Euphytica **116:** 131–142, 2000.

DNA fingerprinting and classification of geographically related genotypes of olive-tree (*Olea europaea* L.)

M. Gonzalo Claros*, Remedios Crespillo, María L. Aguilar & Francisco M. Cánovas
*Departamento de Biología Molecular y Bioquímica, Facultad de Ciencias, Instituto Andaluz de Biotecnología, Universidad de Málaga, E–29071 Málaga, Spain; (*author for correspondence; e-mail: claros@uma.es)*

Received 8 September 1999; accepted 16 March 2000

Key words: AP-PCR, cultivar identification, *Olea europaea*, olive tree, RAPD

Summary

Málaga is a province of Spain where olive-trees are cultivated in a large range of environments, climates and soils. We have developed a reliable and reproducible method to detect RAPD and AP-PCR polymorphisms, using DNA from olive-tree (*Olea europaea* L.) leaves. Starting from their natural orchards, fifty-six olive-tree cultivars throughout Málaga province, including oil and table olive cultivars, were screened and grouped into 22 varieties. A total of 62 informative polymorphic loci that provide 601 conspicuous bands were enough to differentiate the varieties. Clustering analyses managing 3 different pairwise distances, as well as phylogenetic analyses, led to the same result: olive-trees in Málaga can be divided into three main groups. Group I (90% of certainty) contains wild type and two introduced varieties, group II (83% of certainty) covers some native olive-trees, and group III (58% of certainty) is an heterogeneous cluster that includes varieties originating and cultivated in a number of Andalusian locations. Geographic location seems to be the first responsible of this classification, and morphological traits are needed to justify the group III subclustering. These results are consistent with the hypothesis of autochthonic origin of most olive-tree cultivars, and have been used to support a Label of Origin for the olive oil produced by the varieties included in group II.

Abbreviations: AP-PCR – Arbitrarily primed PCR; CTAB – Cetyl-trimethyl-ammonium bromide; PCR – Polymerase chain reaction; PEG – Polyethylene glycol; PVP – Polyvinyl pirrolidone; RAPD – random amplified polymorphic DNA; UPGMA – Unweighted pair group method with arithmetic averages

Introduction

The olive-tree (*Olea europaea* L.) is one of the most characteristic species of the Mediterranean area and nowadays is the only cultivated representative of the genus *Olea*. Since its chromosome complement is 46, it is considered – although not demonstrated – a tetraploid species, and several botanical varieties have usually been recognised (Brousse, 1987). Two of the varieties correspond to cultivated *sativa*, which is propagated vegetatively, and wild type *sylvestris* (syn. *Olea oleaster* Hoffm. et Link), which only reproduces by sexual means and is used for grafting some *sativa* varieties. The appearance and early cultivation of the olive-tree date back to prehistory (Chevalier, 1948), being a product of hybridisation of some living *Olea* species with others which have since disappeared. In fact, the wild olive-trees occupy the same niches than feral forms originated by naturalisation from the cultivated varieties (Zohary & Hopf, 1994).

Accurate and rapid identification of clones, varieties, or species is especially important in vegetatively propagated plants. Differentiation among olive cultivars is traditionally supported by numerous morphological and pomological traits. At least 156 different varieties (Barranco & Rallo, 1984) have been identified in Andalusia (Spain) alone using these descriptions. Unfortunately, phenotypical traits are difficult to evaluate, affected by subjective interpretations, and severely influenced by the environment (such as ag-

ronomic cares) and plant developmental stage. More recently, workers have found allozymes to be useful markers for an objective cultivar identification (Pontikis et al., 1980; Loukas & Krimbas, 1983; Trujillo et al., 1990; Ouazzani et al., 1993; Ouazzani et al., 1996). However, allozymes are products of gene expression, sensitive to environmental influences and tissue-specificity, facts that hamper the interpretation of results. Hence, the *Olea* genus taxonomy is being reviewed (Green & Wickens, 1989). All these challenges can be bypassed by direct analysis of DNA. DNA provides an opportunity for direct comparison of different genetic material independent of any influences since it generates an almost unlimited quantity of markers that are independent of environment and developmental stages. At the moment, there are only a few olive genes cloned (Villalba et al., 1994; Asturias et al., 1997; Batanero et al., 1997; Haralampidis et al., 1998), limiting the preparation of specific molecular markers. Fingerprinting with arbitrary oligonucleotides (Welsh & McClelland, 1990; Williams et al., 1993) allow a rapid, cost-effective analysis of the polymorphism and genetic distances of many individuals like other methods (Demeke et al., 1992). Fingerprinting performed on arbitrarily chosen olive-tree (Bogani et al., 1994; Vergari et al., 1996; Rubio & Arús, 1997; Belaj et al., 1998) suggested varying levels of importance of morphological traits and geographic origin (Fabbri et al., 1995; Wiesman et al., 1998; Angiolillo et al., 1999). In such state of art, there is no doubt that development of molecular markers will shed light on obtaining reliable olive-tree classification to clearly differentiate the varieties, providing new insights in the knowledge of phylogenetic relations of cultivars. In addition, due to the long juvenility phase of olive-tree, breeding strategies are hard to perform (Lavee, 1990; Rallo, 1994). Thus, molecular markers should be helpful since they can shorten the elapsed breeding time and they can diminish the experimental orchard size.

