Plant Mol Biol Rep (2017) 35:575–585
DOI 10.1007/s11105-017-1046-y

ORIGINAL PAPER

Cultivated and Wild Olives in Crete, Greece—Genetic Diversity and Relationships with Major Turkish Cultivars Revealed by SSR Markers

M. Aksehirli-Pakyurek[1,2] · G. C. Koubouris[3] · P. V. Petrakis[4] · S. Hepaksoy[5] · I. T. Metzidakis[3] · E. Yalcinkaya[6] · A. G. Doulis[2]

Published online: 24 August 2017

Abstract The genetic relationships between and within some traditionally grown cultivars of olive tree (*Olea europaea* L.) in Greece (island of Crete) and in Turkey were investigated. Cultivars from Crete included 'Koroneiki', 'Throubolia' and 'Mastoidis', while those from Turkey included 'Samanli' and 'Gemlik'. Cultivars were represented by multiple genotypes of aged trees collected from the field, each one complying with established descriptors. Representative genotypes of wild olive trees from Crete were also employed. A total of 112 genotypes were analysed, employing seven microsatellite (SSR) loci yielding a total of 81 alleles, and reaching a cumulative probability of identity of 6.73×10^{-09} with a mean observed heterozygosity of 0.852. Analysis of molecular variance significantly partitioned genetic diversity between and within cultivars, albeit with no appreciable difference between the two levels of diversity. All cultivar genotypes aggregated along single, cultivar-specific clusters, pointing to human-driven selection. The two Turkish cultivars 'Samanli' and 'Gemlik' were grouped together. The Cretan cultivar 'Throubolia' grouped together with the two Turkish cultivars, indicating germplasm movement across the Aegean Sea during historical times. Some gene flow was observed between the Cretan cultivars and the native wild populations (likely feral forms). SSR alleles were ranked for their efficiency in discriminating the examined materials, thus establishing a molecular key for cultivar identification. An identification process is proposed including a classification binary tree and provided a method for sorting any new unknown material purportedly originating from any of the analysed cultivars.

Keywords Classification binary tree · Cluster · Cultivar · Germplasm movement · Wild olives

✉ A. G. Doulis
andreas.doulis@nagref-her.gr

[1] Present address: Faculty of Agriculture, Department of Horticulture, University of Siirt, 56100 Siirt, Turkey

[2] Institute of Olive Tree, Subtropical Plants and Viticulture, Hellenic Agricultural Organization "Demeter" (ex. NAGREF), Laboratory of Plant Biotechnology & Genomic Resources, Kastorias 32A, 71307 Heraklion, Greece

[3] Hellenic Agricultural Organization (H.A.O.) "Demeter" (ex. N.AG.RE.F.), Laboratory of Olive Cultivation, Institute of Olive Tree, Subtropical Plants and Viticulture, Agrokipion, 73100 Chania, Greece

[4] Hellenic Agricultural Organization "Demeter" (ex. NAGREF), Laboratory of Forest Entomology, Institute of Mediterranean Forest Ecosystems and Forest Products Technology, 115 28 Athens, Greece

[5] Faculty of Agriculture, Department of Horticulture, University of Ege, 35100 Izmir, Turkey

[6] Turkish Agriculture Insurance Pool Management Company, Technical Department, 34768 Istanbul, Turkey

Introduction

The olive (*Olea europaea* L.) is one of the major tree crops in the Mediterranean Basin, where more than 90% of the world's olive oil production is realized (IOC 2016). Olive cultivation has a history of approximately 5000–6000 years (Vossen 2007) with intense presence in nutrition, culture and commerce of many different civilizations throughout history. The dominant theory asserts that the first olive domestication occurred in the Middle East. However, other scenarios of local domestication events, as well as admixture shaping olive germplasm, have also been proposed (Diez et al. 2015). Based on the post-glacial history of oleaster populations, Greece is considered as one of the main regions of primary domestication and secondary diversification of the olive tree

(Besnard and Rubio de Casas 2016). Archaeological evidence reveals that the cultivated olive was initially domesticated from a Near Eastern and Aegean gene pool by Phoenicians and Greeks (Newton et al. 2014). Subsequently, olive was introduced in other Mediterranean areas such as Marseille, France, 2500 years ago (Breton et al. 2009), contributing to the evolution of existing olive germplasm in recent times. The long-standing occurrence of the olive in the Mediterranean is also supported by molecular phylogenetic analyses (Besnard et al. 2013).

Greece is the third largest olive oil-producing country with an average annual production of 320,000 t and over one million hectares of olive trees (IOC 2016). Turkey is the second largest table olive-producing country with an average annual production of 225,000 t and approximately 800,000 ha of olive trees (IOC 2016). Crete is considered among the first areas of olive cultivation in ancient times (Lanza 2011). Indeed, the trade of olive products contributed to the wealth of the Minoan Civilization approximately 3000 BC. Currently, more than 30 million trees are cultivated on the island, the vast majority of which are cvs. 'Koroneiki', 'Mastoidis' and 'Throubolia'.

The genetic characterization of olive germplasm is a fundamental task for establishing the origin of the currently grown cultivars, for selecting cultivars with important agronomic traits, and for developing new cultivars customized to modern plantation systems or adapted to adverse environmental conditions. Additionally, it is a prerequisite for the certification of high-quality true-to-type propagation material. Morphological descriptors and molecular markers have been the main tools used for this purpose. Indeed, microsatellites, or simple sequence repeats (SSRs), are the most widely used molecular markers for the identification of cultivars in gene banks (Trujillo et al. 2013) as well as for establishing the relationships between cultivated and wild olives (Belaj et al. 2011). This is due to their higher reproducibility and non-anonymous and co-dominant nature, especially when compared to previously employed markers such as random amplified polymorphic DNA (RAPD) and amplified fragment length polymorphism (AFLP; discussed in Belaj et al. 2003).

Greece and Turkey have a long history of interactions resulting in similarities in cultural and economic aspects of their modern societies, including agricultural production patterns and cultivars. Consequently, two major Turkish olive cultivars were included in the present study in order to investigate their potential relationship with Greek germplasm. For one of them, cv. 'Gemlik', Yoruk and Taskin (2014) indicated that it is not related to any cultivars from other regions in Turkey. In Greece, studies assessing genetic diversity of olive tree germplasm usually involve only one or a very small number of independent genotypes per cultivar derived from germplasm banks and focus on inter-cultivar rather than on intra-cultivar diversity (Linos et al. 2014; Xanthopoulou et al. 2014; Roubos et al. 2010; Owen et al. 2005; Terzopoulos et al. 2005; Hagidimitriou et al. 2005; Nikoloudakis et al. 2003). At present, the genetic constitution of traditionally employed autochthonous materials in olive orchards, especially before the government-directed introductions of few selected clones, is still critically poorly understood. Indeed, one of the aims of the present paper is to provide an initial genetic characterization of such autochthonous materials. This could, in turn, provide initial insights into the origin of the currently grown cultivars in Crete, Greece, as well as into possible exchanges of olive tree germplasm across the two sides of the Aegean Sea during historical times. Additionally, and to the best of our knowledge, there exist no data on the genetic characterization of wild olives in Greece. However, given the long history of olive growing in Greece, as well as the traditional germplasm selection practices applied, wild olives are expected to represent a valuable source of genetic variability. In fact, some important agronomic traits that are hardly found in olive cultivars have been observed in the wild germplasm (Colella et al. 2008; Mkize et al. 2008; Belaj et al. 2011).

