



Genetic variability is preserved among strongly differentiated and geographically diverse almond germplasm: an assessment by simple sequence repeat markers

Júlia Halász¹ · Ossama Kodad² · Gábor M. Galiba¹ · István Skola³ · Sezai Ercisli⁴ · Craig A. Ledbetter⁵ · Attila Hegedűs¹

Received: 25 September 2018 / Revised: 2 January 2019 / Accepted: 8 January 2019 / Published online: 21 January 2019
© The Author(s) 2019

Abstract

Eighty-six almond accessions of diverse geographic origin, ranging from Central Asia to the USA, were genotyped in 15 simple sequence repeat (SSR) loci to compare genetic diversity parameters, characterize genetic differentiation, and examine factors responsible for the maintenance of genetic diversity and population structure in almond. The mean allele number was 18.86 alleles per locus. All but one primer demonstrated polymorphic information content higher than 0.7. Almond genotypes clustered according to their pedigree and geographic origin. STRUCTURE analysis determined nine genetically distinct subgroups within the studied genotypes including the Kyrgyz, Akdamar, Bademli, Hungarian, Monor, Italian, Moroccan, and Californian accessions and wild species and an admixed group. An AMOVA analysis confirmed that considerable genetic variation occurred within populations (71.30%), and genetic variation among populations was also significant (28.70%). The mean values of the fixation index (F_{ST}) varied between 0.38 and 0.55, indicating marked genetic differentiation among the populations. The among-population genetic differentiation based on allele sizes (R_{ST}) was significantly higher than that based on allele identities (F_{ST}) between the most groups, suggesting that stepwise mutations have also contributed to genetic differentiation. A Mantel test and a neighbor-joining tree showed no significant correlation between the geographic distance and the genetic distance ($R_{xy} = 0.173$, $P = 0.226$) and indicated that geographic distance among the assessed populations has little influence on their genetic differentiation ($R_{xy} = 0.248$, $P = 0.194$). Our data show drift, mutations, and massive gene exchange between several wild species and cultivated *P. dulcis* as crucial components of genetic differentiation. Slight losses of genetic diversity are attributable to geographic isolation, human selection and not to the relatively recent occurrence of self-compatibility. There was no indication of a major decrease in genetic variability in almond germplasm from Asia to Europe. The present results reveal that almond domestication avoided the occurrence of a genetic bottleneck although its risk is present in many subpopulations.

Keywords Almond · Genetic variability · Genetic differentiation · Microsatellites · *Prunus dulcis* · SSR

Communicated by M. Wirthensohn

✉ Júlia Halász
genetics.prunus@gmail.com

¹ Department of Genetics and Plant Breeding, Szent István University Budapest, P.O. Box 53, Budapest H-1518, Hungary

² Département d'Arboriculture, École Nationale d'Agriculture de Meknès, BP S/40, Meknès, Morocco

³ NARIC Fruitculture Research Institute, Research Station of Cegléd, Cegléd, Hungary

⁴ Department of Horticulture, Ataturk University, 25240 Erzurum, Turkey

⁵ Agricultural Research Service, Crop Diseases, Pests & Genetics Research Unit, United States Department of Agriculture, 9611 S. Riverbend Avenue, Parlier, CA 93648-9757, USA

Introduction

Almond [*Prunus dulcis* (Mill.) D.A. Webb; syn. *P. amygdalus* Batsch] is a species of the large and economically important *Rosaceae* family that is commercially grown worldwide for its kernels. The cultivated almond is thought to have originated in the arid mountainous regions of Central Asia (Gradziel 2011). Numerous related wild species (over 20) are also found growing in the mountains and deserts of Central Asia from western China through Iran and Turkey (Martínez-Gómez et al. 2007). Molecular results evidenced a dissemination of the cultivated almond from Central Asia (Iran) to the Eastern Mediterranean and subsequently to

the Western Mediterranean regions, to North America, and finally, to the southern hemisphere including South America and Australia.

Hybridization has played a central role in the evolutionary history of domesticated plants. Nuclear and chloroplast DNA markers supported in a complementary way, *P. fenzliana* as a probable ancestor of the cultivated almond (Zeinalabedini et al. 2010). Delplancke et al. (2012) detected high genetic diversity levels along with substantial and symmetric gene flow between the domesticated *P. dulcis* and the wild *P. orientalis*. Almond is a self-incompatible species which is governed by the highly polymorphic, multiallelic *S*-locus (Dicenta and García 1993a). Due to the genetically controlled self-incompatibility system, almond is one of the most polymorphic cultivated fruit species (Martínez-Gómez et al. 2007). Genetic diversity studies in almond have not revealed a direct relationship between the level of diversity and the origin of the germplasm (Szikriszt et al. 2011).

Molecular markers developed for *Prunus* are particularly useful to study genome evolution and structure, genetic diversity, and for fingerprinting accessions (Martínez-Gómez et al. 2003). Several marker types have been used to study almond species. The last decade showed that the molecular marker technique most intensively used and most appropriate for genetic variability studies in almond was the SSR (simple sequence repeats) analysis because SSRs are extremely abundant, dispersed relatively evenly throughout the genome and show codominant inheritance (Gupta et al. 1996). Since the first SSRs were described in peach (Cipriani et al. 1999), they have been developed in many other *Rosaceae* species, such as apricot, Japanese plum, and cherry (Dirlewanger et al. 2002; Messina et al. 2004; Mnejja et al. 2004). Transferability (being able to use an SSR developed in one species for other species) has been frequently reported, particularly for peach SSRs (Cipriani et al. 1999). The first set of almond SSRs was published by Testolin et al. (2004).

To date, several studies were carried out to characterize the SSR diversity of almond cultivars and genotypes originating in specific geographical regions. Xu et al. (2004) developed SSR markers for the phylogenetic analysis of almond accessions from China and the Mediterranean region. Genetic diversity of the Spanish national almond collection was characterized by Fernández i Martí et al. (2009). Twelve highly polymorphic SSR loci were selected to uniquely identify cultivars commonly grown in California, and to allow an accurate assessment of parent/offspring relationships among them (Dangl et al. 2009). Zeinalabedini et al. (2010) characterized Spanish, French, Italian, American, Iranian, Tunisian, Australian, Ukrainian, Portuguese, and Slovakian almond cultivars by chloroplast and nuclear SSRs. These studies established the value of SSR markers for distinguishing different genetic lineages and detected an extensive gene pool available to almond genetic improvement. Promising Iranian

almond genotypes, wild almonds, and related *Prunus* species were also characterized by SSR and EST-SSR markers (Zeinalabedini et al. 2012; Rahemi et al. 2012). Nuclear DNA markers showed that Moroccan genotypes were genetically different from the tested commercial cultivars and therefore formed a separate genetic pool (El Hamzaoui et al. 2013).

Genetic diversity parameters have not demonstrated marked differences in almond germplasm grown in different regions of the world. This current study was carried out to examine this phenomenon by analyzing accessions from Central Asia to California using highly polymorphic SSR markers. In addition, our analysis also included almonds of Hungarian origin, a previously unexamined region. We also aimed to characterize genetic differentiation and both natural and human imposed factors that are responsible for the maintenance of genetic diversity and population structure in almond.

Materials and methods

Plant material

Eighty-six almond accessions originating from different geographical regions were evaluated in the experiments. The Hungarian, Ukrainian, and Italian cultivars are kept in the collection of the Szent István University, Faculty of Horticultural Science, Department of Genetics and Plant Breeding. Other Hungarian samples were collected from old abandoned orchards in Gellérthegy (47° 29' 8.6748" N and 19° 3' 0.9468" E), Tétényi-fennsík (47° 24' 48.7" E), Monor (47° 20' 53.03" N and 19° 26' 24.464" E), and the germplasm collection of National Agricultural Research and Innovation Centre, Fruitculture Research Institute, Cegléd, Hungary. Leaf samples of the Moroccan accessions were collected from natural populations in the Rif Mountains, north of Morocco (33° 14' 0.456" N and 8° 29' 48.012" W) and Sais Plain in central regions of Morocco (33° 27' 28.8936" N and 5° 11' 1.824" W). Turkish wild almond genotypes originated from Bademli region, Erzurum (40° 27' 13.3704" N and 40° 54' 9.3924" E) and Akdamar Island in Lake Van (38° 20' 30.9084" N and 43° 2' 7.9044" E). Some wild-growing accessions were collected in Kyrgyzstan, near the city Osh (40° 28' 56.1432" N and 72° 43' 20.2764" E). Californian cultivars and three accessions of wild almond species (*Prunus tenella*, *P. arabica*, and *P. webbii*) were obtained from an experimental orchard of Agricultural Research Service (United States Department of Agriculture, Parlier, CA).