In Southern Spain, the province of Málaga is a complex orographic zone that comprises a large range of environments, contrasting climates, and diversity of soils. Hence, the olive-tree orchards should be adapted to diverse conditions, which may determine a high level of variability among cultivars. Despite the economic importance, the number and nature of olive-tree varieties is not well established, since there are native, Andalusian, and foreign, recently-introduced varieties. The present study is an attempt, using RAPD and AP-PCR markers, to identify, characterise, and establish relationships of wild and geographically-related olive-tree cultivars. The relevance of the resulting classification is discussed.

Materials and methods

Plant material

Fifty-six *Olea europaea* genotypes (Table 1) localised in the province of Málaga (Andalusia, Spain) were used in this study, collected from their natural orchards, excluding feral individuals. They include four different wild types of *Olea europaea sylvestris*, and three genotypes of 'Blanqueta' and 'Arbequina' that come from the north-east of Spain. Eighteen of the genotypes are recognised as native to Málaga, belonging to the cultivars 'Verdial de la Axarquía', 'Nevadillo Blanco', 'Picudo de la Axarquía', 'Aloreña', and 'Gordal de Archidona'. Only the last two can be found out of Málaga. The rest are genotypes corresponding to cultivars that are also largely cultivated in Andalusia. With the exception of genotypes called 'Picudo', different genotypes that receive the same name are morphologically similar. The precise geographical location of the collecting areas is available upon request. To represent the maximal genetic diversity occurring in a genotype, 10 different trees were sampled for each site when available.

DNA extraction

Olive-tree leaves were collected in Spring 1998 and stored at –70 °C before DNA extraction. Total DNA was prepared from leaf tissue as follows: 1.0 g of leaves were ground in liquid nitrogen and mixed with 5 ml of extraction buffer (100 mM Tris-HCl pH 8, 1.4 M NaCl, 20 mM EDTA, 2% CTAB, 0.1% $NaHSO_3$). Then it was supplemented with 0.2% 2-mercaptoethanol and 100 mg of insoluble PVP (Sigma). Samples were incubated 30 to 60 minutes at 65–70 °C and then chilled. Chloroform was added (5 ml) and mixed vigorously to homogenisation. The aqueous phase was recovered by centrifugation for 5 min at 5000 × g, and 0.5 ml of a solution containing 10% CTAB and 0.7 M NaCl were added and mixed vigorously. Samples were supplemented with 5 ml chloroform, mixed and centrifuged again for 5 min at 8500 × g. The upper phase was taken and 0.6 volumes of 2-propanol were added by gently mixing to precipitate DNA. Tubes were held at room temperature for enough time to permit DNA to reach the bottom of the

Table 1. Description of cultivars used in this study. Cultivars marked with (^{a}) are considered as native, those marked with (^{b}) are considered as foreign since they were introduced in Malaga during the last 60 years, and those marked with (^{c}) are used for olive consumption. The synonymous column gathers other cultivar names utilised for the same genotype

#	Name	Cultivar # × individual #	Polymorphic bands #	Synonymous
1	Aloreña^{a,c}	2 × 10	29	Manzanillo
2	Acebuche fino	2 × 10	35	Bravío
4	Acebuche basto	2 × 10	31	Bravío
6	Zorzaleño	3 × 10	30	
7	Picudo de El Burgo	1 × 10	27	Picuillo
15	Verdial de la Axarquía^a	6 × 10	29	
17	Blanqueta^b	1 × 10	27	
20	Hojiblanca	5 × 10	25	
22	Gordal de Archidona^{a,c}	5 × 10	25	
28	Picudo de la Axarquía^a	1 × 10	33	Picuillo
31	Nevadillo Blanco^a	4 × 10	24	Lechín, Blanquilla
34	Arbequina^b	2 × 10	24	
37	Manzanillo sevillano^c	2 × 10	18	
42	Lechín de Granada	2 × 10	34	
46	Picudo de Baena	3 × 10	27	Carrasqueño
60	Picual	4 × 10	24	Marteño
64	Hojiblanca Gaona^a	1 × 6 clones	24	
67	Morisco	1 × 10	35	
68	Verdial de Ronda	1 × 10	23	
69	Lechín de Sevilla	5 × 10	24	Lechín Ecijano
70	Picudo de Ronda	1 × 10	22	
77	Chorrúo	2 × 9	31	

