Tree Genetics & Genomes (2016) 12: 114
DOI 10.1007/s11295-016-1075-y

ORIGINAL ARTICLE

Genetic diversity and genetic structure of Persian walnut (*Juglans regia)* accessions from 14 European, African, and Asian countries using SSR markers

Aziz Ebrahimi[1] · **Abdolkarim Zarei**[2] · **Shaneka Lawson**[3] · **Keith E. Woeste**[3] · **M. J. M. Smulders**[4]

Received: 5 August 2016 / Revised: 4 October 2016 / Accepted: 27 October 2016 / Published online: 12 November 2016

Abstract Persian walnut (*Juglans regia* L.) is the world's most widely grown nut crop, but large-scale assessments and comparisons of the genetic diversity of the crop are notably lacking. To guide the conservation and utilization of Persian walnut genetic resources, genotypes (n = 189) from 25 different regions in 14 countries on three continents were sampled to investigate their genetic relationships and diversity using ten microsatellite (SSR) loci. The SSRs amplified from 3 to 25 alleles per locus, with a mean value of 11.5 alleles per locus. The mean values of observed and expected heterozygosity were 0.62 and 0.73, respectively. Based on Nei's genetic identity, accessions from Bratislava (Slovakia) and Antalya (Turkey) showed the lowest similarity (0.36), while accessions from Algeria and Tunisia as well as accessions from Debrecen (Hungary) and Trnava (Slovakia) had the highest similarity (0.97). Two populations from Iran (Alborz and Ardabil) had the highest number of private alleles (7 and 5), but they were quite different as they also had the lowest genetic identity when compared to the remaining populations as well as to each other. Although overall differentiation among regions was relatively low (F_{ST} = 0.07), cluster analysis grouped accessions generally but not completely according to geography. STRUCTURE software confirmed these results and divided the accessions into two main groups, separating accessions collected from Europe and North Africa from those from Greece and the Near East. Results indicate the presence of a likely center of diversity for Persian walnut in Eastern and Southeastern Europe. They also provide information that can be used to devise conservation actions. Notably, the genetic diversity of threatened populations from two regions in Iran should be conserved.

Keywords Microsatellite · Persian walnut · Middle East · Private allele

Aziz Ebrahimi and Abdolkarim Zarei contributed equally.

Communicated by A.M. Dandekar

Electronic supplementary material The online version of this article (doi:10.1007/s11295-016-1075-y) contains supplementary material, which is available to authorized users.

✉ Abdolkarim Zarei
zarei14@gmail.com

[1] Department of Forestry and Natural Resources, Purdue University, 715 State Street, West Lafayette, IN 47907, USA

[2] Department of Biotechnology, College of Agriculture, Jahrom University, PO BOX 74135-111, Jahrom, Iran

[3] USDA Forest Service, Hardwood Tree Improvement and Regeneration Center, Purdue University, 715 State Street, West Lafayette, IN 47907, USA

[4] Wageningen University & Research, Plant Breeding, P.O. Box 386, NL-6700 AJ Wageningen, The Netherlands

Introduction

Persian walnut (*Juglans regia* L.) is a member of the Juglandaceae valued for its edible nut and its timber. The species has a long history of cultivation. It is believed that the use of the species originated in the mountains of Eastern and Central Asia, but an exact location is not known. The seeds were subsequently traded along the Silk Road and other trade routes by ancient tribes and conquerors (McGranahan et al. 1998; Bayazit et al. 2006; Pollegioni et al. 2015), as was also the case for fruit species such as apple (Cornille et al. 2012). Today walnut is grown in over 60 countries throughout the temperate

regions of the globe and is harvested from both cultivated orchards and wild populations (Avanzato et al., 2014). While trees in orchards are clearly cultivated, it is generally unknown to what extent the trees in forests have been planted and/or selected by humans.

Walnuts are monoecious and heterodichogamous, favoring outcrossing over selfing. Watanabe et al. (2016) noted that heterodichogamy is an adaption designed to promote outbreeding and was directly related to future seed set. Until recently, Persian walnut was nearly always propagated by seed (Ebrahimi et al. 2011), which allowed each geographic region to maintain a diverse population. Comparison of genetic diversity and the similarities of geographically separated populations could shed light on the dispersal and distribution of this species over time (Pollegioni et al. 2014).

SSR loci for *Juglans* and specific primers to amplify them have been characterized and used to discriminate among *Juglans* genotypes (Woeste et al. 2002; Dangl et al. 2005; Topcu et al. 2015), to describe hybridization among congeners, and for analysis of genetic diversity and population structure of cultivated *Juglans* populations (Foroni et al. 2007; Aradhya et al. 2010; Ebrahimi et al. 2011; Pollegioni et al. 2011, 2014; Pop et al. 2013; Dogan et al. 2014; Vahdati et al. 2015). Most of these studies focused on genotypes collected for studies of local or regional genetic variability or from breeding populations; studies of diversity and relatedness across continental scales are rare. For example, Foroni et al. (2007) and Vahdati et al. (2015) studied *J. regia* at the provincial scale, Ebrahimi et al. (2011) and Dogan et al. (2014) worked on walnut genotypes at national scale, Aradhya et al. (2010) and Pop et al. (2013) worked on accessions collected in a germplasm collection, while Pollegioni et al. (2014) expanded the area of sampling across numerous Asian countries. Therefore, to our knowledge, no published information exists about the relationships between Persian walnut accessions at a continental scale.

The objective of our study was to examine and compare the genetic diversity and structure of walnut genotypes collected from Iran, which may be where walnut was first cultivated, with samples from the Middle East, the Near East, Europe, and the North African side of the Mediterranean basin. Continental scale comparisons of *J. regia* genetic diversity are needed because germplasm exchange and germplasm conservation are now international. By deepening our understanding of the genetic diversity and structure of walnut on continental scales, comparing regional germplasm pools, and identifying regions of unusual diversity and similarity, our results will provide data to support additional germplasm collection and the designation of areas of special concern for conservation. The results are also expected to benefit breeders who wish to know about sources of genetic variability that differ from their current germplasm.

Materials and methods

Plant materials

Our study incorporated 189 adult walnut trees grown locally in each of 14 countries and 25 regions. We sampled four locations each from Hungary, Slovakia Turkey, and Africa (Algeria, Libya, Morocco, and Tunisia), three from Iran and Western Europe (the Netherlands, Italy, and Spain), two from Central Europe (Germany and Greece), and a single location from Iraq (Table 1). Fifty-seven *J. regia* accessions of Asian origin were compared with 107 adult walnut genotypes that originated from Europe/Africa. Mature, healthy leaves were selected from trees at least 1 km apart, to avoid consanguinity, and then dried with silica gel. The location of each sample was recorded along with detailed population information (Table 1, Fig. S3). Accessions sampled from Asia boasted trunk diameters greater than 50 cm and were estimated to be at least 80 years old. We attempted to collect a minimum of ten samples (with these characteristics) per site; however, remote location and low density populations hindered success at some sites. Therefore, the number of samples collected per region was not equal. To obtain genetic diversity information and for comparison with Asian and European samples, 25 elite genotypes, previously collected at the plant research station repository (Casablanca, Morocco) were also included in this study (Table 1).

Genomic DNA extraction, PCR amplification, and PCR product analysis

A CTAB buffer (Doyle and Doyle 1987) was used to extract genomic DNA from leaves dried with silica gel. PCR was performed using ten pairs of fluorescently labeled SSR primers (WGA1, WGA9, WGA27, WGA32, WGA69, WGA89, WGA118, WGA202, WGA276, and WGA321) that amplified loci shown to be highly polymorphic (Woeste et al. 2002; Dangl et al. 2005). PCR conditions followed those described in Ebrahimi et al. (2011). The PCR products were separated using an ABI 3130xl genetic analyzer (Applied Biosystems Inc., Foster City, CA, USA) with a 36-cm capillary array and a POP7 matrix (ABI). One standard sample (Chandler cultivar) was included in each plate to assist scoring across arrays. Allele sizes were scored using Genemapper 3.7 NT software (Applied Biosystems).