DNA extraction and PCR conditions

Genomic DNA was extracted from fully expanded young leaves using a DNeasy Plant Mini Kit (Qiagen, Hilden,

Germany). DNA concentrations and purification parameters were measured using a Nanodrop ND-1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA). A set of 12 SSRs and 5 EST-SSR primer pairs were selected on the basis of previous reports on different *Prunus* species, and included 9 for peach, 6 for almond, and 2 for plum (Table 1), covering eight linkage groups (G1 to G8). The forward primers were labeled with 6-FAM fluorescent dye for detection in a capillary genetic analyzer. PCR reactions were carried out in a PTC 200 thermocycler (MJ Research, Quebec, Canada) using the program described for the primers. For each 25 µl PCR reaction, approximately 20–80 ng of genomic DNA was used. Each reaction contained 10× DreamTaq™ Green buffer (Thermo Fisher Scientific), 4.5 mM MgCl₂, 0.2 mM of dNTPs, 0.2 µM of the adequate primers, and 0.75 U of DreamTaq™ DNA polymerase (Thermo Fisher Scientific). Finally, KCl and (NH₄)₂SO₄ were added at a ratio optimized for robust performance of DreamTaq™ DNA Polymerase.

Electrophoresis of PCR products and allele sizing

To check the PCR amplifications and determine the approximate sizes of the alleles amplified, 4 µl of the PCR products were separated by electrophoresis in 1.2% TAE agarose gels for 2 h at 100 V with DNA bands being visualized by ethidium bromide staining. Fragment lengths were estimated by comparison with the 1-kb DNA ladder (Promega, Madison, USA). To determine the exact size of the fragments, the fluorescently

labeled products were run on an automated sequencer (ABI Prism 3100 Genetic Analyzer, Applied Biosystems, Foster City, CA). For determination of fragment sizes (genotyping), GENOTYPER 3.7 software and the GS500 LIZ size standard (Applied Biosystems) were used.

Data analysis

Genetic relatedness among genotypes was studied by UPGMA (Unweighted Pair Group Method with Arithmetic averages) cluster analysis using Popgene 1.32 (<http://www.ualberta.ca/~fyeh/>). Bootstrap support for nodes in the UPGMA tree was estimated from 1000 replications using the Paleontological Statistics (Past) software v.2.17c (Hammer et al. 2001). A phylogenetic tree (dendrogram) was constructed using the TREEVIEW program (Page 1996). PopGene 1.32 software (Yeh et al. 1997) was also used for calculation of observed heterozygosity (H_o), expected heterozygosity (H_e), observed number of alleles (N_a), Shannon's information index (I), fixation index (F_{ST}), inbreeding coefficient (F_{IS}), and gene flow ($N_m = 0.25(1-F_{ST})/F_{ST}$). To assess the informativeness of the chosen markers, the average polymorphic information content (PIC) was calculated in each locus according to the formula given by Anderson et al. (1993).

To further analyze the genetic composition of almond accessions, a Bayesian approach was used to estimate the number of clusters with STRUCTURE 2.3.4. software (Pritchard et al. 2000). Because the analyzed genotypes were derived

Table 1 SSR loci from different *Prunus* species analyzed in the almond accessions studied, locus type, linkage group of their localization, annealing temperature, number of alleles detected, observed (H_o) and

expected (H_e) heterozygosity, fixation index (F), and polymorphic information content (PIC) values

| Primer name | Locus type | Linkage group | Species | Reference | Ta (°C) | Number of alleles | H_o | H_e | PIC | F |
|-------------|------------|---------------|---------|-------------------------|---------|-------------------|--------|--------|------|-------|
| ASSR17 | EST-SSR | – | Almond | Xu et al. 2004 | 60 | 13 | 0.5465 | 0.8553 | 0.81 | 0.36 |
| ASSR27 | EST-SSR | – | Almond | Xu et al. 2004 | 55 | – | – | – | – | – |
| ASSR63 | EST-SSR | – | Almond | Xu et al. 2004 | 55 | 8 | 0.6744 | 0.8055 | 0.77 | 0.16 |
| BPPCT 001 | SSR | G2 | Peach | Dirlewanger et al. 2002 | 57 | – | – | – | – | – |
| BPPCT 025 | SSR | G6 | Peach | Dirlewanger et al. 2002 | 57 | 16 | 0.8140 | 0.8917 | 0.83 | 0.09 |
| BPPCT 038 | SSR | G5 | Peach | Dirlewanger et al. 2002 | 57 | 20 | 0.7791 | 0.9149 | 0.89 | 0.15 |
| BPPCT018 | SSR | G6 | Peach | Dirlewanger et al. 2002 | 57 | 16 | 0.6912 | 0.7322 | 0.70 | 0.06 |
| BPTCT007 | SSR | G3 | Peach | Dirlewanger et al. 2002 | 57 | 21 | 0.6860 | 0.9045 | 0.90 | 0.22 |
| CPDCT025 | SSR | G3 | Almond | Mnejja et al. 2005 | 62 | 16 | 0.7907 | 0.8993 | 0.87 | 0.12 |
| CPPCT 044 | SSR | G2 | Peach | Aranzana et al. 2002 | 58 | 17 | 0.6279 | 0.8160 | 0.78 | 0.23 |
| CPPCT006 | SSR | G8 | Peach | Aranzana et al. 2002 | 59 | 18 | 0.7275 | 0.7404 | 0.72 | 0.02 |
| CPPCT022 | SSR | G7 | Peach | Aranzana et al. 2002 | 50 | 16 | 0.6914 | 0.7365 | 0.71 | 0.06 |
| CPSCT 021 | SSR | G2 | Plum | Mnejja et al. 2004 | 46 | 18 | 0.7791 | 0.8600 | 0.85 | 0.09 |
| CPSCT018 | SSR | G8 | Plum | Mnejja et al. 2004 | 52 | 18 | 0.7381 | 0.8124 | 0.79 | 0.09 |
| EPDCU 3083 | EST-SSR | G3 | Almond | Howad et al. 2005 | 57 | 20 | 0.7209 | 0.8625 | 0.85 | 0.16 |
| EPDCU 5100 | EST-SSR | G1 | Almond | Howad et al. 2005 | 57 | 4 | 0.1977 | 0.1822 | 0.17 | –0.09 |
| UDP 96005 | SSR | G1 | Peach | Cipriani et al. 1999 | 57 | 17 | 0.6977 | 0.8889 | 0.81 | 0.22 |

from different breeding programs and natural hybridization, the “admixture ancestry” and the “correlated allele frequency” parameters were used. K was set from 1 to 15, each run was replicated 10 times, with the burn-in period of 100,000 followed by 100,000 Markov Chain Monte Carlo (MCMC) repetitions. Estimation of the best K value was conducted with STRUCTURE Harvester (Earl and vonHoldt 2012) following the Evanno et al. (2005) method. Neighbor-joining clustering of geographic groups based on pairwise Nei's genetic distance values was also carried out using STRUCTURE.

Genetic diversity was partitioned using an analysis of molecular variance (AMOVA) as implemented in GenAlEx 6.503 (Peakall and Smouse 2012) to compute pairwise F_{ST} and R_{ST} values between taxa. Standard errors for F statistics were determined by jackknifing over loci using 999 permutations. A Mantel test was used to estimate the association between pairs of independent dissimilarity matrices: geographic distance and genetic distance, geographic distance and R_{ST} values as well as F_{ST} and R_{ST} values. Pairwise geographic distances among the analyzed populations were calculated according to the latitude and longitude of each site with Vincenty's formula (<http://www.movable-type.co.uk/scripts/latlong-vincenty.html>). The significance of estimates was tested based on 9999 permutations.

Results and discussion

Polymorphism analysis of the SSR markers

The amplification of genomic DNA of 86 genotypes collected in different geographical regions was successful in 15 of the 17 SSR loci, including 11 genomic and 4 EST-SSR primers developed from different *Prunus* species (peach, almond, and Japanese plum). All primer pairs produced a maximum of two alleles per genotype in accordance with the diploid genome of the species. Genotypes showing a single band were considered homozygous for that particular locus. Combining data of all efficient SSR markers, 15 primer pairs produced a total of 238 alleles ranging from four to 21 per locus (Table 1), with sizes from 86 to 230 bp. Genomic SSRs were more polymorphic revealing 193 alleles, while the EST-SSRs were less polymorphic by detecting 45 alleles. The mean value was 18.86 alleles per locus. SSR primers derived from the non-coding DNA region detected an average of 17.54 alleles per locus. This value is much higher than 8.4 reported by Testolin et al. (2004), 4.7 obtained by Martínez-Gómez et al. (2003), 6.3 obtained by Xie et al. (2006), 14.9 shown by El Hamzaoui et al. (2013), and slightly higher than 17.21 alleles per locus reported by Fernández i Martí et al. (2009) in other *Prunus* diversity

studies. However, it is slightly lower than 18.66 per locus reported by Fernández i Martí et al. (2015) and 18 alleles per locus obtained by Distefano et al. (2013). The average number of alleles for EST-SSR primers, originating from coding DNA regions, was 11.25, which is lower than that obtained by genomic SSR markers.

