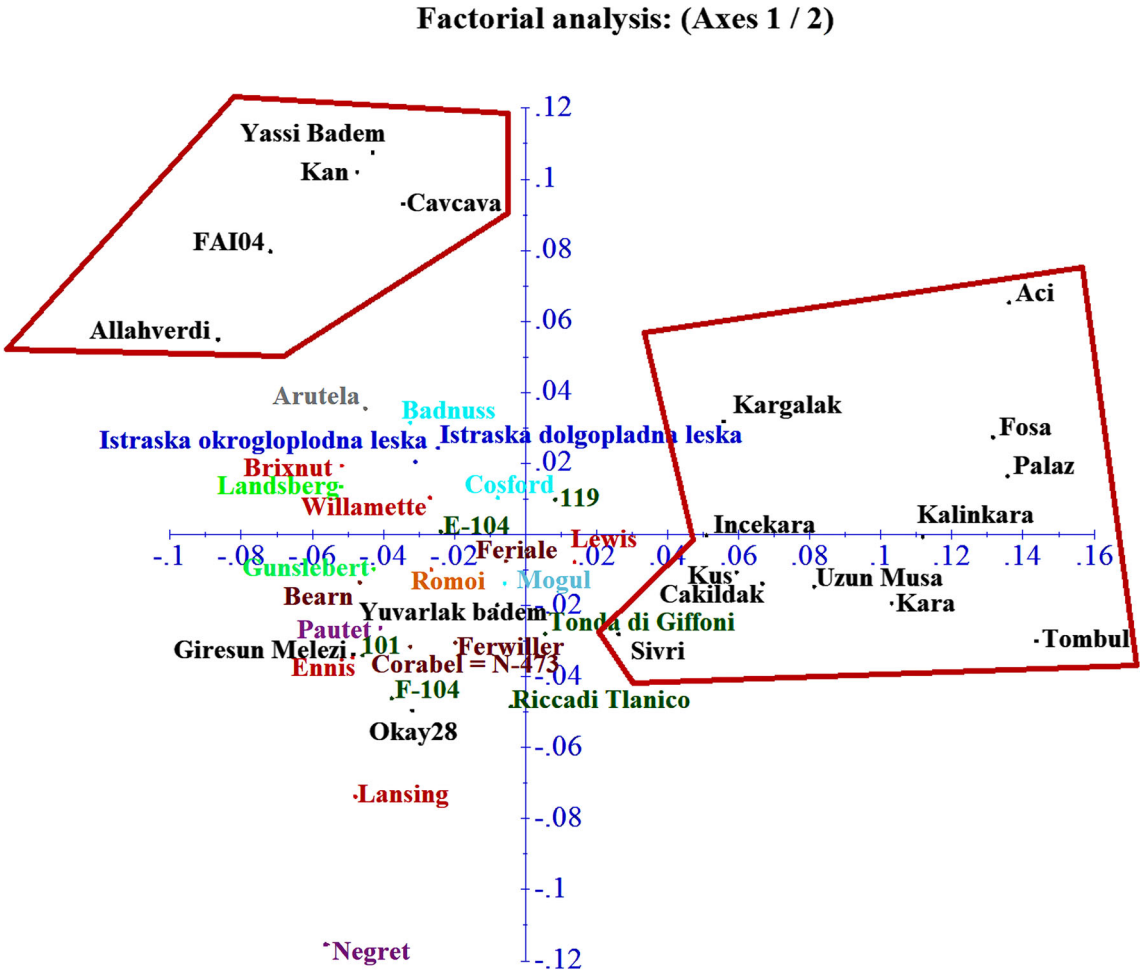

Fig. 3 Principal coordinate analysis of hazelnut accessions according to the first two Eigen vectors which explained 17.2 and 8.7% of the variance, respectively. Most of the Turkish accessions in Cluster A1 and B1 are grouped in red areas on PCoA plot

genome. In other work, "Jefferson," a cultivar resistant to eastern filbert blight, was sequenced and assembled to cover 345 Mb, representing 91% of the genome with 40× coverage (Bhattarai and Mehlenbacher 2017). In the same study, "Tombul" was sequenced with Illumina at a low coverage (10×) sufficient to finding the eastern filbert blight resistance gene using "Jefferson" as a reference. Thus to date, our study has produced the most enlightening genome information for "Tombul."