In the present study in Crete, one of the centres of ancient olive cultivation, nuclear SSR markers were employed for phylogenetic inferences as well as for the evaluation of the degree of genetic variation (i) within and between the three major autochthonous cultivars grown in Crete, (ii) between the same cultivars and wild olive trees in Crete, and (iii) between the same cultivars and two autochthonous cultivars of major significance for olive growing in Turkey.

Materials and Methods

Plant Material

Five olive tree cultivars of high economic importance in Greece and in Turkey and wild individuals growing in Crete, Greece, were included in the present study (Table 1). Representative trees of the Greek cvs. 'Koroneiki', 'Mastoidis' and 'Throubolia' were identified and leaves were collected from the major traditional cultivation areas of each cultivar in the olive growing regions on the island of Crete. The focus of the present work lies in autochthonous germplasm. Consequently, only mature trees (i.e., over 40 years old) from old olive orchards were sampled based on morphological identification. This was done to avoid young plants massively distributed by nurseries during the recent two to three decades following clonal propagation of reference materials. The cv. 'Mastoidis' is mainly cultivated in mountainous areas of western Crete, cv. 'Throubolia' is grown in mountainous areas of central Crete and cv. 'Koroneiki' dominates agricultural lands across the island (Fig. 1). Morphological characteristics of individual trees (mostly of leaves, fruit and crown form) grouped within previously defined ranges for

Table 1 Olive cultivars and wild populations (AGR) analysed, regions of cultivation and uses

Cultivar epithet / population	No. of samples	Country of origin	Cultivation regions (within country of origin)	Sampling regions (within country of origin)	Use
'Koroneiki'	29	Greece	Crete, Peloponnese	Crete	Oil
'Throubolia'	24	Greece	Crete, Aegean Islands	Crete	Table
'Mastoidis'	23	Greece	Crete, Peloponnese	Crete	Double
AGR	17	Greece	–	Crete	–
'Gemlik'	11	Turkey	Marmara	Ismir	Table
'Samanli'	8	Turkey	Marmara	Ismir	Table

each cultivar according to the RESGEN project of the International Olive Council as presented in brief in the World Catalogue of Olive Varieties (2000). Reference genotypes for each cultivar from the National Germplasm Collection of Olive Tree of Greece at the Institute for Olive Tree, Subtropical Plants and Viticulture (HAO Demeter) were also included in the analysis. In addition, leaf samples from mature wild olive trees from Crete were collected and analysed in a way identical to cultivars. Wild olive trees were sampled in the vicinity of old olive orchards (those employed for cultivar characterization) on uncultivated lands with natural vegetation with free bush shape characterized by small endocarp and thin mesocarp. Similar to the Greek cultivars, samples from mature trees over 40 years old of the two major Turkish cvs. 'Gemlik' and 'Samanli' were selected from their main areas of cultivation located in the broader areas of Bursa and Yalova, Turkey.

DNA Extraction

Total genomic DNA was isolated from leaf material using the DNeasy Plant Mini kit (Qiagen, cat. no 69104) according to manufacturer's instructions. Initial grinding was conducted using the automated grinder TissueLyser (Retsch, Germany) in the presence of liquid nitrogen. DNA was quantified using the Hoechst 33258 fluorescence dye (Sigma, No. B2883) on a computerized TD 700 Fluorimeter (Turner Designs, Sunnyvale, CA, USA) against a series of calf thymus DNA standards (Sigma, No. D4764).

Microsatellite Assay

Seven microsatellite loci that have been used in the past for similar analyses of olive germplasm were selected on the basis of their informativeness (values of polymorphic information content; PIC), ease of scoring and observed heterozygosity. These were OeUA-DCA3, OeUA-DCA15 (Sefc et al. 2000), UDO99-019, UDO99-031 (Cipriani et al. 2002), IAS-Oli23, IAS-Oli26 and IAS-Oli27 (Díaz et al. 2006). Polymerase chain reactions were carried out in a 20-μl reaction in a Perkin Elmer 9600 thermocycler including 25 ng of template DNA, 0.2 mM of each dNTP, 0.2 μM of each primer, 2.5 mM $MgCl_2$ and 1 U of DiaTaqDD Polymerase (HyTest, Finland). Thermal cycling included initial denaturation for 5 min at 95 °C, followed by 35 cycles of 95 °C for 30 s, the corresponding annealing temperature for 45 s and 72 °C for 45 s, with a final extension at 72 °C for 10 min. Amplicons were denatured at 97 °C in the presence of formamide for 3 min and loaded on an 8% polyacrylamide gel automated sequencer IR2 (LICOR, NE, USA) at the Microchemistry Laboratory of the Foundation for Research and Technology (FO.R.T.H.), Heraklion, Crete, Greece. The LICOR proprietary software SAGA GT® was employed for IR2 gel characterization. Genotyping

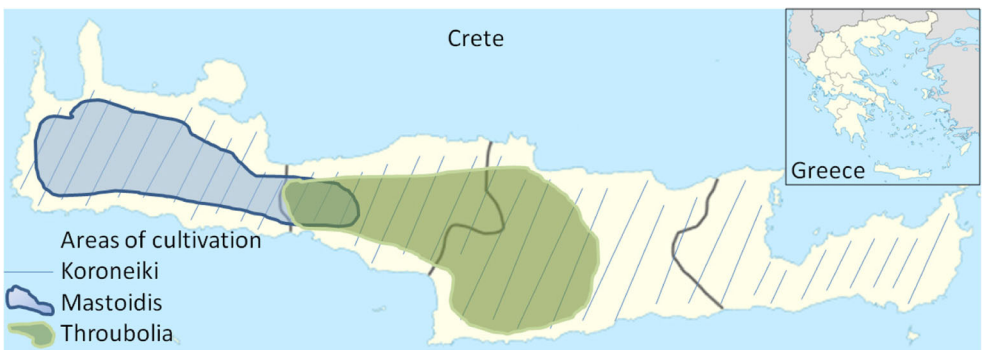

Fig. 1 Main areas of cultivation of olive cvs. 'Koroneiki', 'Mastoidis' and 'Throubolia' in the island of Crete, Greece

was initiated by employing loci with high discriminating power according to the existing literature and adding one locus at a time in a step-wise fashion. The optimal tally of loci (locus1 + locus2 + …) whereupon no further discrimination between genotypes could be attained was retained for subsequent genetic analyses.

Data Analysis

Within SAGA GT, SSR allele sizing and binning were manually performed or verified while additional measures were taken to ensure high accuracy and reproducibility of gel-based SSR data. Such measures involved the use of multiple ladder lanes within a single gel, use of common samples between gels and employing the "de-smiling" function of SAGA GT.

