Genet Resour Crop Evol (2011) 58:519–531
DOI 10.1007/s10722-010-9595-z

RESEARCH ARTICLE

Molecular characterization of genetic diversity, structure, and differentiation in the olive (*Olea europaea* L.) germplasm collection of the United States Department of Agriculture

Anne M. Koehmstedt · Mallikarjuna K. Aradhya · Daniela Soleri · Jenny L. Smith · Vito S. Polito

Received: 25 March 2010 / Accepted: 2 August 2010 / Published online: 26 August 2010

Abstract Fifteen microsatellite loci were used to genotype 108 accessions of cultivated olive, *Olea europaea* L. ssp. *europaea* var. *europaea*, and eight of *O. europaea* L. ssp. *cuspidata* (Wall. ex G. Don) Ciferri, from the germplasm collection of the United States Department of Agriculture in Davis, California. Number of alleles per locus ranged from 3, for locus IAS-pOe12_A, to 16, for locus ssrOeUA-DCA11, with an overall mean of 9.93. Observed heterozygosity ranged from 0.175, for locus UDO99–019, to 0.937, for locus GAPU89, with a mean of 0.640. The cluster analysis using the Unweighted Pair Group Method using Arithmetic mean (UPGMA) method displayed thirteen clusters within seven main groups that can be partially described by common geographic origin or fruit use, though overlap among these groups was common. The locus-wise total gene diversity (H_T) ranged from 0.319, at UDO99–019, to 0.847, at ssrOeUA-DCA3, with an overall mean of 0.696. Most of the gene diversity was partitioned within clusters, with proportions (H_S/H_T) ranging from 0.633, at IAS-pOe12_B, to 0.848 at GAPU89 per locus, with a mean of 0.759. The principal components analysis explained 24.8% of the total variation along the first two components. Projection of accessions onto the first two principal components produced affinities generally in agreement with the results of the UPGMA cluster analysis. The California cultivar 'Mission' clustered closely with Iberian cultivars and may represent clonal selections adapted to local growing conditions. The results show significant diversity but low levels of differentiation among olive cultivars within the collection.

Keywords Cluster analysis · DNA fingerprinting · Microsatellites · *Olea europaea*

Electronic supplementary material The online version of this article (doi:10.1007/s10722-010-9595-z) contains supplementary material, which is available to authorized users.

A. M. Koehmstedt (✉) · M. K. Aradhya · J. L. Smith
National Clonal Germplasm Repository, USDA-ARS, One Shields Avenue, Davis, CA 95616, USA
e-mail: anne.koehmstedt@ars.usda.gov

D. Soleri
Geography Department, University of California Santa Barbara, Santa Barbara, CA 93106-4060, USA

V. S. Polito
Department of Plant Sciences, University of California, One Shields Avenue, Davis, CA 95616, USA

Introduction

Olive, native to the Mediterranean Basin, is believed to have been first introduced into California by the Spanish Franciscan missionaries in the late eighteenth century (Taylor 2000). Compared to leading olive producers such as Spain (2.42 million ha), Italy (1.4 million ha), Tunisia, and Greece (1.03 million ha), California is a minor player in the worldwide table

olive and oil markets (Vossen 2007). Yet, most olives produced in the United States come from California, grown on 12,245 ha in 2007 (U.S. Department of Agriculture 2008). In recent years, California domestic olive oil production has almost doubled, from 936,889 liters in 2000 to 1.514 million liters in 2005, with a projected increase to 3.785 million liters by 2013 (Vossen 2007; Vossen and Devarenne 2005).

The U.S. Department of Agriculture's National Clonal Germplasm Repository for Fruit and Nut Crops (NCGR) maintains a field collection of olive germplasm at the Wolfskill Experimental Orchard (WEO) in Winters, California. This collection contains 140 individual accessions, which includes about 120 uniquely named cultivars of *Olea europaea* L. ssp. *europaea* var. *europaea* from the Mediterranean Basin, South America and California. The collection also contains several accessions of hybrids and the related subspecies, *O. europaea* L. ssp. *cuspidata* (Wall. ex G. Don) Ciferri, which has a native range throughout Africa, and South-Western and Subcontinental Asia (Green 2002).

Several previous olive genetic diversity studies applied molecular marker systems including isozymes (Trujillo et al. 1995), amplified fragment length polymorphisms (AFLP; Sanz-Cortes et al. 2003), random amplified polymorphic DNA [RAPD; Besnard et al. 2001(a)] and microsatellite simple-sequence-repeats (SSR; Lopes et al. 2004) to examine genetic diversity, structure, and differentiation among olive cultivars, as well as to characterize germplasm collections and explore cultivar identity. These studies encompass collections and accessions from throughout the Mediterranean and secondary growing regions such as Australia (Sweeney 2003). Belaj et al. (2003) compared the usefulness of RAPD, AFLP, and SSR markers for identification and genetic differentiation of 32 Spanish and Italian olive cultivars. They concluded that SSR markers, due to their co-dominant nature, high levels of polymorphism and reproducibility, have a higher discriminating power for cultivar identification, and are ideal for olive genome mapping and genetic studies.

This study used microsatellite markers to examine the level of genetic diversity, structure and differentiation within *Olea europaea* accessions maintained in the NCGR collection, based on their genetic affinities, geographic origin and fruit use for oil, table, or dual (oil and table) purposes.

Materials and methods

Plant material collection and DNA extraction

A total of 116 accessions in the NCGR olive collection at the WEO were sampled (Table 1). Fresh, young, fully expanded leaves were collected in 2005 or 2008 for DNA isolation. Homogenization was accomplished by freezing fresh leaf tissue in liquid nitrogen, and then either grinding with a mortar and pestle or in tubes using a Geno/Grinder 2000 (OPS Diagnostics, Bridgewater, NJ). DNA was extracted from homogenized leaf tissue using either QIAGEN DNeasy Plant Mini-prep Kits (QIAGEN, Valencia, CA), or a CTAB-based protocol modified from Doyle and Doyle (1987). DNA was suspended in TE buffer (pH = 8) and digested with RNase A (QIAGEN) at 37 °C for 1 h. DNA samples were diluted to approximately 20 ng/μl before conducting PCR.