SSR Markers Developed by NGS

SSRs are iterations of one to six nucleotide motifs and are found in all prokaryotic and eukaryotic genomes (Zane et al. 2002). SSR markers are very important for plant scientists because they can detect multiple alleles per locus, are highly polymorphic, and can be found throughout the plant genome (Zalapa et al. 2012). Thus, the development of SSR markers, especially for economically important crops, is essential for more efficient plant genome analysis. Traditional SSR development techniques such as library enrichment and Sanger sequencing techniques are low-throughput, labor intensive, expensive, and yield a small number of SSRs (Zalapa et al. 2012). Unlike traditional methods, next-generation sequencing technology is high-throughput, fast, cost-effective, and produces millions of sequence reads (Zalapa et al. 2012). Therefore, traditional technologies have been replaced with next-generation technology in the area of SSR development.

A total of 90,142 non-redundant SSR markers were identified in 29.2% of the *C. avellana* genome. SSR density in these contigs was one SSR in every 1.2 kb (on average). Sequencing of the "Jefferson" cultivar revealed an average SSR density of one SSR in every 1.9 kb of the contigs (Bhattarai and Mehlenbacher 2017), agreeing with our results. The most abundant SSR marker type was mononucleotides which accounted for 60.9% of the identified SSRs, followed by dinucleotides with 26.5% and trinucleotides with 4.4%. Sequencing of "Jefferson" also indicated that mononucleotides were most common (69.3%). The results also agree with Cardle et al. (2000) who also showed that the most common