tube (no more than 5 min). Supernatant was removed and the remaining sediment centrifuged for 15 minutes at room temperature and 8500 × g. The pellet was rinsed with 0.5 ml of 70% ethanol and centrifuged again for 3 minutes. It was dried briefly at vacuum and resuspended in 0.5 ml TE (10 mM Tris-HCl pH 7.4, 1 mM EDTA) in a bath at 55–65 °C to accelerate the process. The tube was centrifuged 5 minutes at 8500 × g to remove any insoluble particles and the supernatant was digested with 0.1 mg/ml RNAse A (Roche Biochemicals) for 45 minutes at 37 °C. Then, DNA was precipitated for 5 minutes at 4 °C with an equal volume of a cold, filtered solution containing 13% PEG-8000 (Sigma) and 1.6 M NaCl. The solution was centrifuged 10 minutes at 12000 × g at 4 °C, and the pellet rinsed with 70% ethanol and vacuum dried. The pellet was resuspended in 0.5 ml TE as previously, its concentration determined using a spectrophotometer at 260 nm, and diluted to a working solution of 20 ng/μl.

Fingerprinting analysis

Decamers with a GC content of 60% were obtained from Operon Technologies (Table 2). The 12 primers that provide repeatable variability were used for RAPD analysis following a modified version of the original procedure (Williams et al., 1993) to optimise results with olive-tree DNA. Reactions were carried out using 200 μl extra-thin tubes in a volume of 20 μl containing 67 mM Tris-HCl pH 8.8, 16.6 mM $(NH_4)_2SO_4$, 0.01% Tween-20, 2 mM $MgCl_2$, 0.1 mM each of dATP, dCTP, dGTP and dTTP, 0.4 mM primer, 0.5 U of EcoTaq (EcoGen, Spain), and 10 ng DNA. Amplification reactions were performed in a GeneAmp PCR system 2400 (Perkin-Elmer) thermal cycler programmed for 1 cycle of 1 min at 91 °C followed by 35 cycles of 1 min at 91 °C, 1 min at 36 °C

Table 2. Sequences of the oligonucleotides utilised for the RAPD and AP-PCR analyses. The polymorphic band name is given as the size in bp

Name	Sequence	Utility	Polymorphic bands (bp)
OPF2		RAPD	345
			1028
OPF6		RAPD	410
			561
OPF8		RAPD	308
			342
			557
			1594
OPF10	GGAACCTTGG	RAPD	488
OPJ1	CCCGGCATAA	RAPD	512
			683
			710
			740
OPJ6	TCGTTCCGCA	RAPD	646
			812
			905
			1278
			1563
			2244
OPJ18	TGGTCGCAGA	RAPD	230
			483
			1635
OPX4	CCGCTACCGA	RAPD	307
			400
			420
			611
			1600
			2017
OPX6	ACGCCAGAGG	RAPD	437
			557
			630
			930
			1259
			1403
			1988
OPAH2	CACTTCCGCT	RAPD	482
			720
			978
			1003
			1452
OPAI5	GTCGTAGCGG	RAPD	725
			1217
			1400
			1656
			2255
OPAI14	TGGTGCACTC	RAPD	485
			555
			775
			886
SP6	TATTTAGGTGACACTATAG	AP-PCR	523
			1107
			1328
T7	TAATACGACTCACTATAGGG	AP-PCR	329
			396
			443
			579
			689
			1204
			1687
OLE2a	TTGGGCGGGTACATGCCC	AP-PCR	644
			865
			1420

and 1.5 min at 72 °C, for denaturing, annealing and primer extension phases respectively. The last cycle was followed by 7 minutes at 72 °C.

AP-PCR reactions were also optimised from the original protocol (Welsh & McClelland, 1990). The reaction was prepared as for RAPD but containing 0.2 mM of each deoxynucleotide and 15 ng of genomic DNA. Primers utilised were those that, like in RAPD, provide repeatable variability results, namely SP6, T7 and OLE2a (Table 2). Amplification reactions were programmed for 2 cycles of 5 min at 91 °C, 5 minutes at 37 °C and 5 min at 72 °C, followed by 35 cycles of 1 min at 91 °C, 1 min at 56 °C and 2 min at 72 °C, for denaturing, annealing and primer extension phases, respectively. The last cycle was followed by 5 minutes at 72 °C.