Amplification of SSR loci and fragment separation

Twenty-seven microsatellite markers previously developed from *O. europaea* ssp. *europaea* were screened for usefulness on a diverse subset of cultivars from the NCGR collection. Based on criteria of PCR amplification consistency, polymorphism, and allele scoring consistency, the fourteen best markers were chosen: EM03, EM030, EM088, EM090 (de la Rosa et al. 2002), GAPU89 (Carriero et al. 2002), IAS-oli16 (Rallo et al. 2000), IAS-pOe12 (Genbank Acc. No. AF356039 2001), ssrOeUA-DCA03, ssrOeUA-DCA08, ssrOeUA-DCA11 (Sefc et al. 2000), UD099-011, UD099-019, UD099-028, and UD099-042 (Cipriani et al. 2002). The forward primer was labeled with one of three fluorescent dyes: 6-FAM, HEX, or NED [Applied Biosystems, Foster City, CA (ABI)]. However, one marker, IAS-pOe12, was found to PCR amplify two loci instead of one, and these two loci were separated and identified for final analysis as IAS-pOe12_A and IAS-pOe12_B, bringing the total number of microsatellite loci analyzed to fifteen.

Each PCR reaction consisted of 1× Standard Taq Buffer [New England BioLabs, Ipswich, MA (NEB)], 2 mM $MgCl_2$, 0.375 mM each dNTP (ABI), 0.075 units/μl Taq DNA Polymerase (NEB),

Table 1 List of accessions sampled in this study, their geographic origin, and fruit use type

Acc. #	Accession name*	Country of origin	Use
DOLE0101	3s5-117	US	Unk
DOLE0137	Adrouppa	Cyprus	Table
DOLE0131	Ageezy Shami	Egypt	Table
DOLE0039	Arbequina	Spain	Oil
DOLE0186	Arbussana	Spain	Oil
DOLE0028	Ascolano	Italy	Table
DOLE0054	Ascolano Dura	Italy	Table
DOLE0044	Asioloni	Italy	Table
DOLE0119	Azapa	Chile	Table
DOLE0145	Balady	Egypt	Dual
DOLE0111	Barouni	Tunisia	Table
DOLE0015	Bidh El Hammam	Tunisia	Dual
DOLE0108	Black Italian	Italy	Unk
DOLE0106	Bouquetier	France	Oil
DOLE0076	Bouteillon	France	Oil
DOLE0007	Chalkidiki	Greece	Table
DOLE0089	Chemlali	Tunisia	Oil
DOLE0091	Chitoni	Tunisia	Dual
DOLE0051	Columello	France	Unk
DOLE0115	Conservolia	Greece	Dual
DOLE0070	Cordovil	Portugal	Dual
DOLE0175	Cypress 31	Unknown	Unk
DOLE0179	Dolce De Napoli	Italy	Unk
DOLE0084	Dolce del Marocco	Italy	Table
DOLE0031	Dwarf D	US	Unk
DOLE0187	Empeltre	Spain	Oil
DOLE0036	Franklin	US	Unk
DOLE0181	Frantoio	Italy	Oil
DOLE0002	Frantojoi	Italy	Oil
DOLE0154	Gaidourelia	Greece	Table
DOLE0207	Galega	Portugal	Dual
DOLE0148	Giaraffa	Italy	Table
DOLE0138	Grappolo	Italy	Oil
DOLE0026	Grossa Di Spagna	Italy	Table
DOLE0035	Grossane	France	Dual
DOLE0139	Grosse Aberkan	Algeria	Dual
DOLE0034	Hamid	Egypt	Table
DOLE0159	Jalut	Syria	Unk
DOLE0147	K18	Israel	Dual
DOLE0065	Kadesh	Israel	Table
DOLE0189	Kalamata	Greece	Dual
DOLE0152	Karolia	Greece	Dual
DOLE0080	Karydolia	Greece	Table
DOLE0149	Koroneiki	Greece	Oil
DOLE0173	Late Blanquette	France	Dual
DOLE0027	Lecci	Italy	Oil
DOLE0184	Leccino	Italy	Oil
DOLE0050	Liguria	Chile	Oil
DOLE0077	Manzanillo	Spain	Dual
DOLE0158	Massabi	Syria	Table
DOLE0153	Mastoides Tsounati	Greece	Dual
DOLE0143	Maurini	Italy	Oil
DOLE0151	Megaron	Greece	Dual
DOLE0160	Memeli	Greece	Dual
DOLE0167	Merhavya	Israel	Table
DOLE0068	Meski	Tunisia	Table
DOLE0180	Midx-elbasan	Albania	Dual
DOLE0057	Mission	US	Dual
DOLE0129	Mission Leiva	Columbia	Dual
DOLE0042	Mission Variant	US	Dual
DOLE0140	Mission-nieland	US	Dual
DOLE0123	Morcal	Spain	Table
DOLE0008	Mostazal	Peru	Dual
DOLE0113	Nab Tamri	Tunisia	Unk
DOLE0166	Nabali	Israel	Dual
DOLE0064	Nevadillo	Spain	Oil
DOLE0183	Nikitiskaja Krupnoplodnaja	Russia	Unk
DOLE0182	Nikitskaya #1	Russia	Dual
DOLE0163	No. 1 Sevillano	Spain	Table
DOLE0164	No. 12 Sevillano	Spain	Table
DOLE0165	No. 31 Sevillano	Spain	Table
DOLE0168	No. 63	Cyprus	Unk
DOLE0169	No. 65a	Cyprus	Unk
DOLE0124	Nuevo Di Sicrone	Italy	Unk
DOLE0162	Obliza	Croatia	Dual
DOLE0003	Oblonga	Italy	Oil
DOLE0022	Ogliarola	Italy	Oil
DOLE0114	Olivo A Prugno	Italy	Table
DOLE0122	Piconia	Greece	Oil
DOLE0073	Picual	Spain	Oil
DOLE0155	Prunara	Italy	Oil
DOLE0071	Redding Picholine	US	Dual
DOLE0079	Redondillo De Lomgrona	Spain	Dual
DOLE0170	Rigali	Italy	Oil
DOLE0171	Ropades	Greece	Unk

Table 1 continued

Acc. #	Accession name*	Country of origin	Use
DOLE0118	Rouget	France	Oil
DOLE0097	Rubra	France	Oil
DOLE0083	Saiali Magloub	Tunisia	Oil
DOLE0144	Salome	Greece	Dual
DOLE0019	San Francesco	Italy	Table
DOLE0013	Sevillano	Spain	Table
DOLE0172	Sevillano-lovisone	US	Table
DOLE0142	Sir George Greys Spanish	Spain	Oil
DOLE0023	Souri	Lebanon	Dual
DOLE0116	ssp. *cuspidata* (DOLE116)	Unknown	n/a
DOLE0132	ssp. *cuspidata* (DOLE132)	South Africa	n/a
DOLE0176.1	ssp. *cuspidata* (DOLE176.1)	Pakistan	n/a
DOLE0176.3	ssp. *cuspidata* (DOLE176.3)	Pakistan	n/a
DOLE0176.6	ssp. *cuspidata* (DOLE176.6)	Pakistan	n/a
DOLE0177.8	ssp. *cuspidata* (DOLE177.8)	Croatia	n/a
DOLE0177.9	ssp. *cuspidata* (DOLE177.9)	Croatia	n/a
DOLE0090	ssp. *cuspidata* (DOLE90)	Unknown	n/a
DOLE0094	Syrogylolia	Greece	Dual
DOLE0185	Thrombolea	Greece	Oil
DOLE0110	Throumbolia	Greece	Oil
DOLE0161	Touffahi	Egypt	Table
DOLE0001	Tragolea	Greece	Dual
DOLE0156	Uc 49-14 (Ascolano × Barouni)	US	Unk
DOLE0126	Uc 52.24.1 (hybrid)	US	Unk
DOLE0063	Vassilika	Greece	Dual
DOLE0103	Verdale	France	Dual
DOLE0102	Verdeal	Portugal	Oil
DOLE0055	Verdena	Spain	Oil
DOLE0086	Yuaca	Peru	Oil
DOLE0025	Zitoum	Morocco	Dual
DOLE0121	Zoragi	Tunisia	Table