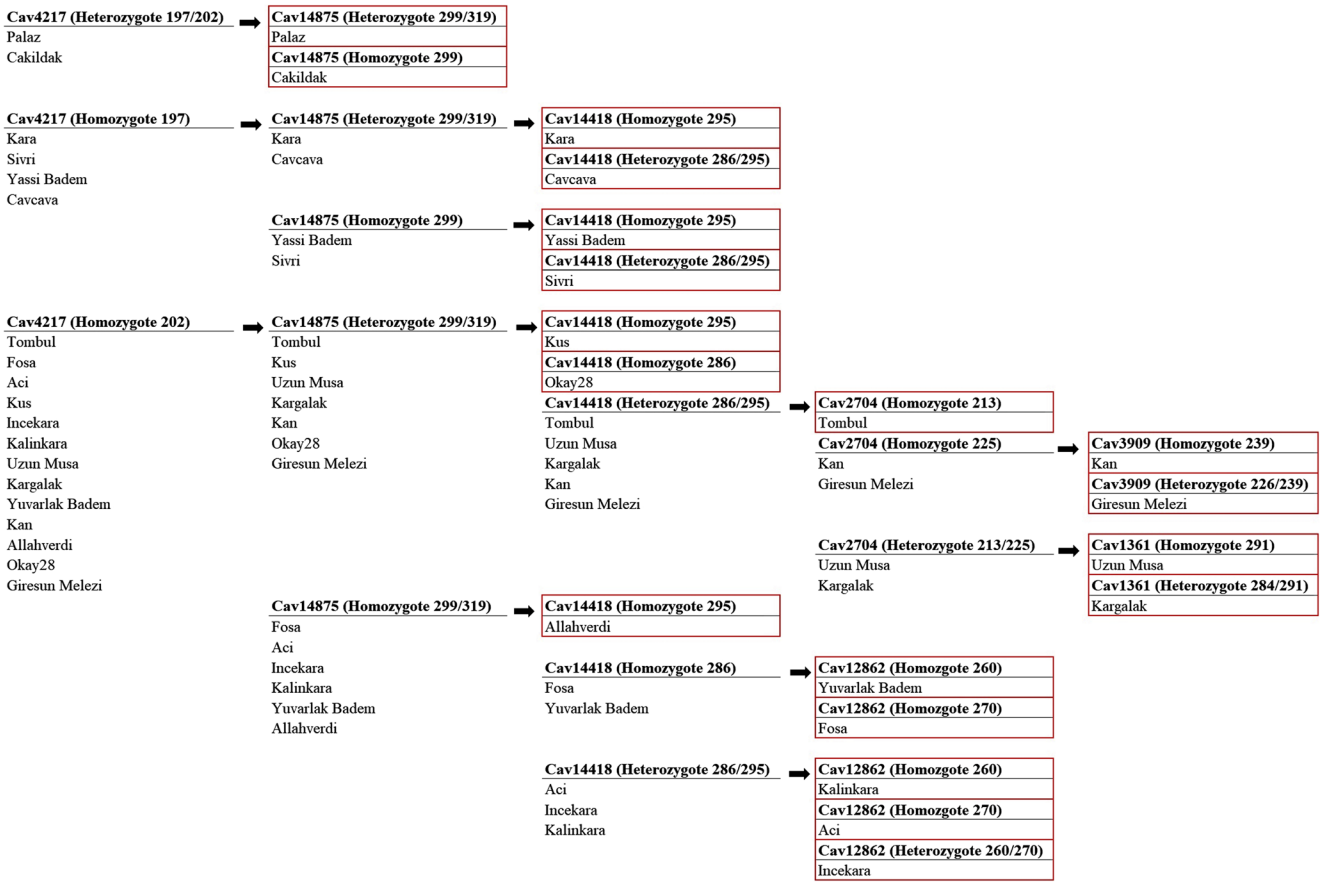

Fig. 4 Identification key displaying the discrimination of Turkish hazelnut cultivars according to seven SSR markers. Red boxes indicate endpoints where each cultivar is discriminated from the others

SSR type in many plants was mononucleotides followed by dinucleotides and trinucleotides. The most abundant mononucleotides in "Tombul" were A/T repeats (99.2%). "Jefferson" also had a majority of A/T mononucleotides compared to G/C ones (Bhattarai and Mehlenbacher 2017). These results agree with other studies which observed that the most common SSR repeat types in plants are A/T repeats (Celik et al. 2014; Gol et al. 2017; Abuzayed et al. 2017). According to the same studies, AT/TA repeats are the second most abundant SSR type in many plants. We also observed the same pattern in our study, 25.4% AT and 24.4% TA SSRs among dinucleotide repeats. The most common tri- and tetranucleotide motifs vary based on the plant species. In this study, the most frequent SSR type was ATT/AAT (32.2%) in trinucleotides and ATTT/AAAT (26.5%) in tetranucleotides. AT-rich trinucleotides were also the most common (43.1%) in "Jefferson" (Bhattarai and Mehlenbacher 2017).

Application of Genomic SSR Markers to Genetic Diversity and Population Structure Analyses

Fifty SSR markers were randomly chosen and tested in 47 hazelnut accessions from ten countries. In this study, the SSR markers that were used in diversity analyses were single or low copy except for markers cavSSR11062 (12 fragments), cavSSR1632 (8 fragments), and cavSSR8737 (7 fragments). The number of alleles ranged from one to 12 with an average of 3.2 alleles. This average allele number was in the range of the previous studies (from 3 to 13.6) which used SSR markers in hazelnut (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. 2017a, b).

A dendrogram was constructed using SSR data and the accessions fell into three subclusters. All of the Turkish accessions were in clusters A1 and B1. The genetic distinctness of the Turkish material suggests that it can be used as a source of diversity for US and European breeding programs. Accessions from Italy, France, and the UK were found in both cluster C with the US cultivars and in cluster B with the other European hazelnut accessions. Trees from Germany clustered together in the dendrogram and the US cultivars were clustered with small groups of accessions from France, Italy, Turkey, and the UK in (sub)clusters A1, B3, and C. This was not surprising because the US cultivars' parents are from these countries (Mehlenbacher 2014). The 47 hazelnut accessions fell into four subpopulations.