Data analysis

The number of alleles observed (A), effective number of alleles (A_E), observed heterozygosity (H_O), expected heterozygosity (H_E), Shannon's information index (I), Nei's genetic distance, and Nei's genetic similarity (Nei, 1972) were calculated using POPGENE version 1.31 software (Yeh et al.

Table 1 Walnut accessions and source information

ID	City	Country	Latitude	Longitude	Elevation (m)	Type of materials	Provider
1-10	Constantine	Algeria	36.3602° N	6.6424° E	694	Selected seedlings	Plant Research Station-Morocco
11-17	Cologne	Germany	50.58° N	6.58° E	91	Selected seedlings	Landscape and Park
18-27	Budapest	Hungary	47.4979° N	19.0402° E	96	Selected seedlings	Institute of Fruit research at Budapest, Hungary
28-39	Debrecen	Hungary	47.5316° N	21.6273° E	121	Selected seedlings	Institute of Fruit research at Budapest, Hungary
40-44	Miskolc	Hungary	48.0964° N	20.7624° E	141	Selected seedlings	Institute of Fruit research at Budapest, Hungary
45-47	Erd	Hungary	47.3920° N	18.9045° E	152	Selected seedlings	Institute of Fruit research at Budapest, Hungary
48-52	Ardabil	Iran	38.2537° N	48.3000° E	1339	Accession	Agriculture Station Research-Iran
53-57	Kurdistan	Iran	35.9554° N	47.1362° E	1538	Accession	Agriculture Station Research-Iran
58-67	Alborz	Iran	35.9960° N	50.9289° E	1341	Accession	Agriculture Station Research-Iran
68-77	Erbil	Iraq	36.2063° N	44.0089° E	420	Accession	Agriculture Station Research-Iran
78-87	Tripoli	Libya	32.8872° N	13.1913° E	81	Selected seedlings	Plant Research Station-Morocco
88-101	Casablanca	Morocco	33.5731° N	7.5898° W	57	Selected seedlings	Plant Research Station-Morocco
102-105	Wageningen	Netherlands	51.9692° N	5.6654° E	9	Selected seedlings	Landscape and botanical park
106-109	Bratislava	Slovakia	48.1486° N	17.1077° E	134	Selected seedlings	Institute of Fruit research at Budapest, Hungary
110-115	Nitra	Slovakia	48.3061° N	18.0764° E	149	Selected seedlings	Local orchard and Landscape
116-120	Levice	Slovakia	48.2174° N	18.5984° E	190	Selected seedlings	Institute of Fruit research at Budapest, Hungary
121-133	Trnava	Slovakia	48.3709° N	17.5833° E	146	Selected seedlings	Institute of Fruit research at Budapest, Hungary
134-143	Sousse	Tunisia	35.8256° N	10.6084° E	4	Selected seedlings	Plant Research Station-Morocco
144-148	Antalya	Turkey	36.8969° N	30.7133° E	64	Accession	Institute of Horticulture Research, Izmir, Turkey
149-154	Diyarbakir	Turkey	37.53° N	40.12° E	674	Accession	Institute of Horticulture Research, Izmir, Turkey
155-158	Sirrt	Turkey	37.9274° N	41.9420° E	896	Accession	Institute of Horticulture Research, Izmir, Turkey
159-161	Izmir	Turkey	38.4237° N	27.1428° E	29	Accession	Institute of Horticulture Research, Izmir, Turkey
162-172	Sykies	Greece	40.64944° N	22.95083° E	164	Accession	Institute of Horticulture Research, Izmir, Turkey
173-181	Bologna	Italy	44.4949° N	11.3426° E	54	Selected seedlings	Plant Research Station-Morocco
182-189	Madrid	Spain	40.4165° N	3.70256° W	667	Selected seedlings	Plant Research Station-Morocco

1997). Polymorphic Information Content (PIC) and F_{Null} estimation were computed by CERVUS version 2.0 (Marshall et al. 1998) software. Number of alleles within a given population or allelic richness were used as measures of a population's potential for breeding. Number of alleles per locus was highly dependent on population size, hence allelic richness was computed using rarefaction with HP-Rare software (Kalinowski 2005). Subsequently, the Inverse Distance Weighted (IDW) interpolation function implemented in the GIS software ArcGIS 9.3 (ESRI, Redlands, CA, USA) was used to infer allelic richness outside the sampled sites and regions.

Cluster analysis was carried out using Nei's genetic distance, and a dendrogram was constructed using Mega software version 6 (Tamura et al., 2012) with bootstrap values. Principal coordinate analysis (PCoA) was conducted using GenAlEx 6.5 (Peakall and Smouse 2012). Analysis of molecular variance (AMOVA) of the germplasm set was performed based on Nei's (Nei 1972) distance matrix using the program GenAlEx 6.5 (Peakall and Smouse 2012). The number of private alleles and rare alleles were estimated using GenAlEx 6.5. Genetic structure analyses were performed with the program STRUCTURE 2.3.4 with 10 runs and 100,000 Markov Chain Monte Carlo (MCMC) repetitions after a burn-in period of 100,000 interactions for each

group number K (where K has values ranging from 1 to 15). The optimum value of K was obtained by calculating the ΔK value to determine the most likely number of groups (Evanno et al. 2005). The results from STRUCTURE were processed with the software STRUCTURE HARVESTER v.0.6.1 (Earl and VonHoldt 2012) to obtain the most likely K value. The K value estimated using STRUCTURE HARVESTER also showed a clear peak at the optimal K value (Fig. S4).

Results

Genetic diversity and population variation

The ten studied SSR loci were polymorphic (Table 2) and produced 115 alleles in 189 accessions. The average number of alleles across marker loci was 11.5 but varied greatly across the loci (SE = 6.5). WGA32, with 25 alleles, was the most varied, while WGA27, with three alleles, was the least variable. The number of effective alleles ranged from 1.9 (WGA27) to 7.6 (WGA276), with a mean value of 4.2 (SE = 1.6). The average PIC was 0.69 although WGA202 displayed the highest value (0.81). Heterozygosity also varied across the ten SSR loci and observed heterozygosities (Ho; 0.62 ± 0.12) were less than expected (He; 0.73 ± 0.09) for all loci. WGA27 revealed the fewest alleles and the lowest H_O (0.43). WGA69 had a significantly lower H_O value (0.56) than expected (0.81), indicating the likely presence of null allele(s) for this locus (F_{Null} = 0.18).

Diversity indices varied across regions (Table 3). The Shannon's information index (I), indicative of genetic diversity, had an average value of 1.59 (SE = 0.47). Accessions collected from Debrecen (Hungary) had the highest number of alleles and number of effective alleles (5.5 and 4.3, respectively), while accessions from Germany had the lowest values for these two indices (2.7 and 1.8, respectively). This pattern was also revealed by Shannon's index (I), which was 1.44 for Debrecen (Hungary) and 0.73 for Germany.

Twenty-eight private alleles were revealed in 13 accession groups. The greatest numbers of private alleles were observed in two populations from Iran (seven alleles, Alborz; five alleles, Ardabil), followed by a single population from Sirrt, Turkey (three alleles). Allelic richness was also interpolated using the IDW interpolation function in ArcGIS. The resulting map (Fig. 1) showed alternating regions of higher and lower allelic richness.

Data generated from Nei's genetic identity tests indicated that accessions from Algeria and Tunisia and those from Debrecen (Hungary) and Trnava (Slovakia) were the most similar (0.97). The most dissimilar accessions were those from Bratislava (Slovakia) and Antalya (Turkey) with value of 0.36 (see Table S1 in Appendix S1). The two populations from Iran with the most private alleles, Alborz and Ardabil, were also highly dissimilar (0.42).