Per locus, allele sizing was based on published repeat patterns. Data matrices were produced, and, as a first step, genetic diversity measures were determined for each employed locus across all fingerprinted genotypes (n = 112) as shown in Table 2. These measures included (i) individual locus polymorphic information content (PIC; Botstein et al. 1980), (ii) observed heterozygosity (H_O) and (iii) expected heterozygosity (H_E). PIC, H_O and H_E were calculated using CERVUS ver. 3.0.3 (Kalinowski et al. 2007). Probability of identity (PI) was calculated using the software Identity 1.0 (Sefc et al. 1997). To test for the overall significance of variance partitioning within and between Cretan and Turkish olive tree cultivars (including the wild populations from Crete), an analysis of molecular variance (AMOVA) was performed using GenAlEx ver. 6.5 as a plug-in module within Microsoft Excel (Peakall and Smouse 2006). The number of alleles per locus (N_a), effective number of alleles (N_e) and Shannon information index (I) and inbreeding coefficient (F_{IS}) were also calculated using GenAlEx. A similarity matrix was produced employing the Lynch distance metric (Lynch 1990), which is a simple band-sharing measure [termed the "Band" coefficient within NTSYSpc ver. 2.21L (Rohlf 2008)]. Subsequently, a genetic similarity tree was constructed employing hierarchical clustering and the agglomerative un-weighted pair group method with arithmetic mean (UPGMA) algorithm (Sneath and Sokal 1973) in NTSYSpc ver. 2.21L (Rohlf 2008).

Classification Binary Tree Analysis

To reveal the relationships among unique genotypes, the data set was trimmed to include all molecularly differentiated individuals by eliminating all but one identical individual. The resulting matrix of 68 genotypes by seven SSR loci was then subjected to classification binary tree (CBT) analysis (Breiman et al. 1984). Similar to the genetic analyses, the CBT input dataset consisted of SSR loci allelic sizes. The output of CBT, called a mobile, initially entails the split of the original sample set into two parts on the basis of a criterion involving one or two discriminatory loci in a simple algebraic expression. All polymorphic loci retained in the construction of a mobile comprise the best diagnostic variable set for the groups already defined prior to the analysis. The groups at the beginning of a CBT construction were the five olive tree cultivars and one wild olive population. The tree of a mobile grows according to splits that produce maximally informative and 'pure' groups according to an impurity function (Breiman et al. 1984). The reduction of error in the classification is monitored by means of an information-theory-compatible Gini-Simpson index in the form of a 'towing' loss function, using the towing coefficient originally proposed by Breiman et al. (1984). CBT analysis was performed using routines and packages within the R environment (R Development Core Team 2008) and used the packages 'rpart' (Therneau and Atkinson 1997; R package version 3.1-52, http://mayoresearch.mayo.edu/mayo/research/biostat/

Table 2 Size ranges, number of observed alleles (N_a), observed (H_O), expected heterozygosity (H_E), polymorphic information content (PIC), probability of null alleles (r) and probability of identity (PI) for seven microsatellite loci used in 112 olive accessions

Locus	Size range	N_a	H_O	H_E	PIC	r	PI
OeUA-DCA3	220–254	13	0.732	0.757	0.714	+ 0.0026	0.1007
OeUA-DCA15	242–264	6	0.679	0.779	0.741	+ 0.0752	0.0844
UDO99-019	101–163	8	0.981	0.637	0.575	− 0.2580	0.1932
UDO99-031	109–157	11	0.674	0.806	0.777	+ 0.0979	0.0641
IAS-Oli23	211–243	13	0.991	0.831	0.806	− 0.0961	0.0515
IAS-Oli26	169–197	11	0.979	0.819	0.790	− 0.0980	0.0586
IAS-Oli27	101–135	19	0.927	0.894	0.881	− 0.0210	0.0212
Mean		11.6	0.852	0.789	0.755		
Min.		6	0.674	0.637	0.575		
Max.		19	0.991	0.894	0.881		
Total		81					6.73×10^{-09}

splusfunctions.cfm) and 'vegan' (Dixon 2003) as implemented by R package version 2.0-3 (http://CRAN.R-project.org/package=vegan). Some algorithms and graphics were applied using SYSTAT.

Results

Overall Microsatellite Diversity

One hundred twelve independent genotypes were fingerprinted using seven SSR loci. Size ranges, number of detected alleles (N_a), observed (H_O) and expected heterozygosity (H_E), polymorphic information content (PIC), probability of null alleles (*r*) and probability of identity (PI) for seven microsatellite loci used in 112 olive genotypes (cultivars as well as wild olive populations from Crete) are presented in Table 2. All seven SSR loci employed herein were highly polymorphic, yielding a total of 81 alleles, and were able to uniquely distinguish 68 out of the 112 genotypes analysed (the former number representing the number of distinct SSR profiles produced in the framework of the present study). Further, all loci were characterized as highly informative since PIC > 0.5.

The number of alleles produced per locus ranged from 6 (OeUA-DCA15) to 19 (IAS-Oli27), with a mean value of 11.6 alleles per locus (Table 2). The overall H_O ranged from 0.674 (UDO99-031) to 0.991 (IAS-Oli23), with an average of 0.852 (Table 2). Expected heterozygosity (H_E) ranged from 0.637 (UDO99-019) to 0.894 (IAS-Oli27), with a mean value of 0.789 (Table 2). Furthermore, the polymorphic information content of each locus (PIC) ranged from 0.575 (UDO99-019) to 0.881 (IAS-Oli27), with a mean value of 0.755. Both H_E and PI values indicated that IAS-Oli27 was the most informative and UDO99-019 was the least informative marker. The probability of null alleles (*r*) had a mean value of − 0.0425 and ranged from + 0.0979 (UDO99-031) to − 0.258 (UDO99-019), while two loci (OeUA-DCA15 and UDO99-031) exhibited an $r > 0.05$ (Table 2). The overall probability of identity (PI) was 6.73×10^{-09}.

When cultivars and wild olive genotypes were considered, 81 alleles were found. When only cultivar genotypes were considered, the number of detected alleles decreased to 58 (data not shown), yielding an average value of 8.3 detected alleles per locus. This represents 23 fewer alleles (28.4%) compared to a combined pool of all cultivars and wild genotypes. Detected private alleles and their sizes are presented in Table 3. A substantial number of private alleles was detected, especially in wild olives (23), followed by Greek and Turkish cultivar genotypes with eight and three private alleles, respectively.