Table 3 Simple sequence repeat (SSR) markers used for the molecular genetic analysis of hazelnuts

Primer name	Forward sequence	Reverse sequence	SSR motifs	PIC	Ho	# of polymorphic fragments/total # of fragments (%)
cavSSR202	CTCAGACACGCTCT CATTTA	AGTAGTAGTGCTCC ACGAAT	$(CT/GA)_{12}$	0.60	0.64	3/3 (100)
cavSSR325	GAGAGAGCTCACAG ACAATT	TTCTTCTCTGGAGG GGATAG	$(AG/TC)_{16}$	0.54	0.61	2/3 (66.7)
cavSSR1361	GATATCACTCACGT CTACCG	GGTCTCTTGGTCTT GATGTT	$(AG/TC)_{14}$	0.58	0.63	1/2 (50)
cavSSR1601	TCTGGAGTTAGCTA CTGTCA	ACTAGTACCTTGGA GTACCC	$(AATTT/TTAAA)_5$	0.45	0.48	1/3 (33.3)
cavSSR1632	GCCTATGTCCCTCT ACAAAG	AGGAAAGTGAAGAT GGTTCC	$(AG/TC)_{12}$	0.93	0.94	8/8 (100)
cavSSR1828	CGGAGTGTTTTAAT GGCATC	TGGTTGGAGAACTG TACATG	$(GA/CT)_{12}$	0.08	0.08	0/5 (0)
cavSSR2135	ATGTAGCGAGCCTT GATAAG	GTTGTCAGGTAGCT TGAAGA	$(TTAA/AATT)_6$	0.41	0.46	2/2 (100)
cavSSR2527	ACCTAGTAGCTGCA TTTAGC	CTACCTCCAGGAGT CAACTA	$(AAT/TTA)_8$	0.46	0.50	2/2 (100)
cavSSR2590	GGTAGGCTGTGTTT TCTGTA	CAGATAGAACGGAC TGGATG	$(TC/AG)_{12}$	0.21	0.23	2/4 (50)
cavSSR2704	GCGGAGTTGGTAGT GATAAT	ATATAGGTATAAGG GGGCCC	$(TAG/ATC)_{10}$	0.72	0.75	2/2 (100)
cavSSR2975	CTGGGCATTTAGGT GTAGTT	GTAGAGAGTGGCCA AAACAT	$(CT/GA)_{12}$	0.44	0.47	4/5 (80)
cavSSR3126	CCGTGAGTTTGTAA GATTGC	AAACCTCTCACTAA GGAGGT	$(GA/CT)_{12}$	0.24	0.28	1/2 (50)
cavSSR3909	AGATGAAGCTGAAG AAAGGG	TATCGCCATCACAC CATTAG	$(GGA/CCT)_8$	0.27	0.29	1/2 (50)
cavSSR4217	GACAGTTGGCATGA AAGATG	GCACTCATCAGAGA GTCAAA	$(ATT/TAA)_{10}$	0.53	0.57	2/2 (100)
cavSSR4769	CCCATGTACGTATT CTCAGG	ATACTGAACCCTTC CGTGTA	$(GCA/CGT)_8$	0.46	0.50	2/3 (66.7)
cavSSR4874	GTCTTGAGAACCTA CACGTT	ACAACATCCGGATA GAAAGG	$(GA/CT)_{13}$	0.59	0.64	1/2 (50)
cavSSR4912	GTTTCCCTTTCCCT CATCAT	CAGTACTGAGGGTT GGATTG	$(GGA/CCT)_8$	0.45	0.48	2/2 (100)
cavSSR6172	TCTGCTTGGAGTGA GGTATA	TCCTTCTGAAGCTC AAGTTC	$(ATA/TAT)_8$	0.54	0.57	2/2 (100)
cavSSR6904	ATCTCCGAGAAAGT CAGAGA	AAGAGCTCTGAGGA TCTGAT	$(GA/CT)_{13}$	0.54	0.61	2/5 (40)
cavSSR7457	CTTGCTTTTAGGAC CTGAGT	CCTGCAATACTAGT GCTTCT	$(AG/TC)_{17}$	0.60	0.65	2/2 (100)
cavSSR7631	TTCCAGGAGCAAGA GATAGA	TTGTAGTTACAGGC AAGACC	$(CT/GA)_{14}$	0.46	0.56	1/3 (33.3)
cavSSR7755	TGAGTATTTGGACC TTGTGG	AAGGAGAAGCTTAC ACTGTG	$(CCA/GGT)_9$	0.12	0.12	0/2 (0)
cavSSR8129	GGTAATTGTTGGAG ACCCAT	CTCTCTCTCCATGT GTCTTG	$(TAT/ATA)_{11}$	0.25	0.27	2/3 (66.7)
cavSSR8344	AAGTTCACGAGTCT AATCCG	GTAGTCACTGCTAT GAGGTG	$(CT/GA)_{12}$	0.54	0.57	5/6 (83.3)
cavSSR8498	GCTAAATTCGCAGA GAGAGA	GCGCGCTTATATAA ATAGGC	$(GA/CT)_{13}$	0.57	0.63	2/3 (66.7)
cavSSR8737	AAAGACTCAAATCT GCTCCC	GAGGTATGCCAACT GAATGA	$(AG/TC)_{13}$	0.57	0.61	2/7 (28.6)
cavSSR9999	CACTCATGGAAGGA GAAACA	TAGCAGAGGAAACA GAACAC	$(TTTC/AAAG)_6$	0.47	0.55	1/2 (50)
cavSSR10247	GGCTCGCTGTAAAG ATGATA	TCCTACAAGCTGTC ATGAAC	$(TC/AG)_{17}$	0.57	0.62	2/2 (100)
cavSSR10870	GGTCAATTGCATAC AGTTGG	TAAAGGGTGAGGTG TAGGAA	$(GA/CT)_{16}$	0.60	0.67	2/3 (66.7)
cavSSR11062	CTCTCAGCAGGAAG AGAATC	CTGAGCTTCTTCTT AAGGCA	$(CTT/GAA)_8$	0.97	0.97	12/12 (100)
cavSSR11181			$(AG/TC)_{14}$	0.70	0.74	4/8 (50)