Amplification RAPD and AP-PCR products were analysed by gel electrophoresis run at 3.5 V/cm in 1.5% agarose in 0.5X TBE buffer, stained with ethidium bromide (0.5 μg/ml) and photographed and recorded under UV light using a UVP Store 500 System with a Sony Videographic printer UP-890CE. Molecular sizes were estimated using phage λ digested with *Eco*RI and *Hin*dIII for bands greater than 1 kb, and pFL61 (Minet et al., 1992) digested with *Hpa*II for bands under 1 kb. All reactions were repeated at least three times using different batches of EcoTaq polymerase and running replicate samples on the upper and lower half of the same 15 $\times$ 20 cm agarose gel (BioRad). Only reproducible bands between 250 and 2500 bp were used in analyses. Each amplification fragment was named by the primer used and its approximate size in base pairs. A reaction mixture without template DNA was seen with each amplification as a negative control.

Data scoring and analysis

Amplified bands generated by RAPD and AP-PCR were visually scored as 1 if present and 0 if absent

only when they are consistent in at least three independent experiments. No score is given for band intensity, the same value being given for strong or weak band presence. Following recent studies that reveal that there is no difference in using all band or only polymorphic bands (Barker et al., 1999), only the last were considered in data sets. The phylogenetic content of the data was mined using likelihood-mapping (Strimmer & von Haeseler, 1997). Since it has been pointed out that different metric to calculate pairwise distances are related by a simple monotonic function (Gower, 1985), we have chosen three independent algorithms to construct pairwise distances: maximum-likelihood (Felsenstein, 1981), Dice (Dice, 1945), and Apostol (Apostol et al., 1993) indexes. The pairwise matrices are available to readers upon request. Unrooted dendrograms (phenograms) were constructed based on the similarity matrix data by applying the unweighted pair group method with arithmetic averages (UPGMA) cluster analysis, and verified with the quartet-puzzling method (Strimmer & von Haeseler, 1996) since this method is as effective as large studies like bootstrap or jack-knife resampling, which require runs on multiple data sets (Felsenstein, 1988). Phylogeny was estimated from the pairwise distances with the Neighbor-Joining method (Saitou & Nei, 1987). Calculations were performed with PHYLIP v3.5 package (Felsenstein, 1993), Puzzle v4.0 (Strimmer & von Haeseler, 1997), and RAPDistance (Amstrong et al., 1994). Graphic display of dendrograms was visualised using TreeView (Page, 1996).

Results

DNA extraction and fingerprinting

We have modified a plant DNA extraction method (Doyle & Doyle, 1987) to avoid the two main contaminants that reduce the reliability of fingerprinting: the polysaccharides that will provoke the loss of larger amplified bands, and the RNA (and small DNA fragments) that is able to non-specifically prime DNA polymerase reactions. Polysaccharides and other organic compounds are removed by means of CTAB and PVP treatments (Claros & Cánovas, 1998). Nucleic acids interference is avoided by precipitation with PEG (small RNA and DNA fragments do not precipitate, results not shown). The procedure provided identical results with DNA samples isolated from either young or old leaves (results not shown) enabling future sampling at any point in the year.

The facts that we use a different DNA extraction protocol and a different DNA polymerase prompted us to develop new fingerprinting conditions. Annealing temperature between 34 and 38 °C, 5 to 500 ng of DNA, and 0.1 to 2 units of EcoTaq polymerase were tested, in the same way as described (Williams et al., 1993; Ferreira & Keim, 1997), to produce repeatable amplifications. Two tests were carried out: first, amplification reactions with DNA obtained from different cultivars to reveal that the same variety provided identical fingerprints. Second, the amplification patterns obtained with different EcoTaq batches to reveal the reproducibility. Two different experiments using DNA prepared with our method are presented in Figure 1, where the polymerisation of high molecular weight bands, the absence of spurious amplifications, and the reproducibility of the results are clearly shown.

From cultivars to varieties: polymorphism identification

The first task was to gather genotypes that belong to the same cultivar taking into account only their DNA fingerprints. RAPD analyses of the 56 genotypes permitted us to distinguish 22 cultivars. Most cultivars with the same name were, as expected, grouped together. In some cases, we have grouped genotypes with different names, enabling to establish several synonymies (Table 1). Conversely, there is one homonymous that is used for a broad group of genotypes, which is ‘Picudo’ or ‘Picuillo’, since we have been able to divide it into 4 completely independent cultivars that have been renamed according to their geographic origin. Similarly, we have differentiated one ‘Hojiblanca’ genotype, that consists in 6 clones of the same tree, whose main morphological difference is its large olive size. Hereafter, we will consider only the 22 distinct cultivars identified (Table 1).