Table 3 Private allele lengths per SSR locus and per cultivar and wild olive tree group (AGR)

Cultivar/population	Locus	Allele length (bp)
'Koroneiki'	OeUA-DCA3	226, 232
	UDO099-019	163
	IAS-Oli27	101, 117
'Throubolia'	IAS-Oli23	225
'Mastoidis'	UDO099-019	151
	IAS-Oli27	107
AGR	OeUA-DCA3	220, 244, 254
	OeUA-DCA15	244
	UDO99-019	141, 147
	UDO99-031	109, 113, 129, 139, 145
	IAS-Oli23	211, 215, 241, 243
	IAS-Oli26	169, 175, 185, 191, 193
	IAS-Oli27	122, 130, 133
'Gemlik'	OeUA-DCA3	248
	IAS-Oli26	197
'Samanli'	OeUA-DCA3	228

Genetic Differentiation and Relationships Between Wild and Cultivated Olives From Greece and Turkey

The distance metric of Lynch (1990) was utilized for clustering of the analysed genotypes (Fig. 2). All five investigated cultivars ('Koroneiki', 'Throubolia', 'Mastoidis', 'Gemlik' and 'Samanli') clearly separated into cultivar-specific groups. Interestingly, the wild olive genotypes did not cluster into a single group but rather as three sub-groups, each joining one of the local cultivar groups of Crete ('Koroneiki', 'Mastoidis' or 'Throubolia'). In fact, the Cretan cv. 'Throubolia' forms one of the two major sub-clusters together with the non-Cretan cvs. 'Gemlik' and 'Samanli'. The other major sub-cluster includes the Cretan cvs. 'Koroneiki' and 'Mastoidis'. For the Cretan cultivars, the average number of alleles per locus was 4.57 for 'Throubolia', 4.14 for 'Koroneiki' and 3.43 for 'Mastoidis'. A higher average number of alleles were detected in wild olives (9.14), and lower values were found in the two Turkish cvs. 'Gemlik' and 'Samanli' (2.86) (Table 4). The average number of effective alleles (N_e) was similar for the Greek cultivars (2.22–2.55) and was higher in wild olives (6.09) and lower in cultivars from Turkey (2.27–2.45). The Shannon information index (I) was higher in wild olives (1.91), followed by Cretan (0.92–0.98) and Turkish cultivars (0.84–0.85). In general, higher values were obtained for observed (H_O = 0.789–1) than for expected (H_E = 0.498–0.809) heterozygosity. For observed heterozygosity, cvs. 'Mastoidis' and 'Gemlik' exhibited the highest (H_O = 1) while cv. 'Throubolia' exhibited the lowest (H_O = 0.738). Wild olives exhibited the highest expected heterozygosity (H_E = 0.809) while cv. 'Throubolia' exhibited the lowest (H_E = 0.498).

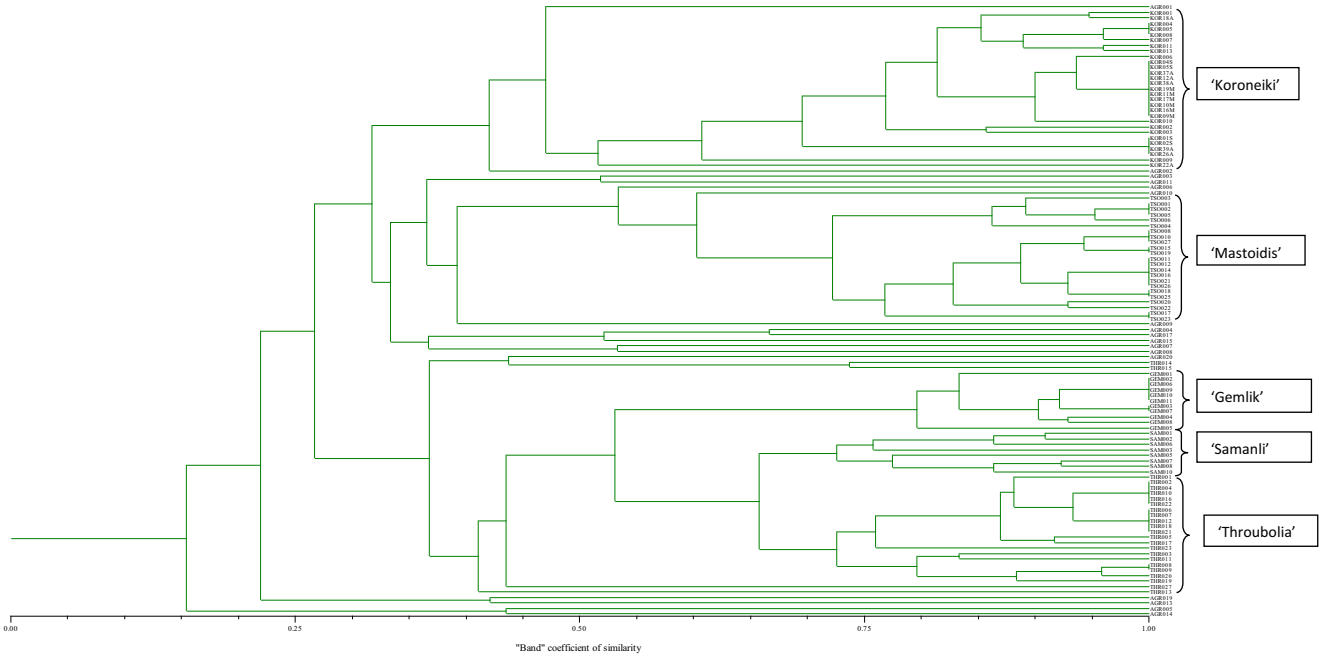

Fig. 2 UPGMA phenogram obtained from 7 SSR loci (81 alleles) using the band similarity coefficient for 112 olive samples of Cretan and Turkish olive and wild olive tree genotypes

On a per-cultivar basis, all H_O were higher than H_E. Specifically, cvs. 'Gemlik', 'Koroneiki', 'Samanli', 'Throubolia' and 'Mastoidis' showed 83.5, 50.6, 69.0, 48.2 and 69.8% increases, respectively. Average values of H_O across all loci were slightly higher (7.3%) in comparison to their respective H_E, indicating the presence of out-breeding. Higher H_O than H_E, determined across all loci, is in turn reflected in the negative inbreeding coefficient (F_{IS}) values. F_{IS} values ranked in diminishing order as follows: (i) Turkish cultivars, (ii) Greek cultivars and (iii) wild olives (Table 4).

Analysis of molecular variance (AMOVA) significantly partitioned genetic variance into two hierarchical levels (Table 5), one among genotypes belonging to the same cultivar (within cultivars; 48%), and another among genotypes belonging to different cultivars (between cultivars; 52%).

Table 4 Genetic diversity indexes of olive cultivars and wild olive populations (AGR) in Crete, Greece, as well as of two major Turkish olive cultivars

Cultivar/population	N_{av}	N_e	I	H_O	H_E	F_{IS}
Koroneiki	4.143	2.224	0.918	0.789	0.524	−0.442
Throubolia	4.571	2.546	0.959	0.738	0.498	−0.427
Mastoidis	3.429	2.515	0.981	1	0.589	−0.723
AGR	9.143	6.093	1.906	0.813	0.809	−0.012
Gemlik	2.857	2.265	0.852	1	0.545	−0.864
Samanli	2.857	2.453	0.842	0.857	0.507	−0.748

N_{av} average number of alleles per locus, N_e number of effective alleles, I Shannon information index, H_O observed heterozygosity, H_E expected heterozygosity, F_{IS} inbreeding coefficient

The mobile of the CBT is shown in Fig. 3 and includes all seven microsatellite loci. The root of the mobile tree includes all five local cultivars as well as the wild population (defined in the ab initio conditions as a singular group), while at the leaf (top) of the tree all five local cultivars and the wild population are recovered. The mobile efficiently describes the loci data in the cultivars while the overall error reduction is 0.97. Table 6 indicates the proportional reduction in error and the improvement achieved at the different steps of the CBT scheme. Six of the loci were used only once, while one microsatellite locus (OeUA-DCA16) appeared twice (once in the separation of cv. 'Koroneiki' from the rest and subsequently to separate cv. 'Throubolia' from cvs. 'Gemlik' and 'Samanli' (see Fig. 3)).