Table 3 (continued)

Primer name	Forward sequence	Reverse sequence	SSR motifs	PIC	Ho	# of polymorphic fragments/total # of fragments (%)
	TACTACTAAGACCCCACCTG	AGTACATGTGTCAACACTCC				
cavSSR11645	TTCTTTGGTGGATGTGAGAG	CTGAAAGAGAGCTTCCATGT	$(TC/AG)_{16}$	0.25	0.27	1/2 (50)
cavSSR12041	ATTCGGCTTGAATCTCTACC	CAATGGCTCTGGTATTCTGT	$(GAT/CTA)_8$	0.18	0.19	1/2 (50)
cavSSR12192	GGGATAACAGACCGAACTAC	GGGGGCAATTAGGTCTTTAA	$(TATG/ATAC)_7$	0.42	0.45	2/3 (66.7)
cavSSR12846	CGTCTATGGTCGTTCAATCT	GTCTCCTTTTTGTATGCACG	$(ATT/TAA)_8$	0.12	0.12	2/3 (66.7)
cavSSR12855	GGTAGTGATGATTGGGTTGT	AATAACCAGTTTCTCCGAGC	$(AG/TC)_{17}$	0.00	0.00	0/5 (0)
cavSSR12862	TAAAATGGGCCTACACTTCC	CCAGTACAGGAAGATACGAA	$(CT/GA)_{13}$	0.47	0.51	2/2 (100)
cavSSR13164	AGAAGAAAGCACTCCTCTTG	CTACCTGCTGTTCCTTTTCC	$(GA/CT)_{13}$	0.20	0.22	1/2 (50)
cavSSR13267	ATATATGCACTGTGGAGGTG	CCCTACTCACTCTATCACCA	$(AG/TC)_{15}$	0.00	0.00	0/3 (0)
cavSSR13350	TTATCCTCAATGCCTTGGAC	AACTTCTTCATCAAGACCCC	$(AG/TC)_{15}$	0.73	0.77	3/3 (100)
cavSSR13386	CCAACGAATCAAAAGACGAG	CCGCCTTCCATATAACTGAA	$(GA/CT)_{14}$	0.04	0.04	0/6 (0)
cavSSR13416	GGGCTTAGCATATGAAGTCA	AGGGTTGTACTACTAGGCAT	$(AG/TC)_{15}$	0.23	0.23	2/3 (66.7)
cavSSR13676	CATCGATGGAGAGGTTAAGG	CATACAAACCTATCCTGGGG	$(TTTC/AAAG)_6$	0.60	0.62	3/3 (100)
cavSSR13891	AAAGGTTGGGATGATGAGTC	ACTCTCCAATCGTATCCTCA	$(AAG/TTC)_{10}$	0.18	0.19	2/2 (100)
cavSSR14219	TATATGGACAGCTGACTCCA	GAGGGAGTTTGTCTGTCTTT	$(AAT/TTA)_9$	0.48	0.57	1/2 (50)
cavSSR14267	CCATCCAGGATCAAGTTGAT	TCAAAGCACCCATACTACAG	$(CATA/GTAT)_6$	0.21	0.23	2/2 (100)
cavSSR14418	GACTGCAAGAATGACAACAG	GTCCTCCTCCTTTTTCGTAG	$(TTGG/AACC)_6$	0.60	0.67	2/2 (100)
cavSSR14875	CACAAGATGATACCCATGCT	TATCAGCTCCTAAAACGACG	$(TACA/ATGT)_6$	0.53	0.61	1/2 (50)
cavSSR14904	GGGTTTTCGATCAGAACAAC	GTCTCGCTCTCTCTCTCTAT	$(AG/TC)_{12}$	0.25	0.26	2/2 (100)
cavSSR14937	TGAGCTCTCTGGTTTCTTTC	ACTGGATCTGCTTTTATGGG	$(CT/GA)_{12}$	0.47	0.54	2/2 (100)