RAPD and AP-PCR amplification products of 15 primers revealed, among the 22 olive-tree cultivars considered in this study, a total of 62 polymorphic amplified fragments. These polymorphic loci provide 601 bands that were conspicuous and highly reproducible, that is, they were always present or absent in three independent experiments. A representative result is displayed in Figure 1. Fingerprinting reliability has been questioned since comigrating bands from different individuals do not necessarily represent homologous amplification products, but the presence of comigrating bands only reduces the absolute similarities, not the relative similarities nor the relationships

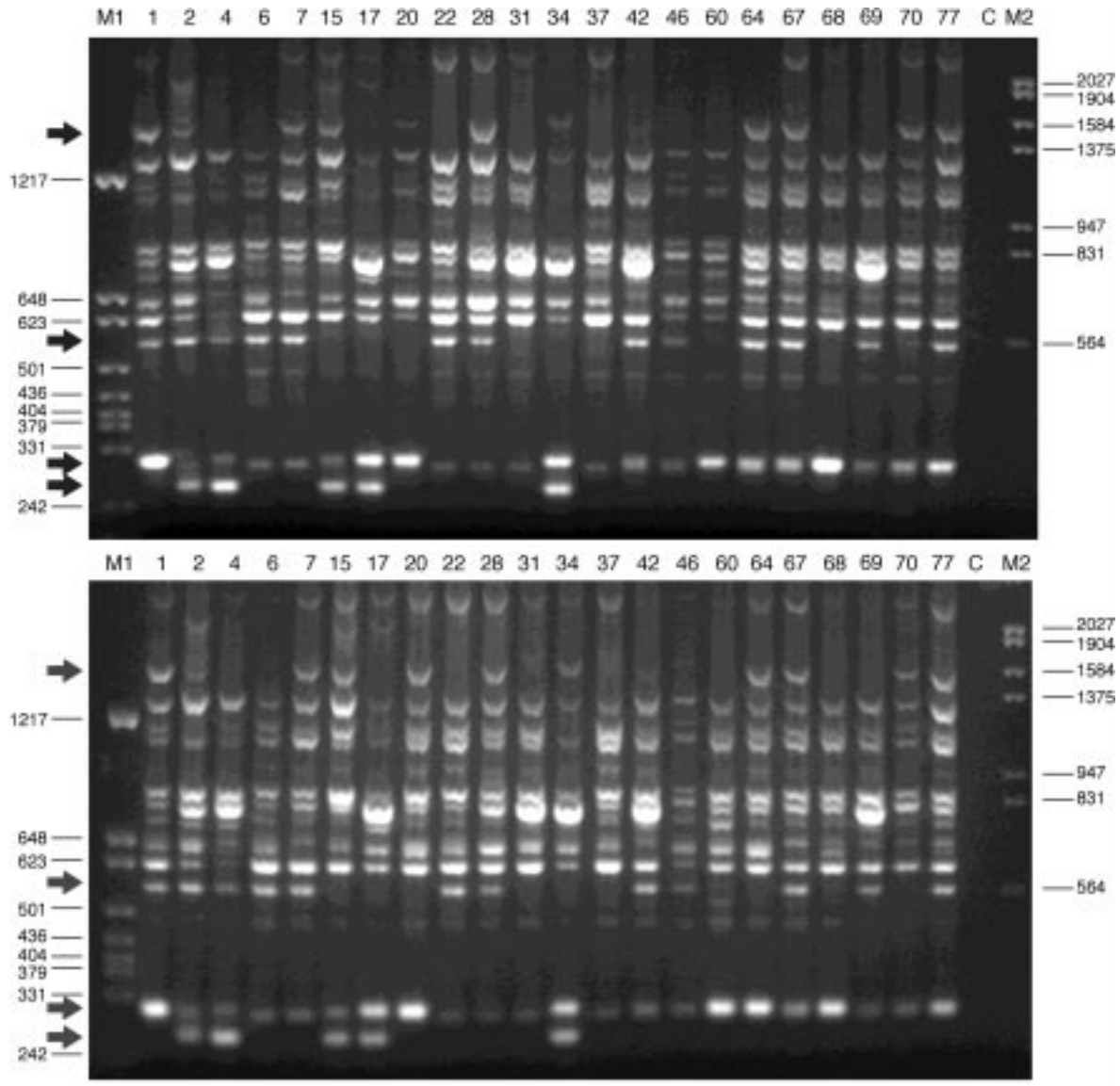

Figure 1. Two replications of the RAPD pattern obtained from the 22 olive cultivars with the primer OPF8 using different DNA preparations and EcoTaq batches. Numbers correspond to cultivars as explained in Table 1. C is negative control (RAPD reaction without template). M1 is the DNA molecular weight marker pFL61-*Hpa*II. M2 is the DNA molecular weight marker λ-*Eco*RI+*Hin*dIII. The arrows are the fragments considered as polymorphic for this study.