Discussion

The main target of this study was the introduction of seven polymorphic SSR markers in order to (1) assess genetic polymorphism and (2) establish phylogenetic relationships within and between wild and cultivated olives in Crete employing as out-group representative genotypes from two major Turkish cvs., 'Samanli' and 'Gemlik'. The high degree of molecular polymorphism (as evidenced by overall PI), the proper grouping of genotypes into cultivar-specific clusters and the fact that gradual increasing of employed loci up to seven did not allow for any further separation of genotypes indicated that seven loci were adequate for the purposes of the present study. Although important, the number of loci used is not always crucial, as few polymorphic SSR markers may discriminate

Table 5 AMOVA summary table for Cretan and Turkish cultivars

Source of variation	Degrees of freedom	Sum of squares	Mean squares	Estimated variance	Total variance (%)	*P* value
Among cultivars	5	310,913	62,183	3, 272	52	0.001
Within cultivars	106	321,668	3, 035	3, 035	48	0.001
Total	111	632,580		6, 306	100	

Probability for PhiPT estimates (*P* value) is based on 999 permutations across the full data set

more than 100 olive genotypes (Baldoni et al. 2009). Bracci et al. (2009) initially employed 12 SSR loci and identified 60 of the 63 olive tree accessions from Liguria (Italy) and the Mediterranean. Nevertheless, they also found that the same resolution could be obtained by employing a subset of six properly selected SSR loci. Previously, Díaz et al. (2006) had found that all pairs of 51 cultivars, originating from different countries and analysed within their study, could be distinguished using just three selected microsatellites (IAS-oli23, IAS-oli26 and IAS-oli27), which were also employed within the present study.

One indication of cultivar genetic isolation and distinctness is the number of private alleles (i.e. alleles uniquely detected among individuals of a specific cultivar/population). Our

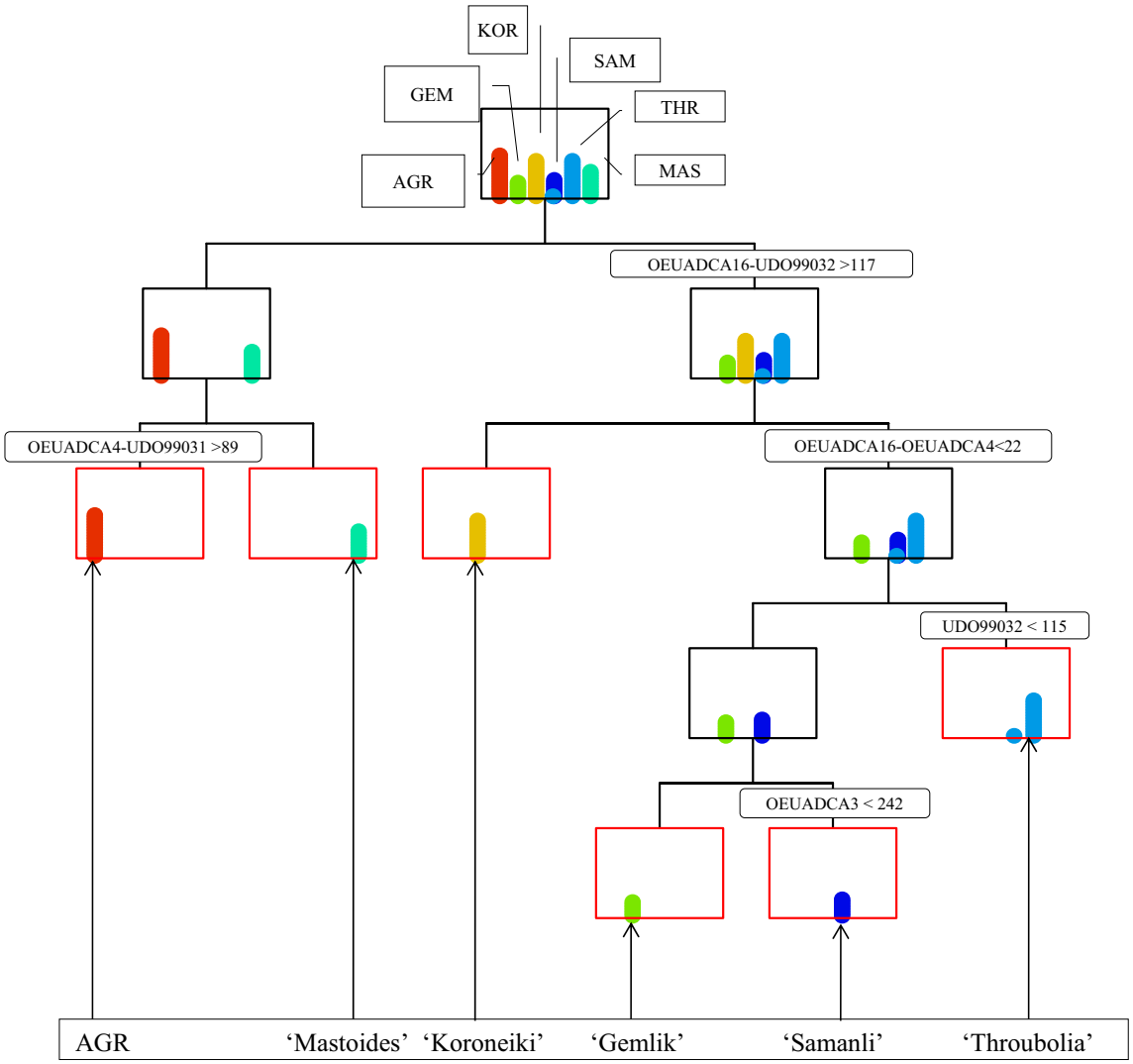

Fig. 3 Mobile constructed by CBT classification of three Greek cultivars ('Mastoides'/MAS, 'Koroneiki'/KOR and 'Throubolia'/THR), two Turkish cultivars ('Gemlik'/GEM, 'Samanli'/SAM) and wild olive genotypes/AGR in Crete. All terminal leaves of the mobile have zero impurities, that is, any cultivar is contained within its own leaf. Moreover the Turkish cultivars (GEM and SAM) which are geographically distantiated from the rest are contained in neighbouring leaves of a single terminal branch, are separated on the basis of the OeUA-DCA3 locus and are distinguished from cv. THR on the basis of the UDO99032 locus

Table 6 CBT analysis of Greek cultivars and wild olive populations in Crete and reference Turkish olive cultivars using the Gini-Simpson index

SSR loci used in splitting	Proportional reduction in error	Improvement achieved
UDO99032-OEUADCA16	0.233	0.233
UDO99031-OEUADCA4	0.455	0.222
OEUADCA16-OEUAD-A4	0.686	0.232
UDO99032	0.850	0.164
OEUADCA3	0.966[a]	0.116