Turkish cultivars were found in both subpopulation 1 (11 accessions), subpopulation 2 (5 accessions), and subpopulation 3 (2 accessions) as well as in the admixed group (2 accessions).

The genetic analyses based on SSR markers revealed some interesting findings. Two Turkish cultivars ("Giresun Melezi" and "Okay28") were developed from hybridization of "Kargalak" and "Tombul." Although "Giresun Melezi" and "Okay28" clustered together in the dendrogram, population structure, and PCoA plots, they did not group most closely with "Kargalak" and "Tombul." Thus, hybridization resulted in new allelic combinations as compared to the parental lines. Such novelty may be especially pronounced in heterozygous breeding material like hazelnut. Another interesting case was the Turkish cultivar "Yuvarlak Badem." This cultivar was found in the same subcluster (B1) as some of the Turkish cultivars but had admixed population structure and grouped with other countries' hazelnuts in the PCoA plot. In addition, "Yuvarlak Badem" has distinctly different nut traits compared to other Turkish cultivars. Its nuts are round and longer than the other material and are harvested earlier (Erdogan et al. 2010). Thus, this cultivar may have originated from Europe.

Hazelnut is a wind-pollinated species and has a self-incompatible mating system. As a result, high levels of genetic diversity are usually expected. However, in our study, the average diversity was low (0.17). In our previous studies, genetic diversity for Slovenian and Turkish cultivars was 0.50 (Ozturk et al. 2017a) and 0.47 (Ozturk et al. 2017b), respectively, and this was considered to be moderate genetic differentiation. However, these previous studies used SSRs that were selected for their high levels of polymorphism in other hazelnut accessions. In contrast,

the current set of tested SSR markers was randomly selected from the thousands of primers that were designed. Low variability is not unexpected because hazelnuts have been selected for similar characters and certain allelic combinations which may reduce their diversity (Campa et al. 2011; Martins et al. 2015; Ozturk et al. 2017a, b). Moreover, cultivars are propagated as clones, a fact that limits their diversity as compared to wild accessions.

Conclusion

New low-copy SSR markers were developed using next-generation sequencing technology and applied to the hazelnut cultivars for validation. Because climatic conditions, altitude, and soil can affect kernel, nut, and agro-morphological traits which often do not become visible until hazelnut is several years old, the SSR markers were assayed on 19 Turkish cultivars to find characteristic molecular markers. As a result, seven single-copy SSR primers were selected to discriminate Turkish cultivars so that hazelnut breeders, farmers, and geneticists can identify true-type hazelnuts and use identical clones. Turkish hazelnuts can be certified using these seven SSR markers, thereby solving an important problem in hazelnut nurseries and orchards. Moreover, these markers will be useful for new cultivar development.