among clusters (Adams & Rieseberg, 1998). Hence, fragment size and frequency can be considered a reliable predictor of homology of closely related individuals (Nicese et al., 1998), provided that they can differentiate between at least two populations. Prior to pairwise distance calculation, each cultivar was subjected to a 5% level χ^2 test to determine whether its 0/1 composition is identical to the average composition of the whole matrix, since samples with deviating composition violate the basic assumption implicit in the pairwise distances. The frequency of band presence was 0.44 and only the 'Manzanillo' variety has a slightly biased composition ($\chi^2 = 1.42$). An evaluation of the tree-likeliness of data was performed with the likelihood-mapping revealing that 55.2% of cases will tend to a tree-like phenogram while 33.6% of them will tend to a star-like distribution. This suggests that some olive cultivars will not be well resolved or that

they share a common, recent ancestor (Strimmer & von Haeseler, 1997).

The polymorphisms were appropriate to differentiate varieties since each one is defined by a set of several markers. The number of polymorphic and reproducible fragments per primer ranged from 1 (primer OPF10) to 7 (primers T7 and OPX6), and the fragment size from 2.2 kb (primer OPJ6) to 0.23 kb (primer OPJ18). Products beyond this range were generally neither clear nor reproducible. Sometimes a polymorphism characterised one cultivar by its presence (OPJ1-740 for 'Acebuche', the wild type) or absence (OPAH2-1452 for 'Blanqueta'). The results obtained comprised a unique fingerprint for each of the 22 olive-tree cultivars, allowing an unequivocal identification of each genotype, even in cases like 'Hojiblanca' and 'Hojiblanca Gaona' varieties which only differ in a single band (marker T7-443).

Clustering the cultivars

Due to the tetraploidy of olive-tree and that RAPD and AP-PCR markers are dominant, only one allele out of four is sufficient to lead to an amplification product. Consequently, we cannot measure any genetic distance between cultivars or varieties, so we must use the terms pairwise distance or dissimilarity value. Pairwise distances were first obtained by a maximum-likelihood criterion due to its conceptual simplicity, its well-defined statistical basis, and its strength against violations of the assumptions used in the underlying model (Huelsenbeck, 1995). Then pairwise distances based on the maximum-likelihood method were clustered with the UPGMA algorithm (Figure 2A), separating the olive-tree cultivars into three main groups. A quartet-puzzling analysis (Strimmer & von Haesler, 1996) was performed to asses the certainty of the main branches. Group I (90% of certainty) comprises the two *sylvestris* cultivars ('Acebuche fino' and 'Acebuche basto') and two cultivars that have been introduced in the last 60 years, whose origin is placed in the north-east of Spain ('Arbequina' and 'Blanqueta'). The mean dissimilarity value of this group is 0.73 ranging from 0.14 to 1.34. Group II (83% of certainty) contains native cultivars that are cultivated in a region that possesses a special climate and soil: 'Verdial de la Axarquía', 'Nevadillo Blanco' and 'Picudo de la Axarquía'. The mean dissimilarity distance of this group is 0.5, ranging from 0.21 to 0.71. The native cultivar 'Aloreña' might be included in this group, increasing the mean dissimilarity value to 0.61. The other native cultivar ('Gordal') is not included in this group, although it seems related to 'Aloreña' (see below). The main difference between the five native cultivars is that those of group II are for oil production while 'Aloreña' and 'Gordal' are table olive cultivars. The third group (58% of certainty) includes the remaining of cultivars found in