*Accession names are verified for spelling based on available passport data for each when received into the NCGR collection

0.05 pmol/μl each primer, and approximately 5 ng/μl DNA. PCR reactions were triplexed, i.e. made with three primer pairs combined in one reaction, each pair labeled with a different fluorescent dye. PCR was performed under the following conditions: 1 cycle of 94 °C for 5 min, 30 cycles of 94 °C for 30 s, 55 °C for 30 s, and 72 °C for 40 s, and then a final elongation of 72 °C for 7 min.

Amplified SSR fragments (approximately 0.8 μl PCR product/sample) were combined with 12 μl Hi-Di Formamide and 0.2 μl HD400 ROX Size Standard (ABI) per sample. Samples were denatured at 95 °C for 5 min. Fragment separation by capillary electrophoresis was conducted on a 3100 Genetic Analyzer (ABI). At least one copy of each available, unique accession was fingerprinted for the fifteen chosen SSR loci.

Data analysis

SSR fragments were sized in nucleotide base pairs and binned into allele categories using GeneScan Analysis Software v. 3.1 and Genotyper v. 2.5 (ABI) (Online Resource 1). Allelic data for all loci were combined and converted to a binary format to produce a rectangular data matrix (1 = presence, 0 = absence).

Using PAUP* v. 4.0 (Swofford 2003), a pairwise genetic distance matrix was computed using the Nei and Li similarity coefficient (Nei and Li 1979), which was then used to perform cluster analysis to generate a phenogram using the Unweighted Pair Group Method using Arithmetic mean (UPGMA) algorithm. Bootstrapping with 1,000 replicates and no resampling was conducted to determine support for each node. Allelic data were grouped into clusters based on the UPGMA cluster analysis. Using BIOSYS-1, v. 1.7 (Swofford and Selander 1989) the clusters were subjected to various within-cluster genetic diversity parameters, such as mean number of alleles/locus, observed and expected levels of heterozygosity, polymorphic index, and contingency chi-squared tests.

Total gene diversity (H_T), gene diversity within clusters (H_S), and proportion of diversity among clusters (G_{ST}) (Nei 1973) were calculated using the

computer program, Dispan: Genetic Distance and Phylogenetic Analysis (Ota 1993). Genetic divergence among groups was examined using the distance-Wagner tree based on the Prevosti distance measure (Swofford 1981; Wright 1978). The program NTSYSpc: Numerical Taxonomy System, v. 2.1.1 (Rohlf 2002) was used for principal components analysis (PCA) using the Dice coefficient distance matrix (Dice 1945). To simplify analysis and according to NTSYSpc requirements, synonymous genotypes, samples with missing data at three or more loci, and the ssp. *cuspidata* accessions were excluded from the data set for PCA.

Results

Genetic diversity, structure, and intraclonal variation within the NCGR collection

Per locus number of alleles ranged from three at IAS-pOe12_A to 16 at ssrOeUA-DCA11, with a mean of 9.93 alleles per locus. Expected heterozygosity ranged from 0.227 at UDO99–019 to 0.860 at ssrOeUA-DCA3 with a mean of 0.687, while observed heterozygosity ranged from 0.175 at UDO99–019 to 0.937 at GAPU89, with a mean of 0.640 (Table 2). Most loci had two or three moderate to high-frequency (20% or higher) alleles. Locus-wise fixation indices were significant at the 5 or 10% level in six loci (EM030, EM088, IAS-pOe12_A, IAS-pOe12_B, ssrOeUA-DCA11, and UDO99–019), all indicating a deficiency of heterozygotes as compared to Hardy–Weinberg (HW) expected levels. Of the 116 total accessions analyzed in the collection, 88 unique, multilocus genotypes were generated.

The cluster analysis using UPGMA method showed low to moderate differentiation within the olive collection with a total of twelve clusters (C1–C12), organized into seven major groups, partially attributable to common geographic origin and/or fruit use (Fig. 1). Most clusters had moderate bootstrap support (50% or above). Accessions from ssp. *cuspidata* clustered as outliers. Contingency chi-squared (X^2) analysis showed significant differences in allele composition and frequencies across all loci among the different UPGMA clusters.

The first group in the UPGMA phenogram is a cluster of oil producing cultivars of French, Greek, South American, and Tunisian origin. Group two is a

Table 2 Per locus number of alleles (A), allele size range in base pairs (bp), unbiased estimate of expected heterozygosity (H_e), direct-count observed heterozygosity (H_o), and fixation index (F)

Locus	A	Size range (bp)	H_e(unb)	H_o(d.c)	F^a
EM03	10	194–217	0.714	0.793	−0.111
EM030	11	181–197	0.706	0.457	0.353**
EM090	5	182–194	0.646	0.745	−0.153
EMO88	9	178–198	0.783	0.500	0.361**
GAPU89	12	156–207	0.790	0.937	−0.186
IAS-oli16	15	136–164	0.781	0.784	−0.004
IAS-pOe12_A	3	106–114	0.362	0.289	0.202*
IAS-pOe12_B	5	117–123	0.458	0.360	0.214*
ssrOeUA-DCA11	16	127–165	0.779	0.543	0.303**
ssrOeUA-DCA3	13	230–260	0.860	0.879	−0.022
ssrOeUA-DCA8	14	112–146	0.786	0.870	−0.107
UDO99-011	13	89–133	0.838	0.931	−0.111
UDO99-019	7	125–165	0.227	0.175	0.229*
UDO99-028	13	119–167	0.798	0.704	0.118
UDO99-042	8	140–164	0.774	0.637	0.177
Mean	9.93	–	0.687	0.640	0.084