Among EST-SSR loci, EPDCU 3083 amplified the highest number of alleles (20), while EPDCU5100 detected the lowest number of alleles, only four (Table 1). The most informative locus was BPTCT007 with a polymorphic information content (PIC) value of 0.90, whereas, the EPDCU 5100 locus was the least informative marker with a PIC value of 0.17. This parameter provides an estimate of the discrimination power of a molecular marker by taking into account not only the number of alleles per locus but also their relative frequencies in each studied population (Anderson et al. 1993).

The average PIC value of EST-SSR loci (0.66) was lower than that of genomic SSR loci (0.80). The present results fit well with previous studies that suggested that EST-SSRs were less polymorphic than their genomic counterparts in other almond genotypes (Rahemi et al. 2012) and other species (Tahan et al. 2009). However, they are in contrast with results of Xie et al. (2006), reporting that genomic and EST-SSR primers amplified similar numbers of alleles. That study analyzed only 38 almond cultivars with five more peach and almond \times peach hybrids. The fewer analyzed genotypes and the inclusion of peach, a more homozygous species into the sample set, or differences in the applied markers may explain the identical performance of genomic and EST-SSRs, which is in contrast with our findings. The results reported here suggest that SSR markers are a very suitable tool for assessing genetic diversity of almond genotypes and genomic SSRs may provide more information than SSR regions in the coding part of the genome.

The values of expected and observed heterozygosity were compared by the fixation index (F) that ranged between -0.09 (EPDCU 5100) and 0.36 (ASSR17) (Table 1). The observed heterozygosity ranged between 0.62 in CPPCT044 and 0.81 in BPPCT 025, with an average of 0.73 across the SSR loci. In the case of EST-SSRs, the average was much lower (0.53) and one locus (EPDCU 5100) revealed an exceptionally low level of heterozygosity (0.19). The average observed heterozygosity value calculated for the SSR makers (0.73) is very similar to the 0.72 value obtained by Fernández i Martí et al. (2009), but higher than the values 0.59 , 0.62 , and 0.62 reported by El Hamzaoui et al. (2013), Martínez-Gómez et al. (2003) and Fernández i Martí et al. (2015), respectively. In general, H_e values were higher than H_o , and hence, the values of fixation index were positive for 14 primers and was only negative for one EST-SSR primer. The average F value was 0.13 , which is lower than 0.23 reported by Fernández i Martí et al. (2015). This low level of heterozygosity deficit is in harmony with the self-incompatibility of this species but it may indicate a modest level of inbreeding (discussed later).

Genetic relationships and similarities among genotypes

The genetic relatedness and diversity among the 86 almond genotypes and cultivars were evaluated using UPGMA cluster analysis based on the similarity matrix generated by the Nei and Li (1979) coefficient. Genetic similarities ranged from 0.03 (*P. tenella* and ‘Ne Plus Ultra’) to 1.00 (three accessions from Akdamar island, Turkey) with an average of 0.29. Almond genotypes clustered according to their pedigree and geographic origin (Fig. 1). The 86 almond cultivars, genotypes, and wild species were classified into two groups of different size, with *P. tenella* forming an outgroup and separated from the rest of the genotypes. The present results indicate that *P. tenella* might not have contributed to the gene pool of modern cultivated almond. *P. tenella* is classified in the section *Chamaeamygdalus* Spach of subgenus *Amygdalus* (Browicz and Zohary 1996). *P. tenella* is considered as a sister to a clade composed of subgenera *Amygdalus* and *Prunus* (Lee and Wen 2001). Bortiri et al. (2006) studied the phylogenetic relationships among 37 *Prunus* species using data from internal transcribed spacer, chloroplastic intergenic spacer sequences, and 25 morphological characters, and confirmed that this species could be considered as a sister to the rest of subgenus *Amygdalus*.

STRUCTURE analysis was carried out to determine the genetic constitution of different groups. The Evanno criterion gave a strong signal for $K = 9$ indicating nine genetically distinct subgroups resided within the studied genotypes (Fig. 2). This analysis detected the Kyrgyz, Akdamar, Bademli, Hungarian, Monor, Italian, Moroccan, and Californian accessions and wild species as genetically distinct groups, all receiving significant bootstrap support (Fig. 1). The results from Bayesian clustering analysis were in harmony with the groupings we detected in the molecular phylogenetic dendrogram (Fig. 1) with some admixed accessions (e.g., all Ukrainian cultivars and some California, Hungarian, and Moroccan accessions).

The second group is formed only by Italian cultivars and one accession of *Prunus webbii*. All these cultivars are self-compatible (Socias i Company 1990). *P. webbii* self-compatibility had been suggested by different authors (Godini 1979; Socias i Company 1990) as a possible origin of self-compatibility in the almond population of Puglia. Godini (2000) also pointed out some morphological similarities between *P. webbii* and some cultivars from Puglia as support for this hypothesis. Further molecular studies have verified this hypothesis, providing unequivocal evidence for gene flow between wild and cultivated almond species (Martínez-Gómez et al. 2003). Later, Bošković et al. (2007) *S*-genotyped Apulian *P. dulcis* and *P. webbii* accessions and speculated that *S_F* might have arisen within *P. dulcis* and its occasional occurrence in *P. webbii* is due to introgression from *P. dulcis*. Our

results seem to support this hypothesis since *P. webbii* carried some Apulian alleles (Fig. 1) and most self-compatible (SC) Italian cultivars did not contain “wild” alleles associated with *P. webbii*. However, regardless of the direction of introgressive pollen transfer between the two species, the clustering of *P. webbii* with the Italian almond cultivars in the present study confirms the hypothesis that when almond cultivars moved toward the Mediterranean, new hybridizations might have occurred, especially with the wild Mediterranean species *P. webbii*, resulting in some of the almond populations found along the northern shore of the Mediterranean sea (Socias i Company 2004).

The analysis of genetic distance shows that ‘Tuono’ and ‘Supernova’ are distinct cultivars. This result agrees with those reported by Fernández i Martí et al. (2009). However, using SSR analysis, Marchese et al. (2008) showed that ‘Supernova’ was very similar to ‘Tuono’. Later, Distefano et al. (2013) and Dicenta et al. (2015) also found ‘Supernova’ and ‘Tuono’ to be identical. Marchese et al. (2008) reported that growers in the Agrigento area, Sicily, had commented on the similarity of the nuts of ‘Supernova’ and ‘Tuono’. While tree characters were also said to be very similar, they also presented small differences in their growth habit and flowering time. These differences support our results and those reported by Fernández i Martí et al. (2009). The cultivars ‘Filippo Ceo’ and ‘Falsa Barese’ exhibited similar SSR profiles. These genotypes also show similar agronomic and phenotypic traits (De Giorgio and Polignano 2001). Our results suggest that these cultivars have the same origin or one may have originated from the other.

The third clade contained almost all Moroccan genotypes included in this study and a *Prunus arabica* accession. The subgroup formed only by Moroccan genotypes agrees with the results reported by El Hamzaoui et al. (2013) on the comparison of almond gene pool of Mediterranean and Californian cultivars. Similar results were reported by Delplancke et al. (2013) who by combining nuclear and chloroplast microsatellites suggested the presence of endemic alleles in Morocco that could either reflect early introductions or relict natural populations of wild almonds in Moroccan glacial refugia (Médail and Diadema 2009). The clustering of Moroccan genotype (K4) with a *P. arabica* accession indicates the contribution of this species in the evolution of cultivated almond. This species is native to Middle East regions (Iraq, Jordan, and Syria) (Meikle 1966). It is probable that in these regions, natural hybridization of *P. dulcis* with other species such as *P. arabica* might have occurred. Recently, Delplancke et al. (2016) using molecular analysis reported that *P. arabica* and *P. dulcis* clustered in the same clade, suggesting that different wild lineages could have spontaneously contributed to the genome of cultivated almond. Furthermore, archeological and ethnographical data reported a long use of wild almonds (*P. orientalis*, *P. arabica*, *P. bucharica*) in different parts of