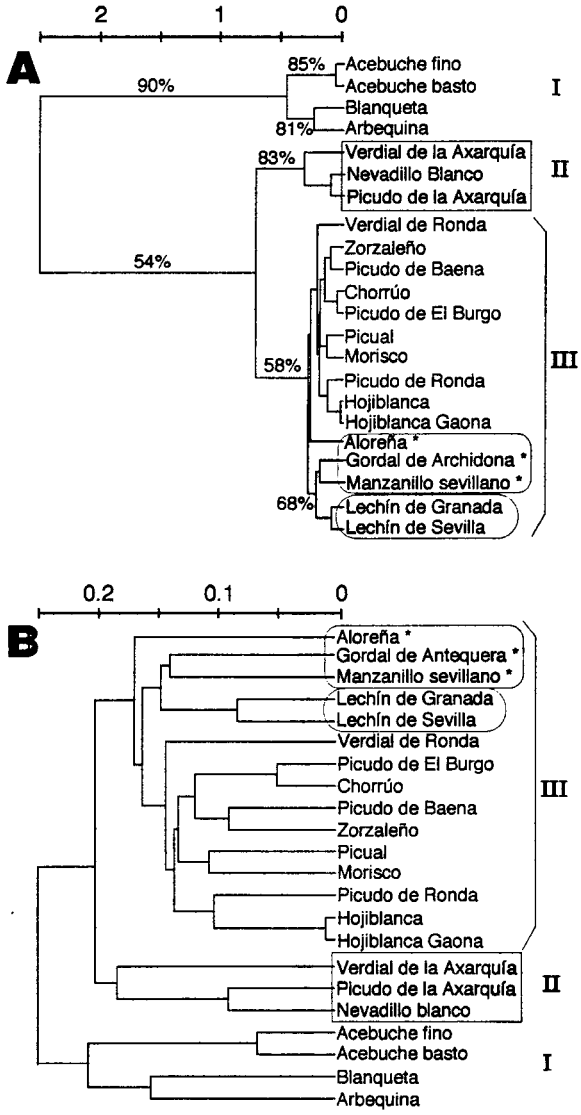

Figure 2. Phenograms of the 22 olive-tree cultivars differentiated in this work, generated by UPGMA cluster analysis of the pairwise distances. Distances were obtained by the maximum-likelihood algorithm (A) or Apostol index (B), giving in A the certainties of the main branches obtained by a quartet-puzzling analysis. Groups I, II and II are identified, subclusters of group III are marked with dashed lines, and cultivars that are used for table olive are marked with a * (see text for details).

Málaga with a mean dissimilarity of 0.62 (this value decreases to 0.47 if 'Aloreña' is not included). The small differences between dissimilarity values of this group accounts for the star-like distribution observed in the likelihood-mapping.

In order to decrease the possible bias introduced by the use of a single criterion for a pairwise analysis, we have performed the same analysis with two other unrelated criteria. Dice index was chosen since it is equivalent to Nei (Nei & Li, 1979) and Jaccard (Jaccard, 1908) indexes which are widely used in phenetic analyses. Apostol's simple matching index was selected since it was designed for RAPD analyses and provides similar results to others like Ochiai index (Ochiai, 1957). Since the differences obtained with Dice and Apostol indexes are minimal (both vary from 0 to 1), only the second is shown in Figure 2B. Both indexes can discriminate the same groups that were described with the maximum-likelihood criterion although the dissimilarity values are different. The mean dissimilarities were 0.37 and 0.32 with Dice index, and 0.34 and 0.29 with Apostol index. The intermediate case of 'Aloreña' is also observed: in Figure 2B the position is clearly intermediate among groups II and III, even if 'Aloreña' seems closer to group III, near to 'Gordal'. From the dissimilarity values of Dice and Apostol it can be suggested that 'Aloreña' and wild type are the most distantly related cultivars (dissimilarity of 0.73).

Some subclustering should be pointed out, with a 68% of certainty, accordingly with the three distance matrices calculated. The three cultivars that are used for table olive ('Aloreña', 'Gordal' and 'Manzanillo') are branched in the same point, thus they can form a subgroup. There are also two cultivars used for olive oil production that are always grouped: the two different 'Lechín' that can be found in very different locations in Málaga. It can be clearly seen in Figure 2 that both subgroups (pickling olive and 'Lechín') appear clearer and in an intermediate position between groups II and III. Conversely, it can be observed that neither the names 'Picudo' nor 'Verdial', based on morphological characteristics of the olive, can be used to group any cultivar.

In the seek of some phylogenetic relation of the 22 olive-tree cultivars identified in this study, the pairwise distances were analysed with the Neighbor-Joining method (Figure 3). It can be observed that groups I, II and III are confirmed. Furthermore, groups I and II seem to respond to the tree-like distribution of cultivars inferred by the likelihood-mapping, while group III seems to account for the star-like tendency. Group II appears in an intermediate stage between groups I and III.

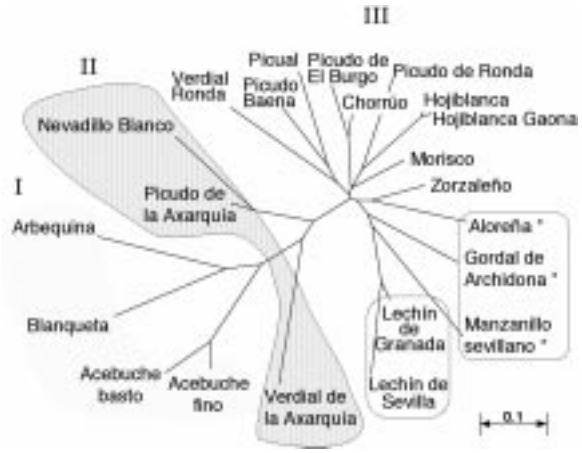

Figure 3. Phylogenetic tree generated with the Neighbor-Joining method with pairwise distances obtained with Dice index. The same three groups and subgroups are identified as described in Figure 2.