[a] Wright's fixation index $= (H_e - H_o)/H_e$; * $P < 0.05$; ** $P < 0.01$

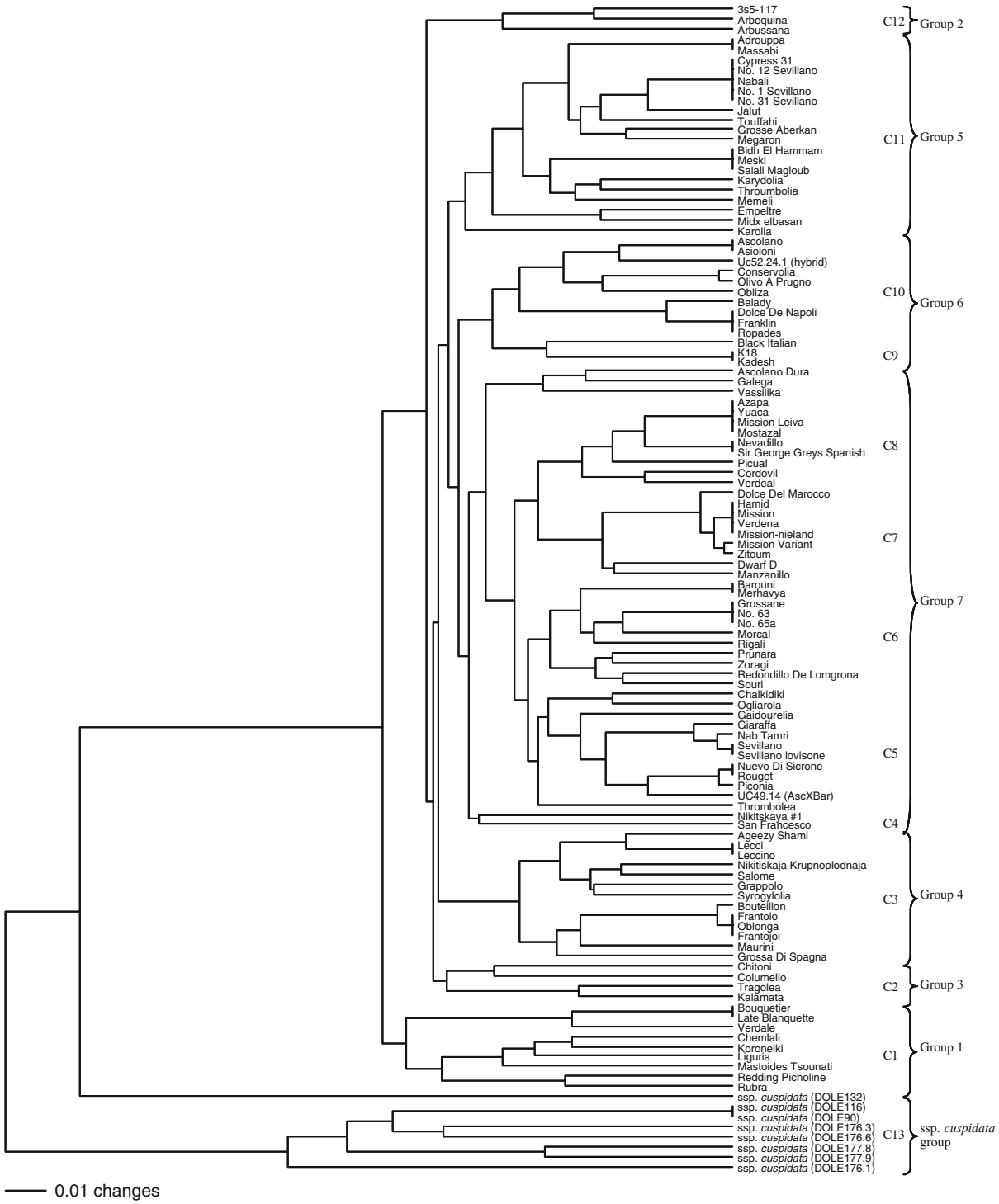

Fig. 1 UPGMA phenogram from Nei and Li (1979) genetic distance measures using fifteen SSR loci on 116 cultivated olive and ssp. *cuspidata* accessions within the NCGR collection

separate, small cluster of the Spanish oil cultivars 'Arbequina' and 'Arbussana', as well as '3s5-117' of unknown origin or use. Group three contains two Greek, one French, and one Italian, dual purpose cultivars. The fourth group is a cluster of mainly Italian oil and Greek dual purpose cultivars. Group five is a medium sized group of cultivars from the Eastern Mediterranean, mainly Greece and Tunisia, of various use types. The sixth group is made up of primarily dual purpose cultivars from throughout the Mediterranean region. Group seven is the largest and comprises most of the Iberian and Italian cultivars of various use types. Seventeen groups of synonyms can also be seen in the UPGMA tree as zero branch lengths (Online Resource 2).

Six instances of closely related accessions, with one or two allelic differences over fifteen SSR loci, were found within the NCGR olive collection. The instances found in this study, however, include accessions that carry unique cultivar names. 'Zitoum' differed from 'Mission' at only one allele by 2 bp at the imperfect dinucleotide repeat locus GAPU89, as did 'Mission Variant', with the addition of a second allelic difference at locus UDO99-028. The accession 'Nab Tamri' differed from 'Sevillano' at one, 2 bp difference allele at locus GAPU89. 'Rouget' differed from 'Piconia' at locus UDO99-028, 'Conserviola' from 'Oliva A Prugno' at ssrOeUA-DCA8, and 'Bouteillon' from 'Frantoio' at UDO99-028. A comparison of the 'Mission' genotype to several Iberian cultivars revealed that all of the latter could contribute at least one allele to the 'Mission' genotype at thirteen or fourteen SSR loci, and at all fifteen loci for 'Sevillano' (Table 5).

Principal components analysis

The first two components of the PCA explained 24.8% of the total variation among a subset of 99 accessions included in the analysis. The first component explained 20.0% of the variation, followed by 4.8% for the second components. Projection of the unique genotypes on a two-dimensional plane, based on the first two principal components, partially confirmed the results of the UPGMA method. The accessions from the UPGMA clusters likewise tended to form their own groups in the PCA, though usually overlapping (Fig. 2). In the top two quadrants of the projection, a loose group of accessions was comprised only from clusters 1, 3, and 10 of the UPGMA phenogram. Overlapping groups of accessions from almost all UPGMA clusters comprise the central to lower quadrants of the projection. They represent various geographic origins and uses.