To assess the overall distribution of genetic diversity within and between groups, an AMOVA was performed. AMOVA analysis indicated that variation among groups (Fst) was 7%, while variation among individuals within groups was 8%, and the remaining 85% of variation was within individuals (Table 4). In contrast to our expectation that all accessions would be from unique seedlings, we found two examples of identical genotypes. Specifically, the profile of the accession 73 from Erbil (Iraq) was identical to accession 155 from Sirrt (Turkey), and accession 76 from Erbil (Iraq) was identical to accession 158 from Sirrt (Turkey).

Population structure

Analyses with STRUCTURE software (at K = 2) identified two groups that roughly corresponded to a group from Asia (Iran and Turkey, plus Greece), and a group of European and North African genotypes (Fig. 2). About 68% (128 genotypes) of the studied walnut accessions shared >80% membership with one of two main clusters and were classified as members of that cluster, while the remaining genotypes (61 accessions) were admixed. Accessions collected from Eastern Europe (Hungary, Slovakia) showed a high degree of admixture (Fig. 4). Of the most admixed genotypes at K = 2 ($30 > Q < 70$), about half were from Slovakia or Hungary. The same trend was observed at K = 8 (see below) (Fig. 3), where 37 of the 49 genotypes that showed pronounced admixture of two populations (Fig. 3, populations four and five, yellow and pink) were from Hungary or Slovakia. Many admixed samples in Fig. 3 were complex admixtures that did not sort strongly into a particular population.

Further analyses using STRUCTURE HARVESTER showed a second clear peak at K = 8 (ΔK = 30.79). Thus, the statistical approach of Evanno et al. (2005) may be problematic in our study, as it most often uses K = 2 as the most likely number of clusters (Pollegioni et al. 2014). Therefore, these data were also analyzed based on K = 8 (Fig. 3). Cluster results showed that the first population (red in Fig. 3) included 12% of accessions and was primarily composed of samples from the European countries Germany, Hungary, the Netherlands, and Slovakia, with a single accession from Algeria. The second population (green) was the smallest one and included only 15 accessions, mostly from Ardabil, Iran, but also a few from Iraq (2), Turkey (1), Algeria (1), and Morocco (1). The third population (dark blue) included accessions from Greece (9), Iraq (6), and Turkey (4). The fourth and fifth populations showed almost complete admixture (yellow and pink), as all accessions in these groups (mostly from Hungary and Slovakia with a few from Greece and Spain) had roughly equal Q values for both groups, and none had a Q value >80%. This complete admixture was also present

Table 2 Information about the ten SSR markers (Woeste et al. 2002) and diversity indices from evaluation of 189 *J. regia* accessions from different

Locus	Number	Type of repeat	SR (bp)	*A*	A_E	PIC	HW	(*I*)	F_{Null}	H_O	H_E	A_{RA}
WGA1	189	(GA)5	176–192	10	3.6	0.68	*	1.49	0.08	0.61	0.73	6
WGA9	189	(GA)16	231–249	7	3.3	0.64	NS	1.31	0.07	0.59	0.70	3
WGA27	189	(GA)30	205–209	3	1.9	0.37	NS	0.70	0.06	0.43	0.48	1
WGA32	189	(CT)19	157–229	25	4.2	0.74	NS	1.97	0.09	0.63	0.76	21
WGA69	189	(GT)13	157–193	10	5.1	0.78	***	1.78	0.18	0.56	0.81	5
WGA89	189	(GA)4	211–226	8	3.2	0.62	NS	1.25	0.07	0.60	0.69	5
WGA118	189	(GA)18	183–293	9	3.7	0.69	*	1.48	0.06	0.65	0.73	4
WGA202	189	(GA)20	198–295	17	5.9	0.81	*	2.08	0.08	0.71	0.83	11
WGA276	189	(GA)14	159–195	18	7.6	0.86	NS	2.34	0.06	0.77	0.87	13
WGA321	189	(GA)14	224–247	8	3.7	0.69	NS	1.47	0.04	0.68	0.73	4
Mean			-	11.5	4.2	0.69		1.59	0.08	0.62	0.73	7.3
SE				6.5	1.6	0.13		0.47	0.04	0.09	0.11	6.26

N accession size, *SR* allele size range (bp), *A* no. of obtained alleles, A_E no. of effective alleles, *PIC* Polymorphic Information Content, *HW* exact test of Hardy-Weinberg equilibrium, *I* Shannon's Information Index, F_{Null} null allele frequency estimated, H_O observed heterozygosity, H_E expected heterozygosity, Nei Nei's (1972) expected heterozygosity, A_{RA} number of rare alleles, *SE* standard error, *NS* non significance

*significance at $P = 0.01$

Table 3 Genetic diversity of walnut accessions from different regions using 10 SSRs

ID	Country, regions	Number	*A*	A_E	(*I*)	H_O	H_E	Nei	N_{PA}
1	Algeria	20	4.2 ± 1.3	3.06 ± 0.84	1.19 ± 0.27	0.62 ± 0.16	0.69 ± 0.09	0.66 ± 0.09	2
2	Germany	14	2.7 ± 0.7	1.83 ± 0.30	0.73 ± 0.18	0.44 ± 0.18	0.47 ± 0.11	0.44 ± 0.10	–
3	Hungary, Budapest	10	4.4 ± 1.8	3.48 ± 1.75	1.26 ± 0.41	0.57 ± 0.23	0.70 ± 0.12	0.66 ± 0.11	1
4	Hungary, Debrecen	12	5.5 ± 3.1	4.33 ± 2.25	1.44 ± 0.52	0.73 ± 0.16	0.75 ± 0.14	0.71 ± 0.13	1
5	Hungary, Miskolc	5	3.9 ± 1.1	3.17 ± 1.11	1.19 ± 0.37	0.66 ± 0.21	0.71 ± 0.19	0.63 ± 0.17	–
6	Hungary, Erd	3	3.3 ± 1.2	2.93 ± 1.19	1.08 ± 0.33	0.67 ± 0.27	0.75 ± 0.13	0.62 ± 0.11	–
7	Iran, Ardabil	5	3.7 ± 1.1	2.97 ± 1.04	1.11 ± 0.40	0.56 ± 0.25	0.67 ± 0.22	0.60 ± 0.20	5
8	Iran, Kurdistan	5	3.2 ± 1.0	2.55 ± 0.74	0.97 ± 0.39	0.62 ± 0.26	0.62 ± 0.23	0.56 ± 0.21	1
9	Iran, Alborz	10	4.0 ± 1.1	3.05 ± 0.83	1.19 ± 0.27	0.69 ± 0.20	0.68 ± 0.10	0.65 ± 0.10	7
10	Iraq	20	4.7 ± 1.8	3.59 ± 1.00	1.33 ± 0.35	0.67 ± 0.19	0.73 ± 0.11	0.70 ± 0.11	1
11	Libya	20	3.9 ± 1.0	2.89 ± 0.67	1.14 ± 0.23	0.61 ± 0.14	0.67 ± 0.09	0.64 ± 0.08	1
12	Morocco	26	4.4 ± 1.4	3.21 ± 0.87	1.24 ± 0.29	0.67 ± 0.16	0.69 ± 0.09	0.67 ± 0.09	2
13	Netherlands	10	3.3 ± 1.4	2.66 ± 0.84	1.02 ± 0.32	0.60 ± 0.25	0.66 ± 0.11	0.60 ± 0.10	–
14	Slovakia, Bratislava	4	3.6 ± 1.4	2.68 ± 1.09	1.03 ± 0.43	0.60 ± 0.21	0.64 ± 0.21	0.56 ± 0.19	–
15	Slovakia, Nitra	6	4.0 ± 1.5	3.29 ± 1.13	1.21 ± 0.38	0.72 ± 0.18	0.72 ± 0.15	0.66 ± 0.13	–
16	Slovakia, Levice	5	3.2 ± 1.1	2.55 ± 0.94	0.98 ± 0.33	0.56 ± 0.31	0.63 ± 0.13	0.57 ± 0.12	–
17	Slovakia, Trnava	13	5.6 ± 2.4	3.67 ± 1.23	1.40 ± 0.39	0.62 ± 0.17	0.72 ± 0.11	0.69 ± 0.11	2
18	Tunisia	20	3.9 ± 1.2	2.79 ± 1.07	1.07 ± 0.36	0.60 ± 0.21	0.62 ± 0.18	0.59 ± 0.17	–
19	Turkey, Antalya	5	3.7 ± 1.6	2.70 ± 1.45	1.01 ± 0.48	0.52 ± 0.30	0.60 ± 0.22	0.54 ± 0.20	1
20	Turkey, Diyarbakir	6	3.9 ± 1.6	2.97 ± 1.34	1.09 ± 0.50	0.43 ± 0.21	0.63 ± 0.24	0.58 ± 0.22	–
21	Turkey, Sirrt	4	3.6 ± 1.4	3.14 ± 1.20	1.12 ± 0.47	0.47 ± 0.36	0.70 ± 0.26	0.62 ± 0.23	3
22	Turkey, Izmir	3	3.4 ± 1.1	2.97 ± 1.05	1.09 ± 0.37	0.53 ± 0.32	0.74 ± 0.18	0.62 ± 0.16	–
23	Greece	22	4.9 ± 2.1	3.85 ± 1.49	1.37 ± 0.45	0.71 ± 0.21	0.72 ± 0.17	0.70 ± 0.16	1
24	Italy	18	3.3 ± 0.7	2.35 ± 0.35	0.97 ± 0.15	0.64 ± 0.20	0.60 ± 0.07	0.57 ± 0.07	–
25	Spain	16	3.9 ± 1.1	2.75 ± 0.66	1.11 ± 0.25	0.68 ± 0.18	0.66 ± 0.10	0.62 ± 0.10	–
	Mean ± SD	11.3 ± 7.1	4.7 ± 1.4	3.24 ± 0.76	1.22 ± 0.23	0.62 ± 0.07	0.67 ± 0.07	0.65 ± 0.8	1.12 ± 1.72