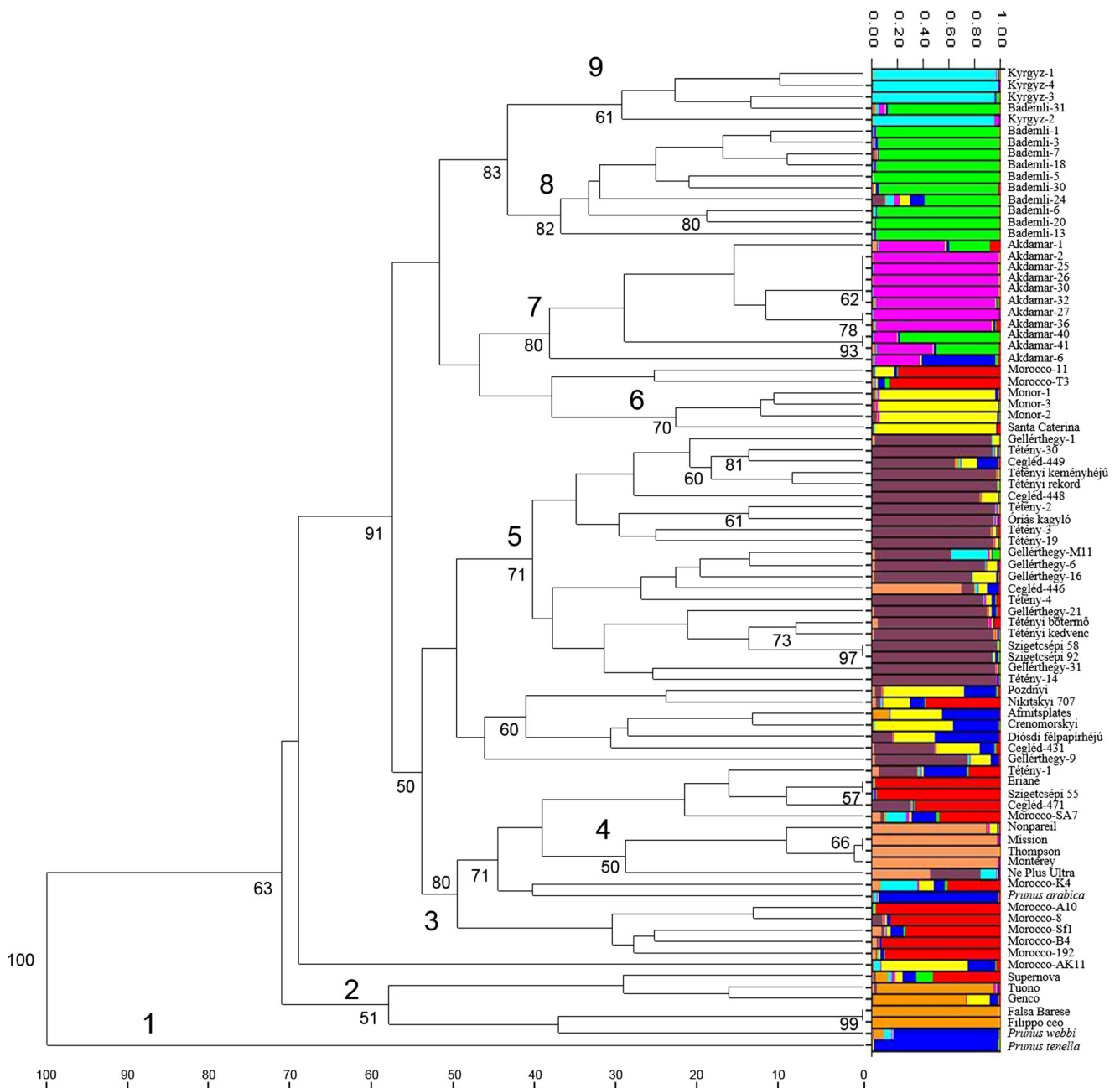


Fig. 1 Genetic distance and structure analysis of the 86 almond cultivars and genotypes studied using 15 SSR markers. The dendrogram was constructed based on UPGMA analysis using the similarity matrix generated by the Nei and Li (1979) coefficients. The numbers at specific nodes indicate percentage of 1000 bootstrap replicates in which a given group was found (< 50% not shown). The bar plots of individual almond

accessions were generated using the STRUCTURE 2.3.4. software. The nine reconstructed populations are distinguished by different colors and their clades are labeled by numbers on the dendrogram. Each accession is presented by a thin bar divided into segments indicating its genetic background. Multiple colors show admixed genetic constitution of each individual

their distribution area (Martinoli and Jacomet 2004). Probably, *P. arabica* or its interspecific hybrids were also introduced into Morocco by the Carthaginians between the fifth and fourth century BP (El Khatib-Boujibar 1983) or by the Arabs during the sixth and seventh centuries (Kester et al. 1991).

California cultivars ('Nonpareil', 'Ne Plus Ultra', 'Mission', 'Monterey', and 'Thompson') formed a separate

subcluster (4). The cultivars 'Ne Plus Ultra' and 'Nonpareil' showed some admixed structure and positioned relatively far compared to the rest of California almonds which were more uniform. 'Ne Plus Ultra', along with 'Nonpareil', are considered founding cultivars of the California almond industry, and originated from the same seedling planting in the late 1800s (Wood 1925). 'Monterey' is an offspring of 'Nonpareil', and the parentage of 'Thompson' is unknown (Brooks and Olmo

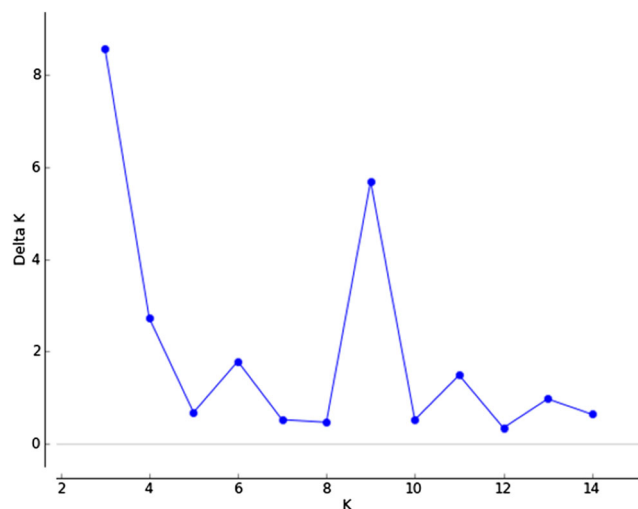


Fig. 2 Estimation of the optimum number of clusters for almond genotypes using the Evanno's method. The graph displays the DeltaK for each K value and indicates $K = 9$ as the uppermost probable number of genetically homogenous groups of the analyzed samples

1952), but our results support previous assumptions based on randomly amplified DNA analysis and S -genotypes pointing to 'Nonpareil' as one of its putative parent (Bartolozzi et al. 1998). 'Mission' also seems to be very close to 'Thompson'. Its putative Mediterranean origin hypothesized by Fernández i Martí et al. (2015) was supported by our results since alleles present in Akdmar Island (South of Turkey) were detected in the US cultivars.

Two sub-clades have an intermixed pattern of cultivars with different origins. One of the sub-clades encompassed Hungarian, French, and Moroccan genotypes while the other grouped some Ukrainian and Hungarian accessions. STRUCTURE analysis identified an admixed gene pool as the reasons for such an interweaving cluster of different cultivars. 'Szigetcsépi 55' showed only minor differences to 'Eriane' in two of the assayed loci. Halász et al. (2010) determined the S -genotypes of two 'Szigetcsépi 55' accessions in a germplasm collection. One of those shared an identical S -genotype with 'Eriane' and our SSR data now support that they must be closely related or identical accessions kept under different names. The placement of 'Eriane' in this subgroup with a Moroccan genotype is an indication of the intense connections among Mediterranean countries and germplasm exchange ranging from Phoenician trading to collecting trips in the second half of the twentieth century (Gradziel 2011).

A subgroup of the Ukrainian cultivars is formed with some Hungarian accessions and 'Afrnitsplates' (of supposed African origin). Ukrainian cultivars ('Pozdnyi', 'Nikitskyi 707' and 'Crenomorskyi'), as well as Hungarian accessions 'Diódsi felpapírhéjú', Cegléd-451, and Gellérthegey-9 carry some alleles from the European and wild groups. Ukrainian almond breeding was based on a germplasm collection having

accessions from France (Yezhov et al. 2005) and several wild species (Gradziel 2011). The Nikitskyi clones were selected at the Nikita Botanical Garden of Yalta but they were reported to be introduced from France, which was also supported by clustering with Iranian and Mallorcan cultivars (Fernández i Martí et al. 2015).

The major part of cluster 5 includes 22 accessions originating from Hungary embracing old cultivars, rootstocks, and naturalized trees from countryside and abandoned orchards in three locations of Hungary. Cultivars with known complete or partial pedigrees were clustered according to their parentage. For example, 'Szigetcsépi 58' and 'Szigetcsépi 92' are seedlings of 'Diódsi felpapírhéjú' while 'Tétényi kedvenc', 'Tétényi keményhéjú', and 'Tétényi rekord' are in half-sib or full-sib relationships (Halász et al. 2010). Several bitter-kernelled accessions are positioned among sweet-kernelled Hungarian cultivars in this cluster. Cegléd accessions are old trees of unknown origin used for producing seed-propagated rootstocks. Gellérthegey accessions are bitter-kernelled almond trees used as rootstock in a peach orchard established in the second half of the nineteenth century (Rapaics 1940). After peach orchards had been devastated, the rootstock trees grew up and formed a picturesque part of the Danube banks in Budapest that is included on the UNESCO World Heritage List. The co-clustering of sweet- and bitter-kernelled almonds is not unexpected since this trait is determined by a single gene with the sweet allele being dominant to the bitter (Dicenta and García 1993b).