The proportional reduction in error achieved by the correspondent loci is given at each node of the mobile of Fig. 3

[a] Overall improvement in error indicating the ability of CBT analysis to describe the data in a manner analogous to the correlation coefficient

approach allowed for the identification of novel as well as private allelic variants. The latter may be useful for the classification of unknown individuals into a cultivar/population, although private alleles are not detected in all of the individuals in a specific cultivar/population (floating frequency < 1). The number of detected alleles per locus compares favourably with the highest encountered in the literature (ranging from 3 to 16; Alba et al. 2009; Bandelj et al. 2002; Belaj et al. 2010; Cipriani et al. 2002; Muzzalupo et al. 2009; Noormohammadi et al. 2014). The present sampling and analytical approach allowed for detection of a high total number of alleles, in turn permitting the calculation of valid population genetic parameters and producing a significant potential for the discrimination of unknown genotypes. The high average number of detected alleles can be further attributed to the pre-existing genetic diversity of the rather diverse olive tree germplasm cultivated in Crete, Greece or Turkey. The average number of detected alleles per locus within wild olive genotypes is higher compared to the average value of each cultivar. For example, in the case of the widely employed locus OeUA-DCA3 (DCA3), the average value of detected alleles across all cultivars was 4.5 while in wild olives a total of 9.1 alleles were detected. In a similar work, Ganino et al. (2007), found six DCA3 alleles analysing eight ancient cultivars from the Emilia region (Italy). On the other hand, Yoruk and Taskin (2014) analysed 200 and four oleaster trees and 27 local cultivars from Turkey identified 21 DCA3 alleles in total (present study: 13) with a range of 230 to 268 bases and 6 UDO099-019 alleles in total (present study: 8) with a range of 121 to 165 bases. Because of the inter-fertile nature of wild olives, it may be proposed that the number of alleles from the original genetic pool of wild olives is higher than that retained within cultivars. It appears that during the historical development of cultivars, selection, inbreeding and clonal selection have all been applied, resulting in the lower genetic diversity within cultivars.

Only two previous works employed some of the same SSR loci while fingerprinting the same cultivars analysed in the framework of the present study. Baldoni et al. (2009) evaluated cvs. ‘Koroneiki’ and ‘Mastoidis’ composed of single genotypes over the locus OeUA-DCA3. They concluded that locus DCA3 was not polymorphic over cv. ‘Koroneiki’ in contrast to our results detecting two different private alleles and five alleles in total (data not shown) across the entire cv. ‘Koroneiki’ germplasm, while for cv. ‘Mastoidis’ Baldoni et al. (2009) detected two alleles (232 and 243) compared to three alleles detected in the present study across the entire cv. ‘Mastoidis’ range. On the other hand, Roubos et al. (2010) analysed 26 different Greek cultivars each represented by four clonal genotypes while they employed agarose gel for allele fragmentation and sizing. For locus DCA3, they determined an allelic range of 226–253 (present study 220–254) and 7 alleles in total (present study 13). It is worth mentioning that ample molecular genetic polymorphisms were detected at the intra-cultivar level in the present study. This is expected since sampling exclusively focused on aged and autochthonous trees by simultaneously avoiding recently established and clonally propagated trees. On-farm selection by farmers of clones with superior characteristics (e.g. high oil content, drupe size, resistance to drought in areas with low precipitation or resistance to fungal infections in areas of high air humidity) might be the source of this intra-cultivar variation. These genotypes could have derived from seedlings by intra-cultivar self-pollination or from long-term epigenetic adaptation of olive trees to local microenvironments of isolated habitats. Similarly, Lopes et al. (2004) identified several Iberian olive cultivars with high levels of intra-varietal polymorphism. This is in sharp contrast with an extensive analysis of cv. Picholine marocaine in Morocco (Khadari et al. 2008).

Our results exhibit lower diversity indexes when compared to the study of Bracci et al. (2009), who analysed cultivars with heterogeneous origins coming from 10 Mediterranean countries and to Sarri et al. (2006) who studied 112 cultivars from 14 Mediterranean countries.

An increased number of private alleles as well as higher diversity indexes were detected in wild Cretan olives (in comparison to cultivated genotypes) in agreement with Belaj et al. (2010). Higher H_O of wild olives is in agreement with the findings of Lumaret et al. (2004) but in contrast with those of Belaj et al. (2010). The latter authors proposed that continuous crossings among existing traditional and newly introduced cultivars allowed for the maintenance of higher H_O in

cultivated rather in wild olives. Since all cultivars showed higher H_O than H_E, some heterozygotic divergence could be inferred.

In our study, the mean calculated H_O value (0.852) was higher than those calculated for accessions in Italy (0.621 on average), Spain (0.74), Slovenia (0.769) and Iran (0.61) (Alba et al. 2009; Bandelj et al. 2002; Belaj et al. 2010; Muzzalupo et al. 2009; Noormohammadi et al. 2007). Further, all genotypes studied clearly clustered according to their expected cultivar group. Wild olives were separated into three subgroups, each assigned to one of the three Cretan cultivars implying in turn that these are not genuine wild olives but rather feral forms resulting from productive cross-pollination between oleasters and domesticated individuals (also discussed by Khadari et al. 2008). Domesticated forms coexist in most of the olive-producing regions of the Mediterranean basin. This link between wild and cultivated olives was realized perhaps through some type of reticulate evolution (i.e. frequent introgressive hybridizations among the two types with traditional on-farm selection providing the evolutionary direction to the cultivated individuals). A contributing factor may be that true seedlings of the examined cultivars produced new fertile olive trees. Before clonal propagation by cuttings or tissue culture became popular, local nurseries and farmers used trunk pieces of either wild olive trees collected from forest areas or cultivated trees to produce new plants, which would be either grafted or grown self-rooted according to fruit yield, size and oil content. Erre et al. (2010), in a study similar to the present one, determined that most Sardinian cultivars shared the same allelic profiles with the ancient cultivated trees, while Bayesian analysis pointed to an autochthonous origin of Sardinian cultivars. Nevertheless, the wild olive trees from Sardinia formed a separate genetic pool. Similar results to ours were found by Yoruk and Taskin (2014) and by Ipek et al. (2009) by comparing SSR profiles of local Turkish cultivars with those of local wild olive trees. The CBT mobile is in agreement with the similarity dendrogram in that cvs. 'Koroneiki' and 'Mastoidis' exhibit higher genetic similarity between them when compared to cv. 'Throubolia' and the two Turkish cultivars. Additionally, cvs. 'Koroneiki', 'Mastoidis' or 'Throubolia' are genetically closer to wild individuals than to the Turkish cultivars. Wild olives are included in the similarity dendrogram and CBT because they may have served as the original gene pool (true oleasters) from which all cultivars were derived. Conversely, many wild olive trees originated from cultivars through the process of gene flow and introgressive hybridization (feral forms), which results in some individual trees being drawn from more than one group (Mekuria et al. 2002). Indeed, these researchers found that the molecular genetic variation of three groups of olive trees—i.e., (i) feral trees in nearby hills, (ii) fence and (iii) cultivated trees—is not significantly different and that the three clusters formed in a RAPD UPGMA phenogram were homogenous. Nevertheless, this grouping provided evidence that gene flow among the three groups is significant (Mekuria et al. 2002). Further, in the present study, inclusion of wild trees together with cultivars in the CBT better shapes the classification space by coagulating cultivars that are similar to wild olive trees. In fact, the entire data space is partitioned into smaller sections—i.e. recursive partitioning of cultivars and/or wild types—where mathematical interactions of loci are more clear (Breiman et al. 1984). The CBT methodology was also successfully employed in works with virgin oil produced by different *O. europaea* cultivars and different oil adulteration treatments (Petrakis et al. 2008; Agiomyrgianaki et al. 2010) by classifying oils characterized by individual compound contents detected with NMR spectroscopy. The decision process produced according to the present CBT provides a method for sorting any new unknown material purportedly originating from any of the presently analysed cultivars. It appears that, based on just seven SSR loci, we have devised a system for sorting the present cultivar collection according to an a priori phenotypic classification and to predict membership of future unknown samples to any of the presently analysed cultivars.