Acknowledgements This research was supported by funding from an Izmir Institute of Technology Scientific Research Project (BAP-2016IYTE71) and The Scientific and Technological Research Council of Turkey (212T201). We are grateful to the Hazelnut Research Institute, Giresun, Turkey, for plant samples.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

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

References

Abuzayed MA, Goktay M, Allmer J, Doganlar S, Frary A (2017) Development of genomic simple sequence repeat markers in faba bean by next-generation sequencing. Plant Mol Biol Report 35:61–71

Akin M, Nyberg A, Postman J, Mehlenbacher S, Bassil NV (2016) A multiplexed microsatellite fingerprinting set for hazelnut cultivar identification. Eur J Hortic Sci 81:327–338

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Boccacci P, Botta R (2010) Microsatellite variability and genetic structure in hazelnut (Corylus avellana L.) cultivars from different growing regions. Sci Hortic 124(1):128–133

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 Hortic 1052:67–74

Balik HI, Beyhan N (2014) Clonal selection of Palaz hazelnut cultivar in Unye district of Ordu province. Anadolu J Agr Sci 29:179

Bassil NV, Botta R, Mehlenbacher SA (2005) Additional microsatellite markers of the European hazelnut. Acta Hortic 686:105–110

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. Genet Resour Crop Evol 60:543–568

Bhattarai G (2015) Microsatellite marker development, characterization and mapping in European hazelnut (*Corylus avellana* L.), and investigation of novel sources of eastern filbert blight resistance in Corylus. Dissertation, Oregon State University

Bhattarai G, Mehlenbacher SA (2017) In silico development and characterization of tri-nucleotide simple sequence repeat markers in hazelnut (*Corylus avellana* L.). PLoS One 12:e0178061

Bhattarai G, Mehlenbacher SA, Smith DC (2017) Eastern filbert blight disease resistance from *Corylus americana* 'Rush'and selection 'Yoder# 5'maps to linkage group 7. Tree Genet Genomes 13:45

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. Hortic Sci 43:667–672

Boccacci P, Botta R (2009) Investigating the origin of hazelnut (*Corylus avellana* L.) cultivars using chloroplast microsatellites. Genet Resour Crop Evol 56:851–859

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

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

Cardle L, Ramsay L, Milbourne D, Macaulay M, Marshall D, Waugh R (2000) Computational and experimental characterization of physically clustered simple sequence repeats in plants. Genetics 156:847–854

Celik I, Gultekin V, Allmer J, Doganlar S, Frary A (2014) Development of genomic simple sequence repeat markers in opium poppy by next-generation sequencing. Mol Breed 34:323–334

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

Erdogan V, Koksal AI, Aygun A (2010) Assessment of genetic relationships among Turkish hazelnut (Corylus avellana L.) cultivars by RAPD markers. Romanian Biotech Lett 15:5591–5601

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 (2017a) http://www.fao.org/faostat/en/#data/TP. Accessed 10.01. 2017

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

Fulton TM, Chunwongse J, Tanksley SD (1995) Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Mol Biol Report 13:207–209

Galderisi U, Cipollaro M, Di Bernanrdo G, De Masi L, Galano G, Cascino A (1999) Identification of hazelnut (*Corylus avellana* L.) cultivars by RAPD analysis. Plant Cell Rep 18:652–655

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

Ghanbari A, Akkak A, Boccacci P, Talaie A, Vezvaie A, Botta R (2005) Characterization of hazelnut (*Corylus avellana* L.) cultivars using microsatellite markers. Acta Hortic 686:111–115

Gökirmak T, Mehlenbacher SA, Bassil NV (2009) Characterization of European hazelnut (Corylus avellana) cultivars using SSR markers. Genet Resour Crop Evol 56(2):147–172

Gol S, Göktay M, Allmer J, Doğanlar S, Frary A (2017) Newly developed SSR markers reveal genetic diversity and geographical clustering in spinach (*Spinacia oleracea*). Mol Gen Genomics 292:847-855

Gupta PK, Balyan HS, Sharma PC, Ramesh B (1996) Microsatellites in plants: a new class of molecular markers. Curr Sci 45:45–54