N no. of accessions from each location, N_A no. of observed alleles, N_E effective number of alleles, *I* Shannon's Information Index, H_O observed heterozygosity, H_E expected heterozygosity, *Nei* Nei's (1972) expected heterozygosity, N_{PA} number of private alleles, *SE* standard error.

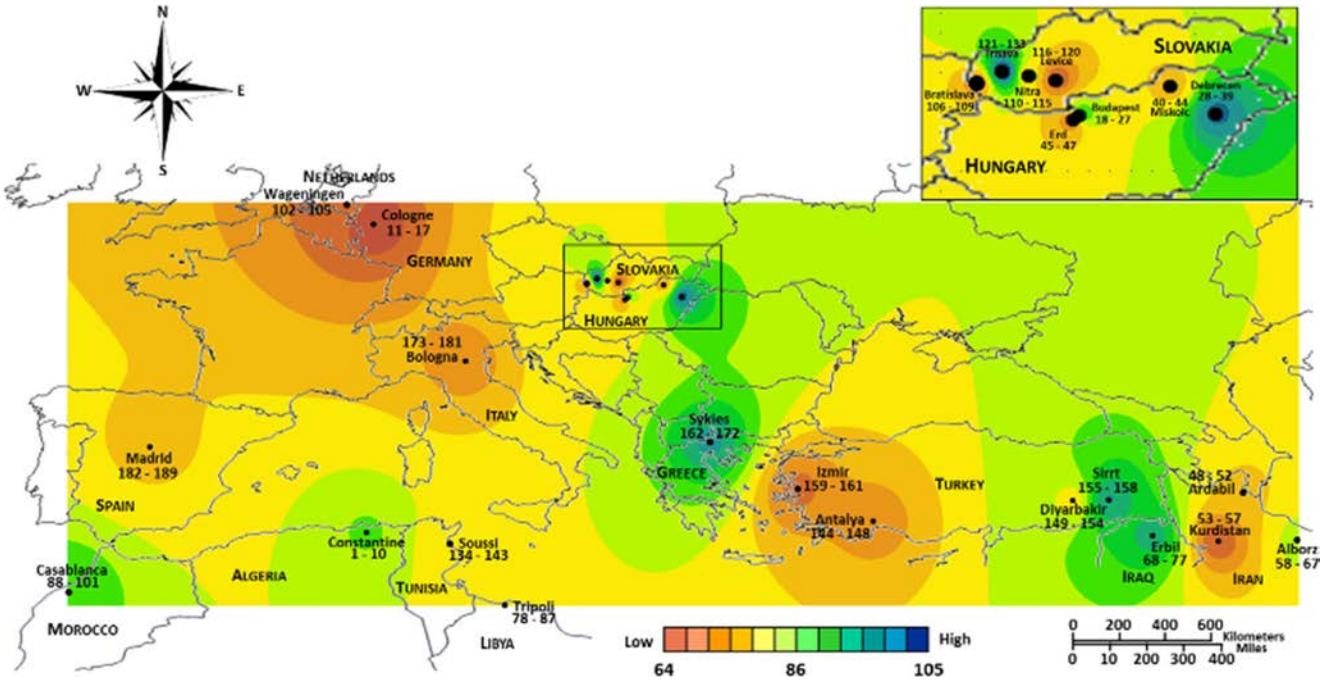

Fig. 1 Geographical variation of allelic richness in *J. regia* across 14 sampled countries

when the analysis was run for $K = 7$ (not shown). The sixth population included accessions collected from Alborz, Iran (10) as well as both Antalya (5) and Diyarbakir (5) in Turkey. In contrast with the two previous subgroups, most of the members of this subgroup had Q values >80%. The seventh population consisted principally of accessions from North African countries and also showed a high level of admixture. The eighth and largest population, nearly 18% of the accessions, included most of the accessions from Italy, and some admixed genotypes from Spain (5), Turkey (2), Germany (2), Slovakia (4), the Netherlands (2), Tunisia (1), Morocco (4), Algeria (3), Libya (3), Iraq (1), and Iran (3).

Neighbor-joining analysis

Neighbor-joining (NJ) based on a matrix of Nei's genetic distance among the 25 sampled locations divided the accessions into five main branches, and within these branches, additional sub-clusters of branches (Fig. 4), although bootstrap values were relatively low. Four of the five main branches appeared to correspond to seven collection sites in Iraq, Iran, Turkey, and Greece (Asia and Near East). The fifth main branch (top~two-third of Fig. 4) was primarily European and North African. The fifth main branch of European and North African sources can be seen as divided into four sub-clusters: (1) a (mostly) Mediterranean/Italian/North African sub-cluster (names of sources are underlined, top of Fig. 4); (2) a sub-cluster of mostly Hungarian/Slovakian sources; (3) a mixed sub-cluster of Northern European, Eastern European, and Iranian sources (names are double-underlined in Fig. 4); and (4) a final sub-cluster containing only the four Bratislava (Slovakia) samples. As an example of the differences among the clustering methods we used, the four samples of sub-cluster 4 (Bratislava, Fig. 4) were separated into two different populations by the $K = 8$ STRUCTURE analysis (Fig. 3), and three different sub-branches when neighbor-joining was performed based on individual genotype (Fig. S5), rather than by the 25 sources (Fig. 4). There were similarities between the $K = 8$ model (Fig. 3) and the neighbor-joining model (Fig. 4) of genotype partitioning; the Ardabil, Iran location (clade 1, bottom of Fig. 4), corresponded to the second population (green) of STRUCTURE (Fig. 3), the second and third

Table 4 Analysis of molecular variance (AMOVA) using 189 accessions from 25 regions using ten SSR markers

Source of variation	Degrees of freedom	Sum of squares	Mean square	Percentage variation (%)	*P* value
Among groups	24	188.976	7.874	7	<0.001
Among individuals within groups	164	604.805	3.688	8	<0.001
Within individuals	189	588.000	3.111	85	<0.001
Total	377	1381.780		100	