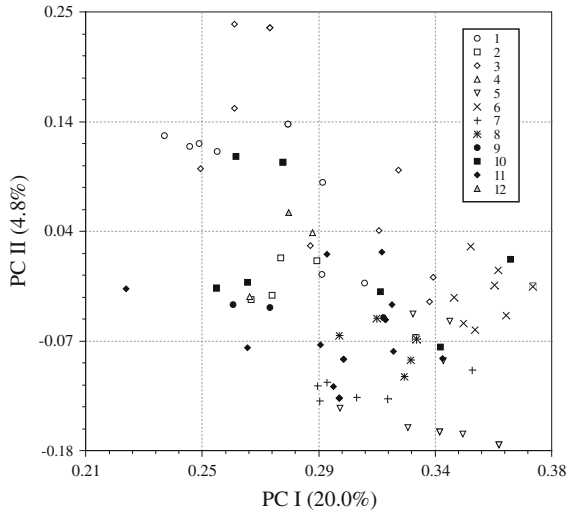

Fig. 2 Two-dimensional projection of olive accessions along the first two principal axes, accounting for 24.8% of the total variation; symbols assigned by UPGMA cluster

Within groups genetic diversity and differentiation

The mean number of alleles per locus among the thirteen clusters found on the UPGMA tree, ranged from 2.1 for C4 to 5.3 for the ssp. *cuspidata* cluster (C13), with an overall mean of 3.6 alleles per locus per cluster. Mean expected heterozygosity (H_e) within groups ranged from 0.529 for C13 to 0.794 for C6, with an overall mean H_e of 0.631. Mean observed heterozygosity (H_o) within groups ranged from 0.422 in C7 to 0.721 for C13, with an overall mean H_o of 0.585 (Table 3). The percentage of polymorphic loci within groups ranged from 80% in C4 to 100% in C1, C3, C12, and C13. Cluster groups 5, 6, 7, 8, and 10 had positive fixation indices, suggesting a deficiency of heterozygotes. In contrast, the ssp. *cuspidata* group (C13) shows an excess of heterozygotes. The remaining clusters and overall mean heterozygosity across all groups approached panmixia.

Table 3 Within-cluster genetic variability at fifteen loci (standard errors in parentheses)

Cluster	N	A[a]	PI[b]	Mean heterozygosity		F[d]
				H_e(unb)	H_o(d.c)[c]	
C1	9.0	5.0(0.4)	100.0	0.689(0.063)	0.701(0.055)	−0.017
C10	12.9	3.9(0.4)	93.3	0.640(0.068)	0.575(0.057)	0.102
C11	19.9	4.5(0.5)	93.3	0.619(0.076)	0.585(0.056)	0.055
C12	2.7	2.7(0.2)	100.0	0.711(0.062)	0.644(0.051)	0.094
C2	4.0	3.2(0.3)	93.3	0.583(0.093)	0.619(0.065)	−0.062
C3	12.9	4.1(0.4)	100.0	0.617(0.074)	0.605(0.035)	0.019
C4	2.0	2.1(0.2)	80.0	0.533(0.091)	0.500(0.078)	0.062
C5	11.9	4.5(0.4)	86.7	0.655(0.088)	0.577(0.064)	0.119
C6	11.0	3.3(0.3)	86.7	0.794(0.088)	0.586(0.065)	0.262
C7	9.0	2.5(0.3)	86.7	0.593(0.119)	0.422(0.061)	0.288
C8	9.0	2.9(0.3)	86.7	0.615(0.112)	0.476(0.070)	0.226
C9	3.0	2.7(0.3)	86.7	0.622(0.107)	0.591(0.077)	0.050
cuspidata (C13)	7.0	5.3(0.6)	100.0	0.529(0.067)	0.721(0.053)	−0.363
Mean	8.8	3.6	91.8	0.631	0.585	0.064

[a] Mean number of alleles per locus

[b] A locus is considered polymorphic if the frequency of the most common allele does not exceed .95

[c] H_e(unb) = unbiased estimate of expected heterozygosity, and H_o(d.c) = direct-count observed heterozygosity (Nei 1978)

[d] Wright's fixation index = $(H_e - H_o)/H_e$

Table 4 Measures of gene diversity and genetic differentiation (Nei 1973) in olive

Locus	Gene diversity			Genetic differentiation	
	H_T	H_S	D_{ST}	Among clusters G_{ST}	Within clusters H_S/H_T
EM03	0.755	0.582	0.173	0.229	0.771
EM030	0.717	0.520	0.198	0.276	0.724
EM090	0.639	0.515	0.123	0.193	0.807
EMO88	0.763	0.561	0.202	0.265	0.735
GAPU89	0.767	0.650	0.116	0.152	0.848
IAS-oli16	0.812	0.659	0.152	0.188	0.812
IAS-pOe12_A	0.367	0.268	0.099	0.270	0.730
IAS-pOe12_B	0.510	0.323	0.187	0.367	0.633
ssrOeUA-DCA11	0.775	0.563	0.212	0.273	0.727
ssrOeUA-DCA3	0.847	0.643	0.204	0.240	0.760
ssrOeUA-DCA8	0.789	0.628	0.161	0.204	0.796
UDO99-011	0.837	0.670	0.166	0.199	0.801
UDO99-019	0.319	0.220	0.099	0.310	0.690
UDO99-028	0.800	0.537	0.263	0.329	0.671
UDO99-042	0.748	0.585	0.163	0.218	0.782
Mean	0.696	0.528	0.168	0.241	0.759

Total gene diversity (H_T), diversity within clusters (H_S), diversity among clusters ($D_{ST} = H_T - H_S$), proportion of diversity among clusters ($G_{ST} = D_{ST}/H_T$), and proportion of diversity within clusters (H_S/H_T)

Table 4 shows genetic differentiation measures within and among olive clusters identified in the UPGMA CA. The different loci varied in the distribution of components of gene diversity. The total gene diversity (H_T) per locus ranged from 0.319 at UDO099-019 to 0.847 at ssrOeUA-DCA3, with a mean H_T of 0.696.

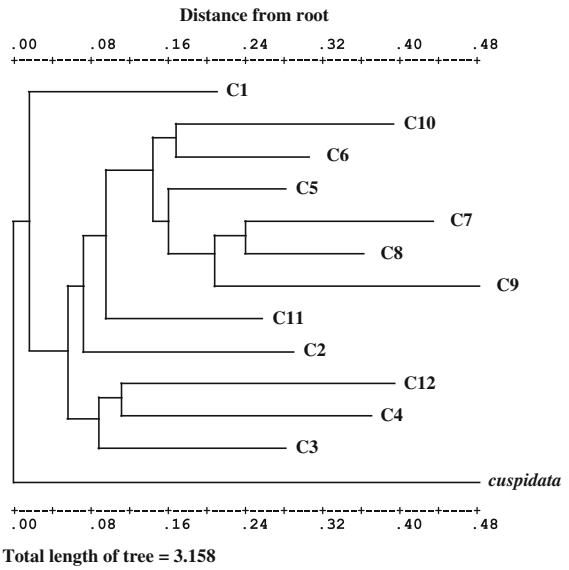

Fig. 3 The distance-Wagner tree of the differentiation of the thirteen UPGMA clusters of olive accessions, cophenetic correlation = 0.978

The locus-wise gene diversity within groups (H_S) ranged from 0.220 at UDO099-019 to 0.670 at UDO099-011, with an overall mean of 0.528. The proportion of diversity partitioned within groups (H_S/H_T) ranged from 63.3% at IAS-pOe12_B to 84.8% at GAPU89, with a mean of 75.9%. The proportion of gene diversity partitioned among groups (G_{ST}) ranged from 15.2% at GAPU89 to 36.7% at IAS-pOe12_B, with a mean G_{ST} of 24.1%.