Ample genetic diversity was detected between and within cultivars. Ganino et al. (2007) analysed ancient trees in the Emilia region in Northern Italy and concluded that the longevity of this species and the low breeding pressure has contributed to the conservation of its variability and that the reduced extent of genetic erosion within its germplasm has allowed the persistence of olive diversity. The elevated genetic diversity is apparent in the similarity dendrogram, whereupon inclusion of all analysed genotypes creates a complex network of linked genotypes. Groups of identical genotypes were also discovered with all analysed cultivars (with the exception of cv. 'Samanli') which can likely be attributed to the presence of clonal materials. This is especially true for the establishment of new olive orchards in Crete over the past 50 years given the preference of farmers for cv. 'Koroneiki' due to its suitability for modern intensive farming. On the other hand, the Cretan cv. 'Throubolia' exhibits higher genetic affinity to the Turkish cvs. 'Gemlik' and 'Samanli' than to the other two Cretan cvs. 'Koroneiki' and 'Mastoidis' in agreement with Owen et al. (2005). Yoruk and Taskin (2014) examined the genetic diversity of cvs. 'Gemlik' and 'Samanli' within the broad genetic pool of Turkish olive cultivars and wild olive trees. They concluded that cv. 'Gemlik' is a widespread cultivar in the Marmara region but that this cultivar is not related to any of the studied cultivars from other regions in Turkey. Further, Yoruk and Taskin (2014) found that the majority of Turkish cultivars were not clearly clustered according to their geographic origin and proposed that humans have moved cultivars to different sites in the country during the past several 1000 years of olive cultivation. Some interesting scenarios regarding the post-glacial persistence of oleasters in Greece and Turkey and the history of subsequent domestication and

selection of cultivars were presented by Besnard and Rubio de Casas (2016). Nevertheless, further detailed research is needed to ascertain the exact paths of material flow in the area.

Acknowledgements The research presented herein was funded by the Greek Secretariat for Research and Technology (GSRT), project Regional Innovation Pole of Crete, "i4Crete, A8—Integrated system of olive oil fingerprinting control and promotion—11RIPC06" to AGD and to ITM which was coordinated by the Heraklion Chamber of Commerce and Industry, Heraklion, Crete, Greece. Ms. M. Aksehirli-Pakyurek's research visit to NAGREF was funded by an Erasmus fellowship administered by the Technological Educational Institute of Crete (Prof. Ioannis Vlahos). Thanks are due to Ms. Maria Pikraki and to Mr. Dimitrios Giakoumakis for their assistance with SSR markers and to Ms. Irene Stratidaki (IMBB-FORTH) for assistance with the LICOR sequencer. Comments of the three anonymous reviewers are greatly appreciated.

References

Agiomyrgianaki A, Petrakis PV, Dais P (2010) Detection of refined olive oil adulteration with refined hazelnut oil by employing NMR spectroscopy and multivariate statistical analysis. Talanta 80:2165–2171. doi:10.1016/j.talanta.2009.11.024

Alba V, Montemurro C, Sabetta W, Pasqualone A, Blanco A (2009) SSR-based identification key of cultivars of *Olea europaea* L. diffused in Southern-Italy. Sci Hortic 123:11–16

Baldoni L, Cultrera NG, Mariotti R, Ricciolini C, Arcioni S, Vendramin GG, Buonamici A, Porceddu A, Sarri V, Ojeda MA, Trujillo I, Rallo L, Belaj A, Perri E, Salimonti A, Muzzalupo I, Casagrande A, Lain O, Messina R, Testolin R (2009) A consensus list of microsatellite markers for olive genotyping. Mol Breed 24:213–231

Bandelj D, Jakse J, Javornik B (2002) DNA fingerprinting of olive varieties by microsatellite markers. Food Technol Biotechnol 40:185–190

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 107:736–744

Belaj A, Munoz-Diez C, Baldoni L, Satovic Z, Barranco D (2010) Genetic diversity and relationships of wild and cultivated olives at regional level in Spain. Sci Hortic 124:323–330

Belaj A, Leon L, Satovic Z, De la Rosa R (2011) Variability of wild olives (*Olea europaea* subsp. *e*uropaea var. sylvestris) analyzed by agro-morphological traits and SSR markers. Sci Hortic 129:561–569

Besnard G, Rubio de Casas R (2016) Single vs multiple independent olive domestications: the jury is (still) out. New Phytol 209:466–470

Besnard G, Khadari B, Navascues M, Fernandez-Mazuecos M, El Bakkali A, Arrigo N, Baali-Cherif D, Brunini-Bronzini de Caraffa V, Santoni S, Vargas P, Savolainen V (2013) The complex history of the olive tree: from Late Quaternary diversification of Mediterranean lineages to primary domestication in the northern Levant. Proc R Soc B 280(1756):20122833. doi:10.1098/rspb.2012.2833

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic-linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Bracci T, Sebastiani L, Busconi M, Fogher C, Belaj A, Trujillo I (2009) SSR markers reveal the uniqueness of olive cultivars from the Italian region of Liguria. Sci Hortic 122:209–221

Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and regression trees. Wadsworth International Group, Belmont ISBN-10: 0412048418

Breton C, Terral JF, Pinatel C, Medail F, Bonhomme F, Berville A (2009) The origins of the domestication of the olive tree. C R Biol 332:1059–1064

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

Colella C, Miacola C, Amenduni M, D'Amico M, Bubici G, Cirulli M (2008) Sources of verticillium wilt resistance in wild olive germplasm from the Mediterranean region. Plant Pathol 57:533–539

Díaz A, de la Rosa R, Martin A, Rallo P (2006) Development, characterization and inheritance of new microsatellites in olive (*Olea europaea* L.) and evaluation of their usefulness in cultivar identification and genetic relationship studies. Tree Genet Genomes 2:165–175

Diez CM, Trujillo I, Martinez-Urdiroz N, Barranco D, Rallo L, Marfil P, Gaut BS (2015) Olive domestication and diversification in the Mediterranean Basin. New Phytol 206:436–447

Dixon P (2003) VEGAN, a package of R functions for community ecology. J Veg Sci 14:927–930 R package version 2.0–3