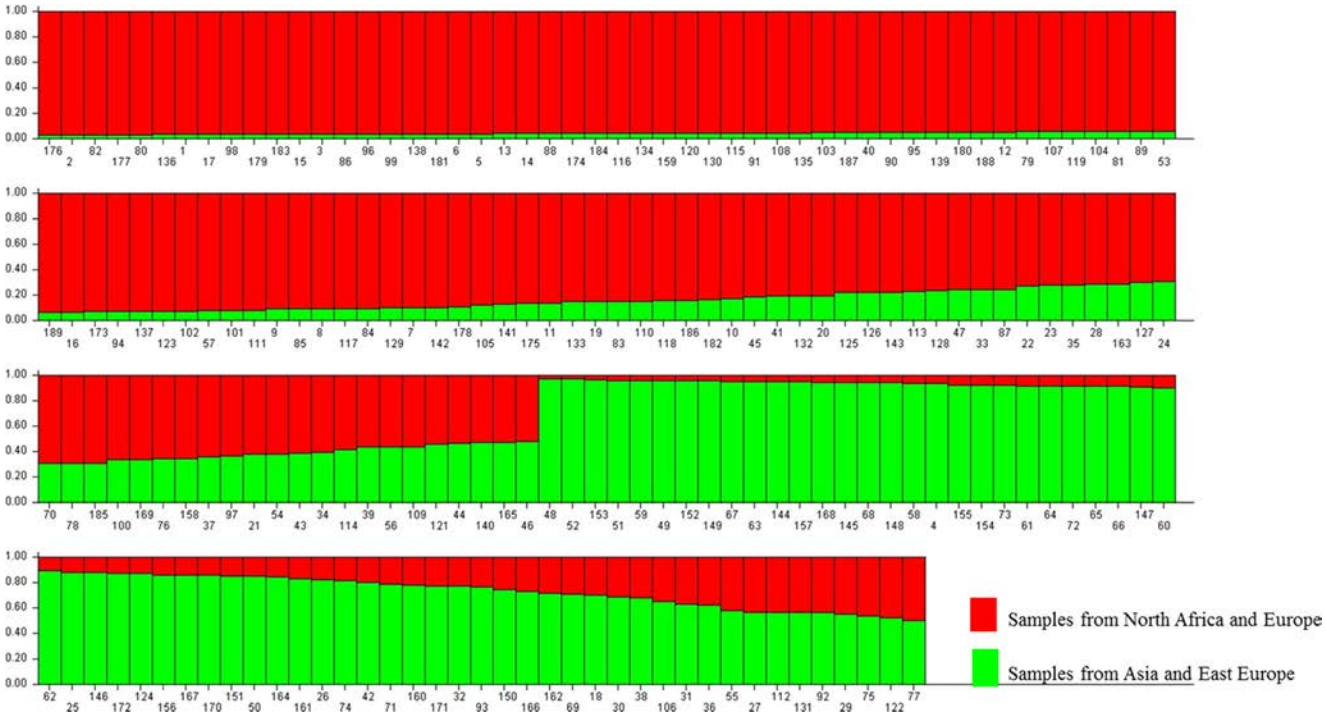

Fig. 2 Population structure of 189 accessions based on 10 SSR markers ($K = 2$) and graph of estimated membership fraction for $K = 2$

branches of the NJ tree (Fig. 4, Antalya and Diyarbakir from Turkey and Alborz from Iran) were joined in the STRUCTURE analysis as population six (light blue), and the fourth clade of Fig. 4 (Iraq, Greece, and Sirrt from Turkey) was concentrated into the third STRUCTURE population (dark blue in Fig. 3).

The NJ analysis showed important areas of overlap with the $K = 8$ partition of STRUCTURE. Branch five of the NJ

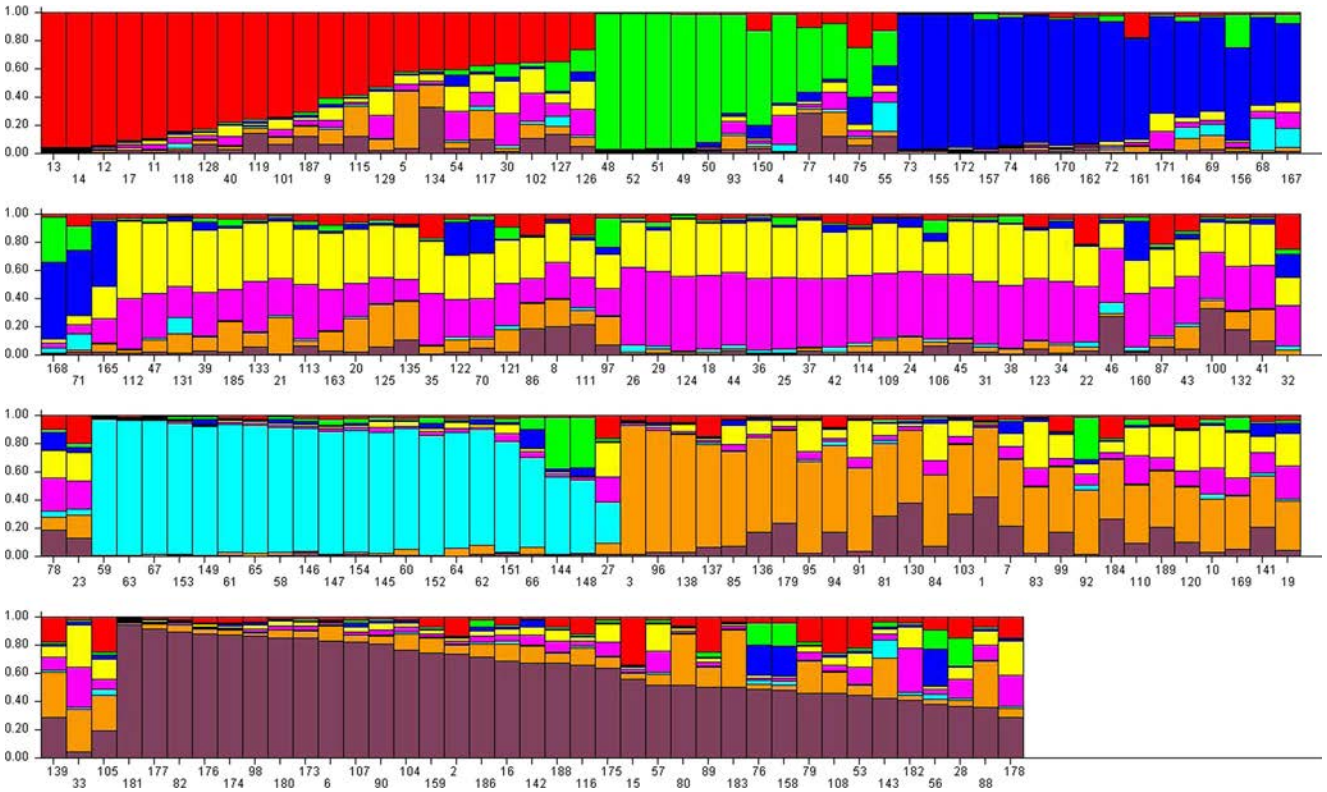

Fig. 3 Population structure of 189 Persian walnut accessions based on ten SSR markers ($K = 8$)