Figure 3 shows a distance-Wagner tree, using the Prevosti distance measure, produced for populations based on the thirteen clusters provided by the UPGMA. The ssp. *cuspidata* cluster again forms an outlier group, and the cultivated olives can be further divided into three main groups. The French and Greek oil cultivar cluster, C1, forms a distinct group. The second group comprises the majority of clusters, including cultivar accessions of various origins and use types. However, two subgroups are formed, one with most of the Iberian and Italian origin cultivars, and the other with two clusters dominated by Eastern Mediterranean (primarily Greek and Tunisian) cultivars. The third group contains a Spanish oil cluster (C12), an Italian oil and Greek dual purpose cultivar cluster (C3), and a pair comprised of a western Asian cultivar of unknown use and the Italian table cultivar 'San Francesco' (C4).

Discussion

Genetic diversity and cultivar identity

The generally high heterozygosity per locus found within the NCGR olive collection is typical in naturally out-crossing, perennial, and clonally propagated species that are highly selected for greater adaptability, vigor, and productivity under cultivation and harbor recessive somatic mutations as heterozygotes (Aradhya et al. 2003). The allelic diversity results generally agree with recent olive SSR studies (Lopes et al. 2004; Belaj et al. 2004; de la Rosa et al. 2002; Sefc et al. 2000; Cipriani et al. 2002; Omrani-Sabbaghi et al. 2007). However, the observed heterozygosity for locus EM030 is lower than found in a previous study (de la Rosa et al. 2002). This may be due to the differences in the selection regime causing subtle differences in the genetic structure and differentiation patterns among the germplasm used. The mean heterozygosity across all loci and among groups approaches HW expectations. This indicates general panmixia within cultivated olive. However, the significant deficiency of heterozygotes found at six loci suggests a difference in selection pressures and responses among loci.

Polyclonal cultivars have been documented in olive. For example, 'Cordovil', from Portugal, may have arisen from accidental seedling propagation and local selection (Lopes et al. 2004). However, it is highly unlikely that differences at one or two microsatellite loci could be the result of sexual reproduction in a highly heterozygous, out-crossing species such as olive (Cipriani et al. 2002). Another possible cause of this type of clonal variation is a somatic mutation event occurring as bud sports in the mother cultivar tree, and subsequent clonal propagation (Cipriani et al. 2002). This may explain the six instances of slight clonal variation found in this study. However, this variation occurs between different accessions not known to be synonyms or closely related–the sole exception is 'Mission Variant' with 'Mission'. These observations suggest misidentification of at least one accession. For example, 'Zitoum' is putatively a synonym of 'Picholine marocaine', a unique cultivar from Morocco (Bartolini et al. 2005), yet the results suggest that the NCGR accession labeled 'Zitoum' is actually an incorrectly named, clonal variant of 'Mission'.

The cases of synonymy or near synonymy encountered in this study may be the result of accession mislabeling, however, the problem may also lie in the assumption that an olive cultivar is a homogeneous entity. Many olive cultivars have a long history of cultivation in the regions where they were selected by local farmers (Breton et al. 2006). Traditional farmers, in the past and still today, develop and name crop cultivars based on distinct phenotypes associated with specific agronomic, morphological, phenological and organoleptic qualities (Gibson 2009), some of which may be achieved via different genotypes, potentially resulting in heterogeneous cultivars. For example, in cassava, generally assumed to be clonally propagated, a study of farmer management in Guyana (Elias et al. 2001) found intravarietal polymorphisms in 72% of varieties analyzed. In the case of olive, the issue is further complicated by: both clonal and seed propagation, changing selection pressures, exposure to new growing environments, gene flow with other cultivars or locally occurring oleasters (*O. europaea* L. ssp. *europaea* var. *sylvestris* (Miller) Lehr.), all of which may result in intracultivar heterogeneity. Additionally, phenotypic or genetic changes may occur, triggering a new name for an extant genotype, or a new genotype for an established name. Finally, germplasm collections may themselves be sources of identification confusions contributing to intracultivar heterogeneity (Lopes et al. 2004).

A good example is the case of 'Oblonga' in the USDA collection. That cultivar was first identified in 1940 as an exceptionally *Verticillium*-resistant volunteer seedling in a Corning, California orchard (Hartmann et al. 1971). Based on RAPD and SSR marker data, however, that accession is potentially a synonym of 'Frantoio', a traditional Italian oil cultivar (Barranco et al. 2000). The current study corroborates those findings. 'Oblonga' and 'Frantoio' were synonymous at all fifteen SSR loci.

'Mission', a widely grown mainstay of the California olive industry, was first grown in California by Spanish Franciscan missionaries in the eighteenth Century, but its precise origin remains unclear (Taylor 2000). It shares the most allelic diversity in the UPGMA cluster analysis with several other ancient and well-established Iberian cultivars: 'Nevadillo', 'Manzanillo', 'Cordovil', 'Picual', 'Verdeal', and one genotype synonymous among several South American cultivar names ('Azapa', 'Yuaca', 'Mission Leiva', and 'Mostazal'). This supports a Spanish origin of 'Mission'. A comparison of the 'Mission' genotype to the above mentioned cultivars and 'Sevillano', another widely grown Spanish cultivar, revealed that all could contribute at least one allele to the 'Mission' genotype at thirteen or fourteen SSR loci, and at all fifteen loci for 'Sevillano' (Table 5), suggesting that 'Mission' could have originated as a seedling selection of the Iberian germplasm imported to the New World by the Spanish missionaries. Another possibility in the cases of both 'Zitoum' and 'Mission' accessions is that these are catch-all names that may have included distinct cultivars that at one time were subsumed under those names for ease or because of broad shared characteristics, as has been observed for *Phaseolus* species and cultivars among farmers in southern Mexico (Soleri et al. *forthcoming*). The leading international database for olive cultivars includes five entries for "*Zeitoun*" (http://www.oleadb.it/scheda.php?name=zeitoun&Submit2=submit). Another Californian cultivar, 'Redding Picholine', is thought to have been imported into California in the late nineteenth century as a rootstock variety initially mistaken for the French 'Picholine' scion cultivar (Taylor 2000). The UPGMA cluster analysis places 'Redding Picholine' into the most basal cultivated olive clade (C1) and is most closely associated with the French cultivar, 'Rubra', suggesting that the former cultivar may derive from French germplasm.