Cluster 6 includes mainly Turkish, some Moroccan, and Hungarian accessions as well as cultivar 'Santa Caterina', probably originating in the Mediterranean region of the Middle East. This grouping suggests that the very old trees sampled in Monor (Hungary) may have relationships with Mediterranean genotypes. Almond was first brought to Hungary (Pannonia) by the Romans (Rapaics 1940; Gradziel 2011) and later by Latin monks and traders in the Middle Ages. It is further supported by etymology since the Latin mandorla/mandola reutilized and altered to "mandula" in Hungarian language.

Turkish genotypes clustered in two groups; accessions from 'Akdamar' island in Lake Van (cluster 7) were separated from those collected in 'Bademli' village (Erzurum) (cluster 8). Bademli-31 was placed among Kyrgyz accessions indicating their genetic relatedness. This is probably explained by its location close to Silk Road trading routes (Gradziel 2011). It is very interesting to consider the separation of the Bademli and Akdamar accessions and that genotypes Akdamar 2; 25–26; and 30, 32 are very similar, as are Akdamar 27 and 36 as well as Akdamar 40 and 41. The latter two and Akdamar 1 seem to share some Bademli alleles, as well. Genetic diversity analysis was carried out to clarify the phylogenetic patterns of Turkish and other almond accessions and genetic differentiation of the major reconstructed populations.

Genetic diversity and differentiation of major reconstructed populations

The genetic diversity parameters were further investigated in the major populations reconstructed by the STRUCTURE analysis. Wild species as single representatives of *P. arabica*, *P. webbii*, and *P. tenella* were neglected in some analyses as were Kyrgyz and Monor groups since we had less than five individuals in those groups. An AMOVA analysis declared that considerable genetic variation occurred within populations (71.30%), and genetic variation among populations was also significant, reaching a value of 28.70% (Table 4). This level of variation among populations is much higher than the value estimated for *P. sibirica* (Wang et al. 2014) or *P. mahaleb* (Jordano and Godoy 2000), and that shown by Fernández i Martí et al. (2015) in almond.

The highest number of alleles for most of the used markers (Table 2) was detected in the Hungarian and Moroccan accessions, while the lowest number of alleles occurred in California cultivars. In Turkish genotypes, the mean number of alleles was lower in the almond population collected from Akdamar Island (4.45) compared to those collected from the Bademli inland region (5.09), and higher than the values of California (3.09) and Italian (3.63) genotypes. Most unique alleles were detected in Hungarian accessions, while unique alleles were not identified in California cultivars. Since allele frequencies are influenced by the number of accessions studied, Akdamar, Bademli, and Morocco can be directly compared since 11 accessions were tested in each of those regions. The genetic diversity of Akdamar almonds seem to be lower compared to Bademli and Moroccan accessions.

The highest observed heterozygosity was shown in the Turkish accessions from Akdamar Island (0.77), while the lowest value occurred in the group of Italian cultivars (0.58). These values remain high across evaluated populations since almond is a self-incompatible species and are similar to the data reported in the literature (Martínez-Gómez et al. 2003; El Hamzaoui et al. 2013; Fernández i Martí et al. 2015). The mean value of the fixation index (F_{ST}) varied between 0.38 (Turkish Akdamar Island population) and around 0.55 for both the Moroccan and Hungarian genotypes. These values are positively high indicating a marked genetic differentiation among the populations. A gene flow value (N_m) greater than one is “strong enough” to prevent substantial differentiation due to genetic drift (Slatkin and Barton 1989). Its value was much lower than one in each population presenting an evidence of restricted gene flow (Table 2), further supporting the basis of strong genetic differentiation among populations.

As stepwise mutation is reflected by R_{ST} values (Slatkin 1995) and it proved to be more efficient to assess population genetic differentiation in *Prunus* species (Mihretie et al. 2015), we determined pairwise R_{ST} values among major reconstructed populations (Table 3). The mean value of genetic

differentiation among all populations was significant ($P < 0.01$) both based on unordered alleles ($F_{ST} = 0.168$) and ordered alleles ($R_{ST} = 0.289$). Only R_{ST} values between pairs of the Hungarian and Admixed (composed of Ukrainian and some Hungarian accessions) as well as the Monor and Admixed groups were not statistically significant ($P \leq 0.01$). Its explanation is illustrated in Fig. 1, showing the cultivars in admixed group within cluster 5 share some common alleles with the Hungarian and Monor groups. These latter two groups are isolated from each other, which is indicated by their strong bootstrap support and their significantly different R_{ST} values. Although F_{ST} was found to be more suitable in a study on *P. sibirica* (Wang et al. 2014), R_{ST} proved to be more sensitive in *P. africana* (Mihretie et al. 2015). In our study, only slight differences were shown between F_{ST} and R_{ST} values, which was confirmed by a Mantel test showing significant association of F_{ST} and R_{ST} values ($R_{xy} = 0.636$, $P = 0.001$). Between the pairs of populations of Kyrgyz and Monor, Monor and Morocco, Akdamar and Italian, Hungarian and Admixed, Monor and Admixed as well as Monor and Wild species, the values for F_{ST} were significantly higher than R_{ST} indicating that drift-driven differentiation is more probable between such populations, while stepwise mutations have also contributed to differentiation between the rest of populations (Hardy et al. 2003) (Table 4).

A Mantel test showed no significant correlation between the geographic distance and the genetic distance ($R_{xy} = 0.173$, $P = 0.226$) as well as between the geographic distance and the R_{ST} values ($R_{xy} = 0.248$, $P = 0.194$), suggesting that geographic distance among the assessed populations has little influence on their genetic differentiation. This is also confirmed by genetic relationships among the almond groups defined by STRUCTURE and assessed based on Nei's genetic distances and the neighbor-joining algorithm (Fig. 3). Populations in relative geographic proximity show considerable genetic distance, such as between Hungarian accessions and those in Monor group (separately clustering from the rest of Hungarian genotypes) or Turkish groups in Bademli and Akdamar.

The inbreeding coefficient (F_{IS}) had low positive values for the Hungarian, Bademli, and Moroccan accessions, as well as self-compatible Italian cultivars, which indicates low-level inbreeding in these populations. Similar findings related to heterozygosity deficiency have been observed in other almond populations (Xie et al. 2006; Fernández i Martí et al. 2009; El Hamzaoui et al. 2013). A deficiency of heterozygosity in the outcrossing tropical species *Sextonia rubra* (Mez) van der Werff was considered to be an effect of biparental breeding, resulting from limited pollen dispersal among relatives (Veron et al. 2005). Half-sibling mating was shown by Hadziabdic et al. (2012) to be responsible for the deficiency in heterozygosity in dogwood accessions growing in a small geographical area.

Table 2 Results of the analysis of molecular variance (AMOVA) for 86 almond accessions grouped in 10 populations

| Source of variation | df | Sum of squares | Estimated variance | Percentage of variation | P value |
|---------------------|-----|----------------|--------------------|-------------------------|---------|
| Among populations | 9 | 61,923.605 | 368.107 | 29% | 0.001 |
| Among individuals | 76 | 75,871.093 | 83.989 | 6% | 0.016 |
| Within individuals | 86 | 71,408.000 | 830.326 | 65% | 0.001 |
| Total | 171 | 209,202.698 | 1282.421 | 100% | |

df, degree of freedom

The deficit of heterozygosity in the analyzed almond trees could be, in part, due to the frequent use of superior almond genotypes for propagation (Lansari et al. 1994; Ercisli 2004; Gradziel 2011). This leads to a loss of genetic diversity and therefore strengthens the hypothesis for the existence of a certain level of inbreeding within local almond tree populations. This might be the only factor in the Moroccan gene pool, which is characterized by high genetic diversity reflected by the average number of alleles and observed heterozygosity. This fact could be due to the multiple and recurrent introductions that avoided the domestication bottleneck and led to higher local diversity (Pickersgill 1998). Delplancke et al. (2013) confirmed that the Moroccan almond gene pool has been formed by numerous exchanges between southern Europe and North Africa. This has led to the observed high diversity and hence Morocco could be considered a hotspot of the secondary diversity of almond.

The Hungarian gene pool showed a high number of alleles indicating great genetic diversity. However, H_o is lower than H_e , and this slight deficiency in heterozygosity might be attributed to human selection and breeding activity performed through several decades on this material. Tétényi and Szigetcsépi cultivars can be traced back to a limited number of offspring from ‘Burbank seedling’, advanced selections Budatétényi 6 and 7, and ‘Diódsi felpapírhéjú’ (Halász et al. 2010). This co-ancestry may result in a slight loss of heterozygosity.