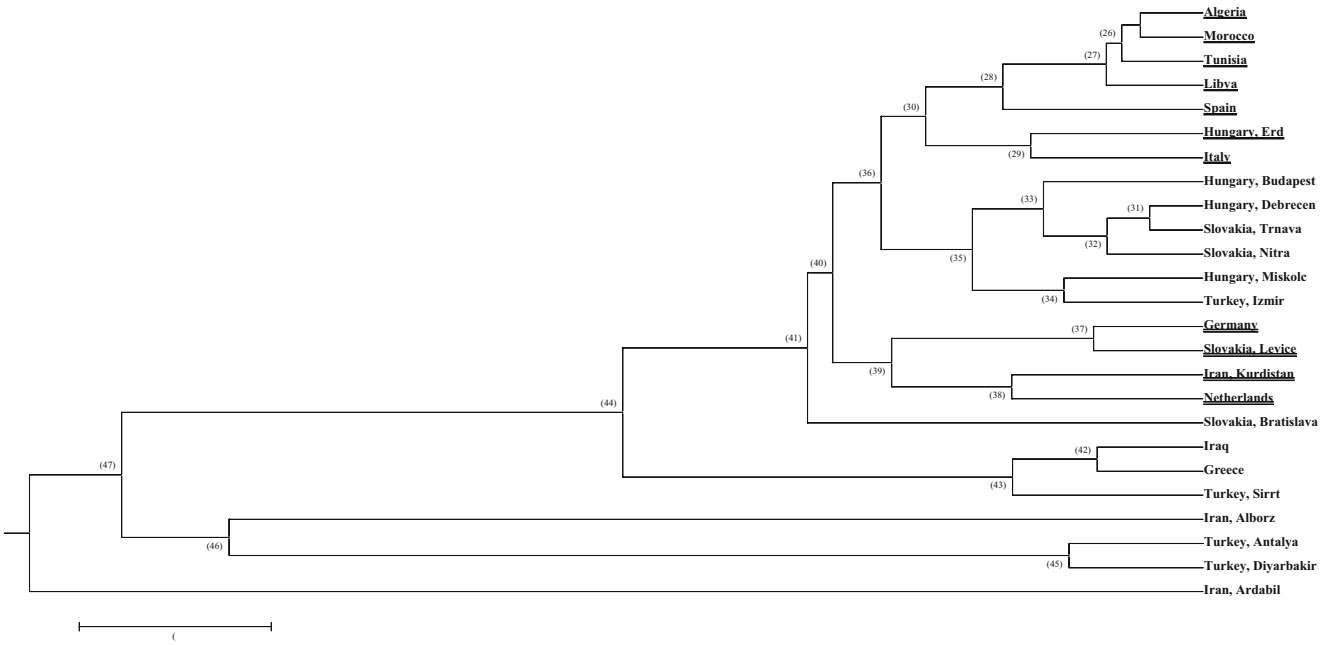

Fig. 4 Neighbor-joining dendrogram of 25 groups of *J. regia* based on Nei's genetic distance at ten SSR loci

tree (as described above, the upper 2/3 of Fig. 4, containing mostly European and North African sources) contained four sub-clusters. The first sub-cluster, which consisted of mostly Mediterranean sources (single underline in Fig. 4), was split into two populations by STRUCTURE, corresponding to population seven (Ochre) and eight (purple). Sub-cluster two of the NJ tree, which was strongly dominated by Hungarian and Slovakian samples, was also split into two populations containing individuals almost entirely equally admixed (pink and yellow, populations four and five in Fig. 3). German and Levice (Slovakia) samples of sub-cluster three of the NJ tree (sources double-underlined in Fig. 4) were concentrated by STRUCTURE into population one (red, Fig. 3). The other members of sub-cluster three (Netherlands and Kurdistan, Iran) were scattered by STRUCTURE across multiple populations, but did not fit well into any of them based on high levels of admixture. The fourth sub-cluster (Bratislava only) had a similar fate, as described previously in the example above. Both the STRUCTURE analyses ($K = 2$, Fig. 2 and $K = 8$, Fig. 3) and the NJ analysis revealed strong separation between most Turkish, Greek, and Iranian sources and those from Europe and North Africa. This latter group, based on NJ, appears to have a major split (Mediterranean versus Eastern European) but is in general highly admixed. Cluster analysis of 189 accessions using neighbor-joining showed that although many accessions were grouped into sub-branches based on their sampling sites, there were some clusters that contained accessions from different regions (Fig. S4). In general, it was difficult to reconcile the NJ tree based on individual genotypes with the STRUCTURE plots or the phonogram based on 25 sources (Fig. 4).

PCoA revealed that about 45% of the genotypic variance could be captured with two eigenvectors (see Table S2 in Appendix S1). A two-dimensional scatter plot involving all 189 accessions showed that the first two PCoA axes accounted for 41.4 and 3.8% of the genetic variation among populations (Fig. 5). Consistent with the fact that most variation was present within populations and only 7% among populations, we observed that accessions from regions largely overlapped, similar to the representation in Fig. S4.

Discussion

A complication of comparing population genetics studies is that allelic numbers are marker dependent. However, the comparisons presented here are meaningful as several of the microsatellite markers we used were also used by Foroni et al. (2007), Ebrahimi et al. (2011), and Pollegioni et al. (2011) making pairwise comparisons possible. For example, WGA69 produced 10 alleles in the present study but this locus had 6 alleles in 59 Turkish genotypes (Dogan et al. 2014), 7 alleles in 31 Iranian genotypes (Ebrahimi et al. 2011), and 10 alleles in 328 seedlings of *J. nigra* (Zhao et al. 2012). In the USDA repository, Aradhya et al. (2010) detected 13 different alleles in 840 *J. regia* accessions. A second marker, WGA276, was also used in previous studies and produced 18 alleles in our study, while 5, 16, and 8 alleles were reported by Ebrahimi et al. (2011), Dogan et al. (2014), and Vahdati et al. (2015), respectively. The wide geographic range from which our accessions were drawn, including especially the Iranian collection, may explain the higher overall number of alleles in this

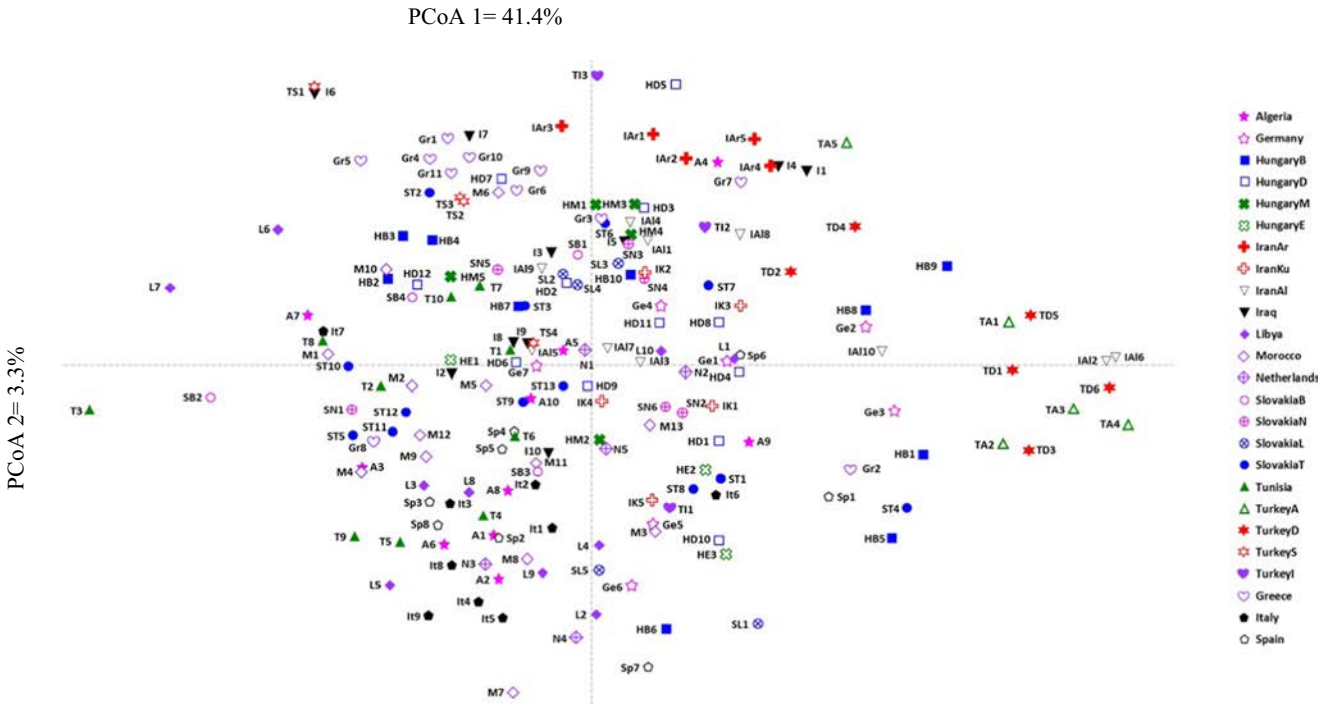

Fig. 5 BiPlot obtained from PCoA analysis of all accessions showing distribution of walnut samples in two main dimensions

study compared to other reports on *J. regia*. Average values of H_O and H_E were 0.60 and 0.73, respectively. Variation among populations essentially resulted from the large variation in the allele frequency distribution detected.