The NCGR olive collection is a useful source of genetically diverse material for use by olive breeders, researchers, growers, and backyard hobbyists to help maintain and improve the olive crop in California. Therefore, it is vital for the genetic identity of olive germplasm to be verified before distribution to these users. Instances of misidentification or ambiguous identity, as with the 'Zitoum' example above, highlight the need for a standardized set of molecular markers used to build a database of verified olive cultivar genotypes (Doveri et al. 2008; Baldoni et al. 2009), a set harmonized with researchers in the Mediterranean and worldwide to ensure robust, globally accurate identities.

Table 5 Comparison of the 'Mission' genotype to that of several Iberian and one South American accession(s) at fifteen SSR loci

Accession	Country of origin	ssrOeUA-DCA11	EM030	GAPU89	UDO99-011	EM090	UDO99 -028	EM03	IAS-oli16	IAS-pOe12_A	IAS-pOe12_B	ssrOeUA-DCA3	EM088	UDO99–019	ssrOeUA-DCA8	UDO99–042
Mission	US (CA.)	141/161	187/187	158/205	121/127	184/186	147/149	211/213	156/164	114/114	121/121	238/252	184/184	129/129	136/138	146/146
Azapa	Chile	141/141	185/187	158/205	111/113	184/184	147/149	209/213	150/154	114/114	121/121	238/248	184/184	129/129	136/140	140/148
Cordovil	Portugal	141/141	185/187	158/175	111/113	184/186	123/149	211/213	154/164	114/114	121/121	238/248	184/184	129/129	136/138	146/146
Manzanillo	Spain	141/161	189/189	158/205	113/127	184/186	147/149	209/213	154/164	114/114	121/123	244/252	186/186	129/129	136/138	140/146
Nevadillo	Spain	141/141	185/187	158/205	113/125	184/184	123/149	209/213	151/156	114/114	121/123	244/248	184/184	129/129	136/140	140/148
Picual	Spain	141/141	185/187	158/175	111/113	184/184	147/149	209/217	151/156	114/114	121/123	238/248	186/186	129/129	136/140	146/146
Sevillano	Spain	161/161	187/189	175/205	113/121	184/184	123/147	209/211	150/164	114/114	121/121	248/252	184/186	129/129	136/136	140/146
Verdeal	Portugal	141/161	185/187	158/175	111/121	184/186	123/131	209/213	154/164	114/114	121/121	238/248	186/186	129/129	136/138	148/148

Genetic structure and differentiation within and among groups

The US *ex situ* germplasm for cultivated olive, represented by the accessions in the NCGR collection, appears to be relatively variable as reflected in the highly dissected nature of the UPGMA phenogram, low level of differentiation among groups, low proportions of variation explained by the first two principal components, and overlapping groups within the PCA. Such low differentiation might be expected from free gene flow among the highly outcrossing olive cultivars and millennia of human exchange of germplasm throughout the Mediterranean (Angiolillo et al. 1999). However, the greater structure and intercultivar variation encountered in surveys of Mediterranean germplasm (Cipriani et al. 2002; Sarri et al. 2006) suggest that our observations may reflect one or more of the following: (a) narrow genetic diversity in a crop species outside its center of origin and diversity; (b) a founder effect in a nonnative and often clonally propagated species; (c) a relatively short and geographically limited experience with olive cultivation; (d) poor representation in the germplasm collection of regionally present genotypes. For example, a small recent study (128 individuals from five different plantings) of *in situ* historic olive trees in central coastal California using the same SSR markers as reported here found four different cultivar genotypes not represented in the USDA accessions (Soleri et al. 2010).

The moderate to high level of gene diversity found overall per locus is typical of outcrossing, clonally propagated crops such as olive (Sefc et al. 2000). Most of this gene diversity is apportioned among cultivars within groups. However, the structure of the distance-Wagner analysis indicates some significant differences in the frequency and composition of alleles among olive groups.

Some Eastern Mediterranean cultivars in the NCGR collection showed higher levels of differentiation than other accessions in the UPGMA (C1, C11). This may support archeological (Zohary and Hopf 2000) and mitotype (Besnard et al. 2001; Breton et al. 2006) evidence of an Eastern Mediterranean center of origin of olive domestication and east to west dispersal pattern. The differentiation of certain oil cultivars, such as the French and Greek oil cultivars in cluster 1 of the UPGMA, suggests

multilocal, strong selection pressure for oil-producing traits. Notably, cluster 1 also forms a separate clade among UPGMA clusters using the distance-Wagner method. Multilocal selection of cultivars was likewise reported using RAPD (Besnard et al. 2001) and SSR (Belaj et al. 2004) data.

Acknowledgments This study was funded by the USDA-ARS Project CRIS 5306-21000-016-001. The research presented in this paper was part of the Master of Science thesis of the senior author in the Horticulture and Agronomy Graduate Group at the University of California, Davis.

References

Genbank Acc. No. AF356039 (2001) http://www.ncbi.nlm.nih.gov/nuccore/19070686. Accessed 09 Feb 2010

Angiolillo A, Mencuccini M, Baldoni L (1999) Olive genetic diversity assessed using amplified fragment length polymorphisms. Theor Appl Genet 98(3–4):411–421

Aradhya MK, Dangl GS, Prins BH, Boursiquot J-M, Walker MA, Meredith CP, Simon CJ (2003) Genetic structure and differentiation in cultivated grape, *Vitis vinifera* L. Genet Res 81(3):179–192

Baldoni L, Cultrera N, Mariotti R, Ricciolini C, Arcioni S, Vendramin G (2009) A consensus list of microsatellite markers for olive genotyping. Mol Breed 24(3):213–231

Barranco D, Trujillo I, Rallo P (2000) Are 'Oblonga' and 'Frantoio' Olives the same cultivar? HortScience 35(7): 1323–1325

Bartolini G, Prevost G, Messeri C, Carignani G (2005) The olive germplasm database. FAO Seed and Plant Genetic Resources Service. http://apps3.fao.org/wiews/olive/oliv.jsp. Accessed 17 March 2009

Belaj A, Satovic Z, Cipriani G, Baldoni L, Testolin R, Rallo L, Trujillo I (2003) Comparative study of the discriminating capacity of RAPD, AFLP and SSR markers and of their effectiveness in establishing genetic relationships in olive. Theor Appl Genet 107(4):736–744

Belaj A, Cipriani G, Testolin R, Rallo L, Trujillo I (2004) Characterization and identification of the main Spanish and Italian olive cultivars by simple-sequence-repeat markers. HortScience 39(7):1557–1561