In the case of Italian cultivars, a heterozygosity deficiency could be explained as a process of partial selfing that occurred among genotypes taking into account that all analyzed cultivars are self-compatible. Godini (2000) and Socias i Company (1990) reported that in the Italian cultivars, self-compatibility (the allele S_f at the S -locus) could have been transferred spontaneously into cultivars from *P. webbii*. This gene flow between crop and wild species could also account for the high diversity observed in self-compatible Italian genotypes. Fernández i Martí et al. (2009) reported that no differences between the heterozygosity levels of self-compatible (released from breeding program in the last two decades) and self-incompatible cultivars were observed. It is not surprising since all self-compatible cultivars are heterozygous for self-compatibility (Socias i Company 1990) and selfing is not used in almond breeding programs because of resultant inbreeding depression symptoms (Martínez-García et al. 2012). The half-sib mating coupled with the relatively recent occurrence of self-compatibility in almond, detected only in Puglia region in Italy, has probably not advanced through enough generations to accumulate lethal and deleterious alleles in the progeny, as observed in a self-compatible breeding program (Socias i Company 2002), and preventing a serious loss of genetic diversity (Szikriszt et al. 2011).

Only California cultivars and Akdamar group showed negative F_{IS} values indicating excess of heterozygotes in those populations. For the California cultivars, all studied

Table 3 Diversity indices of reconstructed major populations with ≥ 5 individuals suggested by the STRUCTURE analysis

| Major populations | Hungary | Californian cultivars | Akdamar Island, Turkey | Bademli region, Turkey | Morocco | Self-compatible cultivars |
|--------------------------------------|---------|-----------------------|------------------------|------------------------|---------|---------------------------|
| Average number of alleles (N_a) | 8.45 | 3.09 | 4.45 | 5.09 | 7.09 | 3.63 |
| Average number of unique alleles | 1.36 | 0.00 | 0.36 | 0.73 | 0.46 | 0.27 |
| Shannon's Information index (I) | 1.65 | 0.88 | 1.17 | 1.28 | 1.60 | 1.06 |
| Expected (H_e) heterozygosity | 0.74 | 0.55 | 0.64 | 0.66 | 0.75 | 0.64 |
| Observed (H_o) heterozygosity | 0.66 | 0.60 | 0.77 | 0.61 | 0.66 | 0.58 |
| Wright's fixation index (F_{ST}) | 0.55 | 0.39 | 0.38 | 0.52 | 0.54 | 0.50 |
| Inbreeding coefficient (F_{IS}) | 0.11 | -0.09 | -0.20 | 0.08 | 0.12 | 0.09 |
| Gene flow (N_m) | 0.21 | 0.39 | 0.42 | 0.23 | 0.21 | 0.25 |
| Number of samples | 27 | 5 | 11 | 11 | 11 | 5 |

Number of alleles (N_a), average number of unique alleles, Shannon's information index (I), expected heterozygosity (H_e), observed heterozygosity (H_o), fixation index (F_{ST}), inbreeding coefficient (F_{IS}), gene flow (N_m), and the number of samples

Table 4 Pairwise F_{ST} values (below diagonal) and pairwise R_{ST} values (above diagonal) between 10 populations

| | KG | WS | HU | MN | CA | AK | BA | MO | SC | ADM |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|----------|
| KG | 0.000 | 0.140 | 0.329 | 0.204 | 0.289 | 0.412 | 0.388 | 0.344 | 0.420 | 0.337 |
| WS | 0.214 | 0.000 | 0.313 | 0.119 | 0.287 | 0.406 | 0.326 | 0.323 | 0.459 | 0.228 |
| HU | 0.150 | 0.114 | 0.000 | 0.159 | 0.228 | 0.334 | 0.277 | 0.176 | 0.466 | 0.038 ns |
| MN | 0.278 | 0.167 | 0.096 | 0.000 | 0.315 | 0.247 | 0.317 | 0.097 | 0.275 | 0.027 ns |
| CA | 0.203 | 0.161 | 0.112 | 0.235 | 0.000 | 0.448 | 0.384 | 0.223 | 0.484 | 0.265 |
| AK | 0.230 | 0.201 | 0.166 | 0.199 | 0.262 | 0.000 | 0.218 | 0.240 | 0.186 | 0.324 |
| BA | 0.151 | 0.150 | 0.126 | 0.172 | 0.223 | 0.168 | 0.000 | 0.209 | 0.453 | 0.332 |
| MO | 0.162 | 0.079 | 0.086 | 0.113 | 0.104 | 0.160 | 0.134 | 0.000 | 0.336 | 0.126 |
| SC | 0.294 | 0.183 | 0.187 | 0.225 | 0.274 | 0.255 | 0.230 | 0.155 | 0.000 | 0.411 |
| ADM | 0.182 | 0.096 | 0.060 | 0.128 | 0.138 | 0.195 | 0.158 | 0.056 | 0.182 | 0.000 |

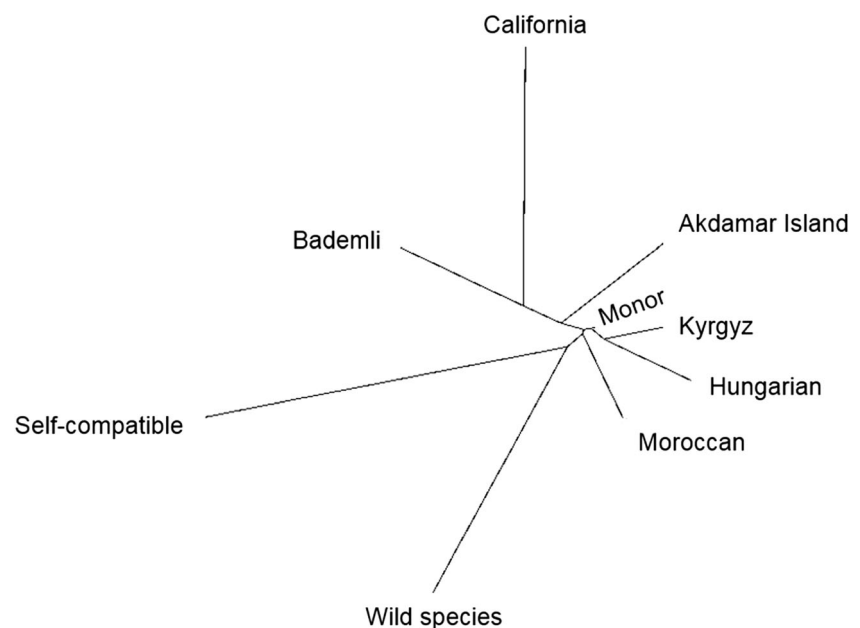
ns, non-significant values at $P \leq 0.01$ significance level

accessions are genetically related and have high degree of co-ancestry (Lansari et al. 1994). The cultivars ‘Monterey’ and ‘Thompson’ are chance seedlings, probably originated from the ‘Nonpareil’ \times ‘Mission’ cross (Asai et al. 1996). Lansari et al. (1994) reported that although many almond cultivars have ‘Nonpareil’ and ‘Mission’ as parents, the mean inbreeding coefficient for US almond cultivars is low, which was also confirmed by our analyses.

The F_{IS} value was the lowest in the Akdamar population, growing in an isolated geographical area. Although Akdamar Island is located approx. 3.5 km from land, a distance that pollinator insects are capable of achieving through flight, bees will rarely fly that distance when over water (Tautz et al. 2004). Ellstrand and Elam (1993) reported that population isolation may lead to stochastic differentiation by genetic drift. In this case, founder effect also contributes to the genetic differentiation with a small number of individuals originally

reaching the island. This is well reflected by the relatively low average allele number and the fact that only some accessions carry alleles that are relatively frequent in another Turkish almond population (Bademli). However, in a self-incompatible species, frequency-dependent selection acting on the S -locus will definitely have an effect on other loci, as well. It will allow mating only between trees with differing S -genotypes that help maintain high diversity under these specific conditions.

All these data indicate that genetic diversity parameters in almond do not show a biased variation along the dissemination routes from its putative center of origin in Central Asia (Martínez-Gómez et al. 2007). This is a sharply different pattern compared to what was shown for apricot, which is characterized by a considerable loss of genetic diversity in Western Europe and the Mediterranean basin (Bourguiba et al. 2012). This phenomenon was found to be linked to changes in the

Fig. 3 Neighbor-joining clustering of geographic groups based on pairwise Nei's genetic distance values

matting strategy (Halász et al. 2007), with self-compatibility arising putatively in the Eastern part of Turkey (Halász et al. 2013). As this important character spread rapidly into the Western regions, it induced a domestication bottleneck. A slight loss of heterozygosity occurred in the almond germplasm in most regions, which might be attributed to several factors. However, the extent of inbreeding is limited, even in the group of self-compatible cultivars. This might be explained by the relatively recent occurrence of this trait in almond and a noted inbreeding depression after selfing, which makes breeders avoid the use of self-crosses.