There were differences among the 25 studied *J. regia* populations for diversity indices (Table 3). Accessions from Debrecen (Hungary), which were exclusively from a gene bank, showed the highest values for the majority of diversity indices. Values for Shannon's index (I) were also high in Trnava (Slovakia), and both H_O and (I) were high for Greece. Thus, parts of Eastern and Southeastern Europe may contain unusually genetically diverse walnuts. Christopoulos et al. (2010) noted high levels of differentiation among local populations of *J. regia* in Greece, and reported that Greek samples were more genetically diverse than the other international cultivars they analyzed. The relatively high levels of genetic diversity and allelic richness we observed in walnut genotypes from Greece might reflect the genetic admixture of Asian and European gene pools, resulting from millennia of human trade across the Eurasian continent, as also suggested by Pollegioni et al. (2014). Alternatively, high levels of genetic diversity and allelic richness in Hungary, Slovakia, and Greece may reflect the presence of a separate center of diversity for walnut in the Carpathian Mountains. Carpathian walnuts are reported to be the most cold-hardy *J. regia* in the world (Mitra et al. 1991; Domoto 2002), so aside from a putative Carpathian center of diversity, it is possible the extreme conditions there resulted in strong selective pressure that influenced the walnut gene pool of this region.

Thirteen sources of *J. regia* had at least one private allele. Accessions from Iran (Alborz, Ardabil) contained the highest number of private alleles by far (12 alleles). Only 11% of the accessions came from Iran, but they contained 46% of the private alleles. In recent decades, most of the wild populations of walnut in Iran have been used extensively by local peoples for nut and wood production (Jafari-Sayadi et al. 2012). Examination of 39 wild populations of Persian walnut in Asia Pollegioni et al. (2014) reported the highest number of private alleles (8 alleles) in a wild population from Alborz, Iran. Surprisingly, this population was one of most sparsely populated (with only 12 accessions of 926) among the 39 analyzed populations. These authors reported that despite heavy interference by humans, high levels of genetic diversity still exist within Persian walnut population across the Asian range. In Iran, there is a long history of walnut cultivation and trade, and wide variety of both cultivated and wild forms of *J. regia* are found, despite selective cutting, general deforestation, and forest fragmentation, which has led to decreased population sizes. Luckily, because walnut, unlike many tree crops, has been propagated by seed almost exclusively throughout its history (Ebrahimi et al. 2007; Ebrahimi et al. 2011), a high percentage of the genetic variability of a region is maintained within walnut individuals (Table 4). So the remaining mature trees in Alborz and other fragmented forests represent the retention of a valuable genetic resource. This pattern has been observed in other forest species as well (e.g., *Boswellia papyrifera*, Addisalem et al. 2016). Based on these observations, we suggest that Iran's *J. regia* germplasm is unusually

rich and distinct from the other walnut accessions studied in this work and merits urgent conservation actions.

Accessions from neighboring countries were often similar. Accessions analyzed from Algeria were similar to those from Morocco, and Hungarian and Slovakian accessions were also similar, indicating a high level of germplasm exchange between these countries. German accessions were most similar to those germplasms from Levice (Slovakia) indicating that the *J. regia* cultivated in Germany might have originated elsewhere, or that Germany and Slovakia share a common, as yet uncollected, germplasm source. It is worth noting that many of the studied accessions were not wild populations, so they may have been affected by the introduction of improved genotypes from breeding programs in other regions, most likely in the form of seeds. Some germplasm introductions are region-specific, but others are designed to add genetic diversity from distant and distinct sources. By comparing walnut genotypes from Iran with some foreign cultivars, Ebrahimi et al. (2011) observed relatively high similarities between Iranian genotypes and some international cultivars and concluded that this reflected plant material exchange between countries. Similarly, Nicese et al. (1998) reported that "PI-159568" from the Iranian plateau was most likely in the pedigree of "Serr", an elite clonal selection from the breeding program in California, based on their genetic similarity. Foroni et al. (2005) reported a relatively high similarity between "Sorrento" and two other cultivars ("Serr" and "Hartley"). In our study, two accessions from Erbil (Iraq) were identical to two accessions from Sirrt (Turkey), likely the result of clonal propagation and the exchange of scion wood between these two neighbor regions. Many of our samples were taken from germplasm collections used for breeding (Table 1), and the international exchange of germplasm explains why genotypes from the same source in our study were dispersed among multiple populations (Fig. S5, Fig. 3) and also explains why bootstrap values separating sources were low (Fig. 4). For example, some accessions from Kurdistan, Iran, and Izmir, Turkey grouped with European accessions.

We found a high level of variation within groups (85%) and only 7% of variation was among groups (Table 4), which is similar to the level found by Aradhya et al. (2010) (86%), Christopoulos et al. (2010) (89%), and Wang et al. (2008) (81% within Chinese walnut populations). Relatively low levels of differentiation among samples from different regions can also be seen in the genetic identity matrix (Table S1) and the NJ analysis (Fig. 4), in which samples from the western Mediterranean (Algeria, Morocco, Tunisia, Libya, Spain, and Italy) were clustered. Partitioning of genetic variability within and among different walnut growing regions can provide useful data for optimizing sampling strategies in walnut cultivar improvement programs (Christopoulos et al. 2010). On the other hand, when we analyzed substructures in the NJ tree (Fig. 4), accessions from different locations within a country were sometimes strongly separated. This was more likely when samples were not from germplasm collections, or, perhaps, when sample sizes were small (Bratislava, Slovakia). Accessions from Ardabil, Iran, were separated from the other Iranian accessions. It is noteworthy that Ardabil's climate is the coldest in the region and, because of high mountains, the location is naturally secluded. This natural isolation is a factor that may restrict pollen flow and seed dispersal.

Conclusion

Accessions of 189 *J. regia* individuals collected from 25 different regions within a total of 14 countries were each analyzed using ten microsatellites. We observed high levels of diversity among the collected accessions as a whole and among the collection regions. The mean number of observed alleles was relatively high, and two thirds of these were rare and private alleles, which could be used to identify and distinguish local germplasm in need of conservation. Accessions from Iran had the greatest number of private alleles with those from Ardabil being the most distinct. Accessions from Eastern and Southeastern Europe were also highly diverse compared to other studied accessions, lending credence to the idea that there are at least two main different centers of diversity for Persian walnut, one from Southwest Asia and another from the Carpathian region in Europe. Accessions from Eastern Europe showed the highest degree of admixture, possibly implying that these regions act as a hybridization zone for Asian and European *J. regia*. Cluster analysis divided the genotypes into two main groups in accordance with their geographical distribution, and accessions from Asia Minor and Greece were separated from Europe and North Africa. Because the plant material sampled in this research was representative of walnut germplasm grown in different regions of walnut cultivation, the results provide important new information about the distribution of germplasm among nations and the importance of maintaining the genetic diversity within this prized hardwood tree species. Nearly all of our samples were derived from accession and germplasm that may or may not have value to breeders based on the tree's phenotypes, including nut quality, kernel quality, disease resistance, frost resistance, and phenology. Breeders will need information about all these traits before deciding if germplasm from another location might be useful, and consultation with local experts is the best way to obtain information about sources described in this manuscript (Germain 2004).