Besnard G, Baradat P, Berville A (2001) Genetic relationships in the olive (*Olea europaea* L.) reflect multilocal selection of cultivars. Theor Appl Genet 102(2-3):251–258

Breton C, Besnard G, Berville AA (2006) Using multiple types of molecular markers to understand olive phylogeography. In: Zeder MA, Bradley DG, Emshwiller E, Smith BD (eds) Documenting domestication: new genetic and archeological paradigms. University of California Press, Berkeley, pp 143–152

Carriero F, Fontanazza GF, Cellini F, Giorio G (2002) Identification of simple sequence repeats (SSRs) in olive (*Olea europaea* L.). Theor Appl Genet 104:301–307

Cipriani G, Marrazzo MT, Marconi R, Cimato A, Testolin R (2002) Microsatellite markers isolated in olive (*Olea europaea* L.) are suitable for individual fingerprinting and reveal polymorphism within ancient cultivars. Theor Appl Genet 104:223–228

De la Rosa R, James CM, Tobutt KR (2002) Isolation and characterization of polymorphic microsatellites in olive (*Olea europaea* L.) and their transferability to other genera in the Oleaceae. Mol Ecol Notes 2:265–267

Dice LR (1945) Measurement of the amount of ecological association between species. Ecology 26:295–302

Doveri S, Gil FS, Diaz A, Reale S, Busconi M, Machado AD, Martin A, Fogher C, Donini P, Lee D (2008) Standardization of a set of microsatellite markers for use in cultivar identification studies in olive (*Olea europaea* L.). Sci Hortic (Amsterdam, Neth.) 116(4):367–373

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

Elias M, Penet L, Vindry P, McKey D, Panaud O, Robert T (2001) Unmanaged sexual reproduction and the dynamics of genetic diversity of a vegetatively propagated crop plant, cassava (*Manihot esculenta* Crantz), in a traditional farming system. Mol Ecol 10:1895–1907

Gibson RW (2009) A review of perceptual distinctiveness in landraces including an analysis of how its roles have been overlooked in plant breeding for low-input farming systems. Econ Bot 63:242–255

Green PS (2002) A revision of *Olea* L. (Oleaceae). Kew Bull 57(1):91–140

Hartmann HT, Schnathorst WC, Whisler JE (1971) 'Oblonga', a clonal olive rootstock resistant to Verticillium wilt. Calif Agr 25(6):12–15

Lopes MS, Mendonca D, Sefc KM, Gil FS, Machado AD (2004) Genetic evidence of intra-cultivar variability within Iberian olive cultivars. HortScience 39(7):1562–1565

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci 70:3321–3323

Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics 89(3):583–590

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci 76:5269–5273

Omrani-Sabbaghi A, Shahriari M, Falahati-Anbaran M, Mohammadi SA, Nankali A, Mardi M, Ghareyazie B (2007) Microsatellite markers based assessment of genetic diversity in Iranian olive (*Olea europaea* L.) collections. Sci Hortic (Amsterdam, Neth.) 112(4):439–447

Ota T (1993) DISPAN: genetic distance and phylogenetic analysis. Pennsylvania State University. Available at http://www.bio.psu.edu/People/Faculty/Nei/Lab/software.htm. Accessed 07 May 2009

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

Rohlf FJ (2002) NTSYSpc: Numerical Taxonomy System, ver. 2.1.1. Exeter Publishing, Ltd, Setauket

Sadiki M, Jarvis D, Rijal DK, Bajracharya J, Hue NN, Camacho-Villa T, Burgos-May LA (2006) Variety names: an entry point to crop genetic diversity and distribution in agroecosystems? In: Jarvis D, Padoch C, Cooper D (eds) Managing biodiversity in agricultural ecosystems. Columbia University Press, New York, pp 34–76

Sanz-Cortes F, Parfitt DE, Romero C, Struss D, Llacer G, Badenes ML (2003) Intraspecific olive diversity assessed with AFLP. Plant Breed 122(2):173–177

Sarri V, Baldoni L, Porceddu A, Cultrera NGM, Contento A, Frediani M, Belaj A, Trujillo I, Cionini PG (2006) Microsatellite markers are powerful tools for discriminating among olive cultivars and assigning them to defined populations. Genome 49:1606–1615

Sefc KM, Lopes MS, Mendonca D, Rodrigues Dos Santos M, Laimer Da Camara Machado M, Da Camara Machado A (2000) Identification of microsatellite loci in olive (*Olea europaea*) and their characterization in Italian and Iberian olive trees. Mol Ecol 9:1171–1193

Soleri D, Koehmstedt A, Aradhya M, Polito V, Pinney K (2010) Comparing the historic olive trees (*Olea europaea* L.) of Santa Cruz Island with contemporaneous trees in the Santa Barbara, CA area: a case study of diversity and structure in an introduced agricultural species conserved in situ. Genet Resour Crop Evol 57:973–984

Sweeney S (2003) NOVA—the National Olive Variety Assessment Project. RIRDC Publication No 03/054. Rural Industries Research and Development Corporation, Canberra, Australia. http://www.rirdc.gov.au/reports/NPP/03-054sum.html. Accessed 16 April 2009

Swofford DL (1981) On the utility of the distance Wagner procedure. In: Funk VA, Brooks DR (eds) Advances in Cladistics: proceedings of the first meeting of the Willie Hennig society. New York Botanical Garden, Bronx, pp 25–43

Swofford DL (2003) PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts

Swofford D, Selander RB (1989) BIOSYS-1 A computer program for the analysis of allele variation in population genetics and biochemical systematics. Release 1.7. Illinois Natural History Survey, Urbana, Illinois

Taylor JM (2000) The olive in California: history of an immigrant tree. Ten Speed Press, Berkeley, California

Trujillo I, Rallo L, Arus P (1995) Identifying olive cultivars by isozyme analysis. J Am Soc Hortic Sci 120(2):318–324

U.S. Department of Agriculture (2008) Fruit and Tree Nuts Situation and Outlook Yearbook. Market and Trade Economics Division, Economic Research Service, U.S. Department of Agriculture

Vossen P (2007) Olive oil: history, production, and characteristics of the world's classic oils. HortScience 42(5):1093–1100

Vossen P, Devarenne A (2005) California olive oil industry survey statistics 2004. University of California Cooperative Extension, Sonoma County. http://cesonoma.ucdavis.edu/files/27314.pdf. Accessed 09 Aug 2010

Wright S (1978) Evolution and the genetics of populations, volume 4, variability within and among natural populations. University of Chicago Press, Chicago

Zohary D, Hopf M (2000) Domestication of plants in the old world: the origin and spread of cultivated plants in west Asia, Europe, and the Nile, Valley. Oxford University Press, New York