Genetic differentiation was shown to be significant in most assayed populations. In most cases, it is driven by drift but mutations might also contribute. The gene flow between cultivated almond and wild species growing in close proximity of the tested populations (e.g., *P. webbii* and *P. arabica*) was shown to be a crucial factor for preserving genetic diversity and differentiation among populations. Several former studies support the bidirectional gene flow among different almond-related species (Godini 2000; Bošković et al. 2007; Zeinalabedini et al. 2010; Delplancke et al. 2012). In addition, a *P. webbii* S-allele was also detected in a Bademli accession (data not shown). This might be additional proof of gene flow, as this study identified common SSR alleles shared by the Italian self-compatible cultivars and *P. webbii* (Fig. 1). Seeds of almond-related wild species were also collected by people, which has resulted in genetic impacts as verified in a recent study (Sorkheh et al. 2017). Based on such findings, almond could efficiently preserve its genetic variability along its dissemination routes from the center of origin and throughout the continents, which can be attributed to self-incompatibility and both massive and symmetrical gene exchange between wild species and cultivated *P. dulcis*. However, drift, especially in isolated geographic regions, was shown to induce a moderate loss of genetic diversity and resulted in consequences similar to those of human imposed selection during cultivar improvement due to the frequent use of superior almond genotypes in crosses. Such impacts may lead to the loss of genetic diversity and our results suggest a moderate level of inbreeding in most of the assayed and genetically differentiated subpopulations. Thus, the present results indicate almond domestication avoided the occurrence of a genetic bottleneck, although its risk is present in many subpopulations.

Acknowledgements Open access funding provided by Szent István University (SZIE). The authors thank Ádám Gutermuth and Róbert Rácz-Szabó for providing plant material and Endre Tóth for his kind help.

Data archiving statement The authors declare that all the work described in this manuscript followed standard *Tree Genetics and Genomes* policy. All parameters for the genomic and EST-SSR primers used in this study are available in Table 1.

Author contribution statement Júlia Halász had the original concept, performed part of the PCRs and data evaluation, and drafted the manuscript. Gábor M. Galiba participated in wet lab work. István Skola,

Ossama Kodad, Sezai Ercisli, and Craig A. Ledbetter provided specific knowledge on the Hungarian, Moroccan, Turkish, and Californian almond germplasm, respectively, and contributed to data interpretation. Attila Hegedűs performed bioinformatics analysis. Ossama Kodad and Attila Hegedűs wrote parts of the paper and carefully revised the manuscript.

Funding information This research was supported by the Higher Education Institutional Excellence Program (1783-3/2018/FEKUTSTRAT) awarded by the Ministry of Human Capacities, Hungary, within the framework of plant breeding and plant protection research of Szent István University. This work was financed by the OTKA K 112554 grant and the Moroccan-Hungarian Inter-Institutional S&T Cooperation Program grant (TÉT-12-MA-1-2013-0021) of the National Research and Technology of Morocco (CNRST) and of the National Research, Development and Innovation Office of Hungary.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Anderson JA, Churchill GA, Autrique JE, Tanksley SD, Sorrells ME (1993) Optimizing parental selection for genetic linkage maps. *Genome* 36:181–186
- Aranzana MJ, Garcia-Mas J, Carbo J, Arús P (2002) Development and variability analysis of microsatellite markers in peach. *Plant Breed* 121:87–92
- Asai WK, Micke WC, Kester DE, Rough D (1996) The evaluation and selection of current varieties. In: Micke WC (ed) *Almond production manual*, vol 3364. University of California Publication, California, pp 52–60
- Bartolozzi F, Warburton ML, Arulsekhar S, Gradziel TM (1998) Genetic characterization and relatedness among California almond cultivars and breeding lines detected by randomly amplified polymorphic DNA (RAPD) analysis. *J Am Soc Hortic Sci* 123:381–387
- Bortiri E, Heuvel BV, Potter D (2006) Phylogenetic analysis of morphology in *Prunus* reveals extensive homoplasy. *Plant Syst Evol* 259: 53–71
- Bošković RI, Tobutt KR, Ortega E, Sutherland BG, Godini A (2007) Self-(in) compatibility of the almonds *P. dulcis* and *P. webbii*: detection and cloning of 'wild-type S f' and new self-compatibility alleles encoding inactive S-RNases. *Mol Genet Genomics* 278:665–676
- Bourguiba H, Audergon JM, Krichen L, Trifi-Farah N, Mamouni A, Trabelsi S, D'Onofrio C, Asma BM, Santoni S, Khadari B (2012) Loss of genetic diversity as a signature of apricot domestication and diffusion into the Mediterranean Basin. *BMC Plant Biol* 12:49
- Brooks RM, Olmo HP (1952) *Register of new fruit and nut varieties*, 2nd edn. University of California Press, Berkeley, California, pp 1–7
- Browicz K, Zohary D (1996) The genus *Amygdalus* L. (Rosaceae): species relationships, distribution and evolution under domestication. *Genet Resour Crop Evol* 43:229–247
- Cipriani G, Lot G, Huang WG, Marrazzo MT, Peterlunger E, Testolin R (1999) AC/GT and AG/CT microsatellite repeats in peach [*Prunus*

- persica* (L.) Batsch]: isolation, characterisation and cross-species amplification in *Prunus*. Theor Appl Genet 99:65–72
- Dangl G, Yang J, Golino D, Gradziel T (2009) A practical method for almond cultivar identification and parental analysis using simple sequence repeat markers. Euphytica 168:41–48
- De Giorgio D, Polignano GB (2001) Evaluating the biodiversity of almond cultivars from germplasm collection field in Southern Italy. Sustain Global Farm 56:305–311
- Delplancke M, Alvarez N, Espindola A, Joly H, Benoit L, Brouck E, Arrigo N (2012) Gene flow among wild and domesticated almond species: insights from chloroplast and nuclear markers. Evol Appl 5: 317–329
- Delplancke M, Alvarez N, Benoit L, Espindola A, Joly IH, Neuenschwander S, Arrigo N (2013) Evolutionary history of almond tree domestication in the Mediterranean basin. Mol Ecol 22: 1092–1104
- Delplancke M, Yazbek M, Arrigo N, Espindola A, Joly H, Alvarez N (2016) Combining conservative and variable markers to infer the evolutionary history of *Prunus* subgen. *Amygdalus* sl under domestication. Genet Resour Crop Evol 63:221–234
- Dicenta F, García JE (1993a) Inheritance of self-compatibility in almond. Heredity 70:313–317
- Dicenta F, García JE (1993b) Inheritance of the kernel flavour in almond. Heredity 70:308–312
- Dicenta F, Sánchez-Pérez R, Rubio M, Egea J, Batlle I, Miarnau X, Palasciano M, Lipari E, Confolent C, Martínez-Gómez P, Duval H (2015) The origin of the self-compatible almond ‘Guara’. Sci Hortic-Amsterdam 197:1–4
- Dirlwanger E, Cosson P, Tavaud M, Aranzana M, Poizat C, Zanetto A, Arús P, Laigret F (2002) Development of microsatellite markers in peach [*Prunus persica* (L.) Batsch] and their use in genetic diversity analysis in peach and sweet cherry (*Prunus avium* L.). Theor Appl Genet 105:127–138
- Distefano G, Caruso M, La Malfa S, Ferrante T, Del Signore B, Gentile A, Sottile F (2013) Genetic diversity and relationships among Italian and foreign almond germplasm as revealed by microsatellite markers. Sci Hortic-Amsterdam 162:305–312
- Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361
- El Hamzaoui A, Oukabli A, Charafi J, Moumni M (2013) Moroccan almond is a distinct gene pool as revealed by SSR. Sci Hortic-Amsterdam 154:37–44
- El Khatib-Boujibar N (1983) Le Maroc et Carthage. Le Memorial du Maroc (I). Nord Organisation ed, 140
- Ellstrand NC, Elam DR (1993) Population genetic consequences of small population size: implications for plant conservation. Annu Rev Ecol Syst 24:217–242
- Ercisli S (2004) A short review of the fruit germplasm resources of Turkey. Genet Resour Crop Evol 51:419–435
- 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
- Fernández i Martí A, Alonso JM, Espiau MT, Rubio-Cabetas MJ, Socias i Company R (2009) Genetic diversity in Spanish and foreign almond germplasm assessed by molecular characterization with simple sequence repeats. J Am Soc Hortic Sci 134:535–542
- Fernández i Martí A, Forcada CF, Kamali K, Rubio-Cabetas MJ, Wirthensohn M (2015) Molecular analyses of evolution and population structure in a worldwide almond [*Prunus dulcis* (Mill.) DA Webb syn. *P. amygdalus* Batsch] pool assessed by microsatellite markers. Genet Resour Crop Evol 62:205–219
- Godini A (1979) Ipotesi sulla comparsa dell’autocompatibilità nel mandorlo. Riv Sci Tecn Agr 19:3–10
- Godini A (2000) About the possible relationships between *Amygdalus webbi* Spach and *Amygdalus communis* L. Nucis Newslett 9:17–19
- Gradziel TM (2011) Origin and dissemination of almond. Hortic Rev 38: 23–82
- Gupta PK, Balyan IS, Sharma PC, Ramesh B (1996) Microsatellites in plants: a new class of molecular markers. Curr Sci India 70:45–54
- Hadziabdic D, Wang X, Wadl PA, Rinehart TA, Ownley BH, Trigiano RN (2012) Genetic diversity of flowering dogwood in the Great Smoky Mountains National Park. Tree Genet Genomes 8:855–871
- Halász J, Pedryc A, Hegedüs A (2007) Origin and dissemination of the pollen-part mutated SC haplotype which confers self-compatibility in apricot (*Prunus armeniaca*). New Phytol 176:792–803
- Halász J, Fodor A, Pedryc A, Hegedüs A (2010) S-genotyping of Eastern European almond cultivars: identification and characterization of new (S36–S39) self-incompatibility ribonuclease alleles. Plant Breed 129:227–232
- Halász J, Hegedüs A, Szikriszt B, Ercişli S, Orhan E, Ünlü HM (2013) The S-genotyping of wild-grown apricots reveals only self-incompatible accessions in the Erzincan region of Turkey. Turk J Biol 37:733–740
- Hammer Ø, Harper DAT, Ryan PD (2001) PAST: paleontological statistics software package for education and data analysis. Palaeontol Electron 4:1–9
- Hardy OJ, Charbonnel N, Freville H, Heuertz M (2003) Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. Genetics 163:1467–1482
- Howad W, Yamamoto T, Dirlwanger E, Testolin R, Cosson P, Cipriani G, Monforte AJ, Georgi L, Abbott AG, Arús P (2005) Mapping with a few plants: using selective mapping for microsatellite saturation of the *Prunus* reference map. Genetics 171:1305–1309
- Jordano P, Godoy JA (2000) RAPD variation and population genetic structure in *Prunus mahaleb* (Rosaceae), an animal-dispersed tree. Mol Ecol 9:1293–1305
- Kester DE, Gradziel TM, Grasselly C (1991) Almonds (*Prunus*). Genet Res Temper Fruit Nut Crops 290:701–760
- Lansari A, Iezzoni AF, Kester DE (1994) Morphological variation within collections of Moroccan almond clones and Mediterranean and North American cultivars. Euphytica 78:27–41
- Lee S, Wen J (2001) A phylogenetic analysis of *Prunus* and the Amygdaloideae (Rosaceae) using ITS sequences of nuclear ribosomal DNA. Am J Bot 88:150–160
- Marchese A, Bošković RI, Martínez-García PJ, Tobutt KR (2008) The origin of the self-compatible almond ‘Supernova’. Plant Breed 127: 105–107
- Martínez-García PJ, Dicenta F, Ortega E (2012) Anomalous embryo sac development and fruit abortion caused by inbreeding depression in almond (*Prunus dulcis*). Sci Hortic-Amsterdam 133:23–30
- Martínez-Gómez P, Sozzi GO, Sánchez-Pérez R, Rubio M, Gradziel TM (2003) New approaches to *Prunus* tree crop breeding. J Food Agric Environ 1:52–63
- Martínez-Gómez P, Sánchez-Pérez R, Dicenta F, Howad W, Arús P, Gradziel T (2007) Almond. In: Kole C (ed) Genome mapping and molecular breeding in plants. Springer-Verlag, Berlin Heidelberg, pp 229–242
- Martinoli D, Jacomet S (2004) Identifying endocarp remains and exploring their use at Epipalaeolithic Öküzini in southwest Anatolia, Turkey. Veg Hist Archaeobotany 13:45–54
- Médail F, Diadema K (2009) Glacial refugia influence plant diversity patterns in the Mediterranean Basin. J Biogeogr 36:1333–1345
- Meikle RD (1966) Contributions to the flora of Iraq. Kew Bull 19:229
- Messina R, Lain O, Marrazzo MT, Cipriani G, Testolin R (2004) New set of microsatellite loci isolated in apricot. Mol Ecol Notes 4:432–434
- Mihretie Z, Schueler S, Konrad H, Bekele E, Geburek T (2015) Patterns of genetic diversity of *Prunus africana* in Ethiopia: hot spot but not point of origin for range-wide diversity. Tree Genet Genomes 11:118