Erre P, Chessa I, Munoz-Diez C, Belaj A, Rallo L, Trujillo I (2010) Genetic diversity and relationships between wild and cultivated olives (*Olea europaea* L.) in Sardinia as assessed by SSR markers. Genet Resour Crop Evol 57:41–54

Ganino T, Beghe D, Valenti S, Nisi R, Fabbri A (2007) RAPD and SSR markers for characterization and identification of ancient cultivars of *Olea europaea* L. in the Emilia region, Northern Italy. Genet Resour Crop Evol 54:1531–1540

Hagidimitriou M, Katsiotis A, Menexes G, Pontikis C, Loukas M (2005) Genetic diversity of major Greek olive cultivars using molecular (AFLPs and RAPDs) markers and morphological traits. J Am Soc Hortic Sci 130:211–217

International Olive Council (2000) World catalogue of olive varieties. Madrid, Spain.http://www.internationaloliveoil.org/store/index/665-world-catalogue-of-olive-varieties-publications

International Olive Council (2016) Market newsletter. No 108 – September 2016.http://www.internationaloliveoil.org/news/view/686-year-2016-news/762-market-newsletter-september-2016

Ipek A, Barut E, Gulen H, Oz AT, Tangu NA, Ipek M (2009) SSR analysis demonstrates that olive production in the southern Marmara region in Turkey uses a single genotype. Genet Mol Res 8:1264–1272

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol 16:1099–1106

Khadari B, Charafi J, Moukhli A, Ater M (2008) Substantial genetic diversity in cultivated Moroccan olive despite a single major cultivar: a paradoxical situation evidenced by the use of SSR loci. Tree Genet Genomes 4:213–221

Lanza F (2011) Olive: a global history. Reaktion Books, London ISBN: 1861898681

Linos A, Nikoloudakis N, Katsiotis A, Hagidimitriou M (2014) Genetic structure of the Greek olive germplasm revealed by RAPD, ISSR and SSR markers. Sci Hortic 175:33–43

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

Lumaret R, Ouazzani N, Michaud H, Vivier G, Deguilloux MF, Di Giusto F (2004) Allozyme variation of oleaster populations (wild olive tree) (*Olea europaea* L.) in the Mediterranean basin. Heredity 92:343–351. doi:10.1038/sj.hdy.6800430

Lynch M (1990) The similarity index and DNA fingerprinting. Mol Biol Evol 7:478–484

Mekuria GT, Collins G, Sedgley M (2002) Genetic diversity within an isolated olive (*Olea europaea* L.) population in relation to feral spread. Sci Hortic 94:91–105

Mkize N, Hoelmer KA, Villet MH (2008) A survey of fruit-feeding insects and their parasitoids occurring on wild olives *Olea europaea*-ssp. *cuspidate*, in the Eastern Cape of South Africa. Biocontrol Sci Tech 18:991–1004

Muzzalupo I, Stefanizzi F, Salimonti A, Falabella R, Perri E (2009) Microsatellite markers for identification of a group of Italian olive accessions. Sci Agric (Piracicaba, Braz) 66:685–690. doi:10.1590/S0103-90162009000500014

Newton C, Lorre C, Sauvage C, Ivorra S, Terral JF (2014) On the origins and spread of *Olea europaea* L. (olive) domestication: evidence for shape variation of olive stones at Ugarit, Late Bronze Age, Syria—a window on the Mediterranean Basin and on the westward diffusion of olive varieties. Veg Hist Archaeobotany 23:567–575

Nikoloudakis N, Banilas G, Metzidakis J, Gazis F, Hatzopoulos P (2003) Discrimination and genetic diversity among cultivated olives of Greece using RAPD markers. J Am Soc Hortic Sci 128:741–746

Noormohammadi Z, Hosseini-Mazinani M, Trujillo I, Ratio L, Belaj A, Sadeghizadeh M (2007) Identification and classification of main Iranian olive cultivars using Microsatellite markers. Hortscience 42:1545–1550

Noormohammadi Z, Trujillo I, Belaj A, Ataei S, Hosseini-Mazinan M (2014) Genetic structure of Iranian olive cultivars and their relationship with Mediterranean's cultivars revealed by SSR markers. Sci Hortic 178:175–183

Owen CA, Bita EC, Banilas G, Hajjar SE, Sellianakis V, Aksoy U, Hepaksoy S, Chamoun R, Talhook SN, Metzidakis I, Hatzopoulos P, Kalaitzis P (2005) AFLP reveals structural details of genetic diversity within cultivated olive germplasm from the Eastern Mediterranean. Theor Appl Genet 110:1169–1176

Peakall R, Smouse PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol Ecol Notes 6:288–295

Petrakis PV, Agiomyrgianaki A, Christophoridou S, Spyros A, Dais P (2008) Geographical characterization of Greek virgin olive oils (Koroneikicv.) using H-1 and P-31 NMR fingerprinting with canonical discriminant analysis and classification binary trees. J Agric Food Chem 56:3200–3207

R Development Core Team (2008) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna ISBN 3–900051–07-0, URL http://www.R-project.org/

Rohlf FJ (2008) NTSYSpc: numerical taxonomy system, ver. 2.21q. Exeter Publishing Ltd, Setauket

Roubos K, Moustakas M, Aravanopoulos FA (2010) Molecular identification of Greek olive (*Olea europaea*) cultivars based on microsatellite loci. Genet Mol Res 9:1865–1876

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 geographically defined populations. Genome 49: 1606–1615

Sefc KM, Steinkellner H, Wagner HW, Glössl J, Regner F (1997) Application of microsatellite markers to parentage studies in grapevine. Vitis 36:179–183

Sefc KM, Lopes S, Mendonca D, Dos Santos MR, Machado MLD, Machado AD (2000) Identification of microsatellite loci in olive (*Olea europaea*) and their characterization in Italian and Iberian olive trees. Mol Ecol 9:1171–1173

Sneath PHA, Sokal RR (1973) Numerical taxonomy: the principles and practice of numerical classification. W. H Freeman and Company, San Francisco

Terzopoulos PJ, Kolano B, Bebeli PJ, Kaltsikes PJ, Metzidakis I (2005) Identification of *Olea europaea* L. cultivars using inter-simple sequence repeat markers. Sci Hortic 105:45–51

Therneau TM, Atkinson EJ (1997) An introduction to recursive partitioning using the RPART routine. Technical report 61, Mayo Foundation, Rochester, USA

Trujillo I, Ojeda MA, Urdiroz NM, Potter D, Barranco D, Rallo L, Diez CM (2013) Identification of the worldwide olive germplasm bank of Cordoba (Spain) using SSR and morphological markers. Tree Genet Genomes 10:141–155

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

Xanthopoulou A, Ganopoulos I, Koubouris G, Tsaftaris A, Sergendani C, Kalivas A, Madesis P (2014) Microsatellite high-resolution melting (SSR-HRM) analysis for genotyping and molecular characterization of an *Olea europaea* germplasm collection. Plant Genet Resour 12: 273–277

Yoruk B, Taskin V (2014) Genetic diversity and relationships of wild and cultivated olives in Turkey. Plant Syst Evol 300:1247–1258