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

Haq SU, Jain R, Sharma M, Kachhwaha S, Kothari SL (2014) Identification and characterization of microsatellites in expressed sequence tags and their cross transferability in different plants. Int J Genomics 2014:863948

Hokanson SC, Lamboy WF, Szewc-McFadden AK, McFerson JR (2001) Microsatellite (SSR) variation in a collection of *Malus* (apple) species and hybrids. Euphytica 118:281–294

Hormaza JI (2002) Molecular characterisation and similarity relationships among apricot (*Prunus armeniaca* L.) genotypes using simple sequence repeats. Theor Appl Genet 104:321–328

Islam A (2003) Clonal selection in 'Uzunmusa' hazelnut. Plant Breed 122:368–371

Islam A, Ozgüven AI (2000) Clonal selection in the Turkish hazelnut cultivars grown in Ordu province. In V International Congress on Hazelnut 556:203–208

Kafkas S, Doğan Y, Sabır A, Turan A, Seker H (2009) Genetic characterization of hazelnut (Corylus avellana L.) cultivars from Turkey using molecular markers. HortSci 44:1557–1561

Koressaar T, Remm M (2007) Enhancement and modifications of primer design program Primer3. Bioinformatics 23:1289–1291

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359

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

Lucchetti S, Pastore G, Leoni G, Arima S, Merendino N, Baima S, Ambra R (2018) A simple microsatellite-based method for hazelnut oil DNA analysis. Food Chem 245:812–819

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10–12

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 Report 33:1712-1720

Mehlenbacher SA (2014) Geographic distribution of incompatibility alleles in cultivars and selections of European hazelnut. J Am Soc Hortic Sci 139:191–212

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 (2017a) Molecular genetic diversity and association mapping of nut and kernel traits in Slovenian hazelnut (*Corylus avellana*) germplasm. Tree Genet Genomes 13:16

Ozturk SC, Balik HI, Balik SK, Kizilci G, Duyar O, Doganlar S, Frary A (2017b) Molecular genetic diversity of the Turkish national hazelnut collection and selection of a core set. Tree Genet Genomes 13:113

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

Peterschmidt BC (2013) DNA markers and characterization of novel sources of eastern filbert blight resistance in European hazelnut (*Corylus avellana* L.). MS Thesis, Oregon State Univ., Corvallis

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

Rallo P, Dorado G, Martín A (2000) Development of simple sequence repeats (SSRs) in olive tree (*Olea europaea* L.). Theor Appl Genet 101:984–989

Sathuvalli V, Mehlenbacher SA, Smith DC (2017) High-resolution genetic and physical mapping of the eastern filbert blight resistance region in 'Jefferson' hazelnut (L.). Plant Genome 10

Solar A, Stampar F, Usenik V (1996) Identification of some hazelnut cultivars (*Corylus avellana* L.) with isozyme analysis. IV International Symposium on Hazelnut 445

Sosinski B, Gannavarapu M, Hager LD, Beck LE, King GJ, Ryder CD, Rajapakse S, Baird WV, Ballard RE, Abbott AG (2000) Characterization of microsatellite markers in peach [*Prunus persica* (L.) Batsch]. Theor Appl Genet 101:421–428

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I (2009) AbySS: a parallel assembler for short read sequence data. Genome Res 19:1117–1123

Squirrell J, Hollingsworth PM, Woodhead M, Russell J, Lowe AJ, Gibby M, Powell W (2003) How much effort is required to isolate nuclear microsatellites from plants? Mol Ecol 12:339–1348

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

Yamamoto T, Kimura T, Sawamura Y, Kotobuki K, Ban Y, Hayashi T, Matsuta N (2001) SSRs isolated from apple can identify polymorphism and genetic diversity in pear. Theor Appl Genet 102:865–870

Zalapa JE, Cuevas H, Zhu H, Steffan S, Senalik D, Zeldin E, McCown B, Harbut R, Simon P (2012) Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. Am J Bot 99:193–208

Zane L, Bargelloni L, Patarnello T (2002) Strategies for microsatellite isolation: a review. Mol Ecol 11:1–16