- Mnejja M, Garcia-Mas J, Howad W, Badenes ML, Arús P (2004) Simple-sequence repeat (SSR) markers of Japanese plum (*Prunus salicina* Lindl.) are highly polymorphic and transferable to peach and almond. *Mol Ecol Notes* 4:163–166
- Mnejja M, Garcia-Mas J, Howad W, Arús P (2005) Development and transportability across *Prunus* species of 42 polymorphic almond microsatellites. *Mol Ecol Notes* 5:531–535
- Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc Natl Acad Sci U S A* 76: 5269–5273
- Page RDM (1996) Tree View: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* 12:357–358
- Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28:2537–2539
- Pickersgill B (1998) Crop introductions and the development of secondary areas of diversity. In: Prendergast HD (ed) *Plants for food and medicine*. Royal Botanic Gardens, Kew, pp 93–105
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Rahemi A, Fatahi R, Ebadi A, Taghavi T, Hassani D, Gradziel T, Folta K, Chaparro J (2012) Genetic diversity of some wild almonds and related *Prunus* species revealed by SSR and EST-SSR molecular markers. *Plant Syst Evol* 298:173–192
- Rapaics R (1940) *A magyar gyümölcs*. Királyi Magyar Természettudományi Társulat, Budapest
- Slatkin M (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics* 139:457–462
- Slatkin M, Barton NH (1989) A comparison of three indirect methods for estimating average levels of gene flow. *Evolution* 43:1349–1368
- Socias i Company R (1990) Breeding self-compatible almonds. *Plant Breed Rev* 8:313–338
- Socias i Company R (2002) The relationship of *Prunus webbii* and almond revisited. *FAO-CHIEAM Nucis Newsletter* 11:17–19
- Socias i Company R (2004) The contribution of *Prunus webbii* to almond evolution. *Plant Genet Resour News* 14:9–13
- Sorkheh K, Dehkordi MK, Ercisli S, Hegedűs A, Halász J (2017) Comparison of traditional and new generation DNA markers declares high genetic diversity and differentiated population structure of wild almond species. *Sci Rep* 7:5966
- Szikriszt B, Hegedűs A, Halász J (2011) Review of genetic diversity studies in almond (*Prunus dulcis*). *Acta Agron Hung* 59:379–395
- Tahan O, Geng Y, Zeng L, Dong S, Chen F, Chen J, Song Z, Zhong Y (2009) Assessment of genetic diversity and population structure of Chinese wild almond, *Amygdalus nana*, using EST-and genomic SSRs. *Biochem Syst Ecol* 37:146–153
- Tautz J, Zhang S, Spaethe J, Brockmann A, Si A, Srinivasan M (2004) Honeybee odometry: performance in varying natural terrain. *PLoS Biol* 2:e211
- Testolin R, Messina R, Lain O, Marrazzo MT, Huang WG, Cipriani G (2004) Microsatellites isolated in almond from an AC-repeat enriched library. *Mol Ecol Notes* 4:459–461
- Veron V, Caron H, Degen B (2005) Gene flow and mating system of the tropical tree *Sextonia rubra*. *Silvae Genet* 54:275–280
- Wang Z, Kang M, Liu H, Gao J, Zhang Z, Li Y, Wu R, Pang X (2014) High-level genetic diversity and complex population structure of Siberian apricot (*Prunus sibirica* L.) in China as revealed by nuclear SSR markers. *PloS One* 9:e87381
- Wood MN (1925) Almond varieties in the United States. *US Dep Agric Tech Bul* 1282:1–142
- Xie H, Sui Y, Chang FQ, Xu Y, Ma RC (2006) SSR allelic variation in almond (*Prunus dulcis* Mill.). *Theor Appl Genet* 112:366–372
- Xu Y, Ma RC, Xie H, Liu JT, Cao MQ (2004) Development of SSR markers for the phylogenetic analysis of almond trees from China and the Mediterranean region. *Genome* 47:1091–1104
- Yeh FC, Yang RC, Boyle TB, Ye ZH, Mao JX (1997) POPGENE, the user-friendly shareware for population genetic analysis. Molecular biology and biotechnology centre, University of Alberta, Canada, 10:295–301
- Yezhov VN, Smykov AV, Smykov VK, Khokhlov SY, Zurov DE, Mehlenbacher SA, Molnar TJ, Goffreda JC, Funk CR (2005) Genetic resources of temperate and subtropical fruit and nut species at the Nikita Botanical Gardens. *HortSci* 40:5–9
- Zeinalabedini M, Khayam-Nekoui M, Grigorian V, Gradziel TM, Martínez-Gómez P (2010) The origin and dissemination of the cultivated almond as determined by nuclear and chloroplast SSR marker analysis. *Sci Hortic-Amsterdam* 125:593–601
- Zeinalabedini M, Sohrabi S, Nikoumanesh K, Imani A, Mardi M (2012) Phenotypic and molecular variability and genetic structure of Iranian almond cultivars. *Plant Syst Evol* 298:1917–1929