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

Data archiving If the manuscript is accepted, genotype data will be submitted to TreeGenes database.

Disclaimer Mention of a trademark, proprietary product, or vendor does not constitute a guarantee or warranty of the product by the U.S. Dept. of Agriculture and does not imply its approval to the exclusion of other products or vendors that also may be suitable.

References

Addisalem AB, Bongers F, Kassahun T, Smulders MJM (2016) Genetic diversity and differentiation of the frankincense tree (*Boswellia papyrifera* (Del.) Hochst) across Ethiopia and implications for its conservation. For Ecol Manag 360:253–260. doi:10.1016/j.foreco.2015.10.038

Aradhya M, Woeste K, Velasco D (2010) Genetic diversity, structure and differentiation in cultivated walnut (*Juglans regia* L). Acta Hort 861: 127–132

Avanzato D, McGranahan GH, Vahdati K, Botu M, Iannamico L, Van Assche J (2014) Following walnut footprints (*Juglans regia* L.) cultivation and culture, folklore and history, traditions and uses. Scripta Horticulturae, p 442

Bayazit S, Kazan K, Gülbitti S, Çevik V, Ayanoğlu H, Ergül A (2006) AFLP analysis of genetic diversity in low chill requiring walnut (*Juglans regia* L.) genotypes from Hatay, Turkey. Sci Hort 111: 394–398

Christopoulos MV, Rouskas D, Tsantili E, Bebeli PJ (2010) Germplasm diversity and genetic relationships among walnut (*Juglans regia* L.) cultivars and Greek local selections revealed by Inter-Simple Sequence Repeat (ISSR) markers. Sci Hort 125:584–592

Cornille A, Gladieux P, Smulders MJM, Roldán-Ruiz I, Laurens F, Le Cam B, Nersesyan A, Clavel J, Olonova M, Feugey L, Gabrielyan I, Zhang XG, Tenaillon MI, Giraud T (2012) New insight into the history of domesticated apple: secondary contribution of the European wild apple to the genome of cultivated varieties. PLoS Genet 8:e1002703. doi:10.1371/journal.pgen.1002703

Dangl GS, Woeste K, Aradhya MK, Koehmstedt A, Simon C, Potter D, Leslie CA, McGranaham G (2005) Characterization of 14 microsatellite markers for genetic analysis and cultivar identification of walnut. J Amer Sco Hort Sci 130:348–354

Dogan Y, Kafkas S, Sutyemez M, Turemis N (2014) Assessment and characterization of genetic relationships of walnut (*Juglans regia* L.) genotypes by three types of molecular markers. Sci Hort 168: 81–87

Domoto PA (2002) Walnut tree named 'Domoto'. United States Plant Patent № US PP12, 898 P2

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19:11–15

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. doi:10.1007/s12686-011-9548-7

Ebrahimi A, Vahdati K, Fallahi E (2007) Improved success of Persian walnut grafting under environmentally controlled conditions. Int J Fruit Sci 6(4):3–12

Ebrahimi A, Fatahi R, Zamani Z (2011) Analysis of genetic diversity among some Persian walnut genotypes (*Juglans regia* L.) using morphological traits and SSRs markers. Sci Hort 130:146–151. doi:10.1016/j.scienta.2011.06.028

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

Foroni I, Rao R, Woeste K, Gallitelli M (2005) Characterization of *Juglans regia* L. through SSR markers and evaluation of genetic relationships among cultivars and the 'Sorrento' landrace. J Hortic Sci Biotechnol 80:49–53

Foroni I, Woeste K, Monti LM, Rao R (2007) Identification of 'Sorrento' walnut using simple sequence repeats (SSRs). Genet Resour Crop Evol 54:1081–1094. doi:10.1007/s10722-006-9187-0

Germain, E (2004) Inventory of walnut research, germplasm and references. FAO Regional Office For Europe, REU Technical Series 66, Rome, Italy, p 264

Jafari Sayadi MH, Vahdati K, Mozafari J, Mohajer MRM, Leslie CA (2012) Natural hyrcanian population of Persian walnut (*Juglans regia* L.) in Iran. Acta Hortic 948:97–101. doi:10.17660/ActaHortic.2012.948.10

Kalinowski ST (2005) HP-rare: a computer program for performing rarefaction on measures of allelic diversity. Mol Ecol Notes 5:187–189

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. Mol Ecol 7:639–655

McGranahan GH, Charles A, Leslie CA, Philips HA, Dandaker A (1998) Walnut propagation. In: Ramos D (ed) Walnut production manual. University of California, DANR Publ, Davis, pp. 71–83

Mitra SK, Rathore DS, Bose TK (1991) Walnut. Temperate fruits, vol 27. Published by Horticulture and Allied Publishers, Calcutta, pp. 377–414

Nei M (1972) Genetic distance between populations. Am Nat 106:283–292

Nicese FP, Hormasa JI, McGranaham GH (1998) Molecular characterization and genetic relatedness among walnut (*Juglans regia* L.) genotypes based on RAPD markers. Euphytica 101:199–206

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

Pollegioni P, Woeste K, Olimpieri I, Marandola D, Cannata F, Malvolti ME (2011) Long term human impacts on genetic structure of Italian walnut inferred by SSR markers. Tree Genet Genomes 7:707–723

Pollegioni P, Woeste KE, Chiocchini F, Olimpieria L, Tortolano V, Clark J, Hemery GE, Mapelli S, Malvoti ME (2014) Landscape genetics of Persian walnut (*Juglans regia* L.) across its Asian range. Tree Genet Genomes 10:1027–1043

Pollegioni P, Woeste K, Chiocchini F, Del Lugo S, Cannata F, Olimpieri I, Tortolano V, Clark J, Hemery GE, Mapelli S, Malvolti EM (2015) Ancient humans influenced the current spatial genetic structure of common [Persian] walnut populations in Asia. PLoS One 7:707–723. doi:10.137/journal.pone.0135980

Pop IL, Vicol AC, Botu M, Raica PA, Vahdati K, Pamfil D (2013) Relationships of walnut cultivars in a germplasm collection: comparative analysis of phenotypic and molecular data. Sci Hort 153: 124–135

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2012) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729

Topcu H, Coban N, Woeste K, Sutyemez M, Kafkas S (2015) Developing new microsatellite markers in walnut (*Juglans regia* L.) from *Juglans nigra* genomic GA enriched library. EKIN 1-2:93–99

Vahdati K, Mohseni Pourtaklu S, Karimi R, Barzehkar R, Amiri R, Mozaffari M, Woeste K (2015) Genetic diversity and gene flow of some Persian walnut populations in southeast of Iran revealed by SSR markers. Plant Syst Evol 301:691–699. doi:10.1007/s00606-014-1107-8

Wang H, Pei D, Gu R, Wang B (2008) Genetic diversity and structure of walnut populations in central and southwestern China revealed by microsatellite markers. J Am Soc Hortic Sci 133:197–203

Watanabe S, Noma N, Nishida T (2016) Flowering phenology and mating success of the heterodichogamous tree *Machilus thunbergii* Sieb. et Zucc (Lauraceae). Plant Species Biology 31:29–37

Woeste K, Burns R, Rhodes O, Michler C (2002) Thirty polymorphic nuclear microsatellite loci from black walnut. J Hered 93:58–60

Yeh FC, Yang RC, Boyle TBJ, Ye ZH, Mao JX (1997) POPGENE, the user-friendly shareware for population genetic analysis. Molecular Biology and Biotechnology Center, University of Alberta, Edmonton, Alberta, Canada

Zhao P, Zhang S, Woeste K (2012) Genotypic data changes family rank for growth and quality traits in a black walnut (*Juglans nigra* L.) progeny test. New For. doi:10.1007/s11056-012-9343-7