Discussion

Method reliability

Although olive-tree is an ancient cultivated crop, very little information is known at the molecular level about the different cultivars used. In an attempt to shed light on cultivar identification and biodiversity, we have developed a reproducible method for olive-tree DNA extraction, and used it to find RAPD and AP-PCR markers useful for olive-tree identification. The advantage of fingerprint analyses for the indirect selection of traits is diminished by the problem of reproducibility between laboratories. This fact is mostly dependent on exacting PCR conditions, and the quantity and quality of DNA used. That is why DNA of sufficient quality for PCR amplification may not provide reproducible results with RAPD and AP-PCR (Jones et al., 1997). We have developed a DNA extraction method that increases the reproducibility of the fingerprinting as reflected in Figure 1. The presence of bands larger than 1.5 kb confirms that polysaccharides, if present, do not interfere with polymerisation (Pandey et al., 1996). The absence of spurious bands suggests that the non-specific priming of RNA is effectively absent while this is not the case when PEG precipitation is not carried out. These results minimise doubts about the method's reproducibility under dif-

ferent experimental conditions. In conclusion, we have developed a reliable and reproducible method to detect RAPD and AP-PCR polymorphisms in olive-tree leaves.

Several reasons have led us to use bulk, instead of individual, DNA to characterise the olive-tree cultivars. (i) A genotype is considered as a group of individuals that has been selected for expression of specific traits in a background of otherwise randomly distributed genetic variation. Then, a bulk sample of 10 individuals may be expected to represent the markers linked to these traits (Yang & Quiros, 1993). This is especially evident in the case of olive-tree cultivars that are mainly propagated vegetatively or by grafting, and whose genotype variation seems to be low (Ouazzani et al., 1996). Therefore, comparison of RAPD profiles of bulk DNA samples of several genotypes for a single cultivar may reveal markers that distinguish between them in a significant number of cultivar members. (ii) It can be assumed that polymorphisms revealed in a fingerprint are present in at least 35% of the individuals assessed (Dulson et al., 1998). Finally, (iii) it has been described that 20 bands and 10 individuals are needed to detect a difference between two populations with a type-I error of 5%, a power of 90%, and distance less than 0.1 (Ghérardi et al., 1998). Taking into account all these considerations, we have sampled 10 individuals in each geographic site and we have analysed up to 62 polymorphic bands to obtain significant phenetic distances representative of a cultivar genetic background. We have defined the fingerprint of each genotype by multiple bands presumably at multiple genetic loci, or at least, at multiple DNA regions. This is important for cultivar characterisation since wild type and 'Verdial de la Axarquía' varieties could be defined by the presence of a single marker, 'Blanqueta' by the absence of one marker, and the rest by a set of several markers. The high level of polymorphism probably reflects the outcrossing nature of olive, in accordance with results described in (Fabbri et al., 1995) for other olive cultivars and confirms results obtained with allozymes (Pontikis et al., 1980; Loukas & Krimbas, 1983; Trujillo et al., 1990; Ouazzani et al., 1993; Ouazzani et al., 1996). It is noteworthy that this work shares several primers with other previous articles (Fabbri et al., 1995; Rubio & Arús, 1997), but does not always provide identical results. The differences can be due to homonyms in studied cultivars, the processivity of the EcoTaq polymerase we have used, and the decreased presence of polysaccharides and RNA in our DNA samples. The most striking differences were obtained with primer OPJ6 which was described as giving two polymorphisms of 1.5 and 1.4 kb (Rubio & Arús, 1997) while we have found additional polymorphic bands (Table 2) that Rubio and Arús do not describe.

Olive-tree cultivar classification

Three unrelated algorithms have been used to cluster the cultivars. The results obtained with all of them were nearly indistinguishable and of similar accuracy (Figures 2 and 3). This strongly supports the consistency of the obtained distance matrices. In any case, the phenograms and phylogenetic trees separate the cultivars into three main groups. The validation obtained by a quartet-puzzling strategy strengthens the evidence for the existence of these groups (Figure 2A). Group I is the most differentiated one, containing the wild type olive-tree and the two outsider cultivars 'Arbequina' and 'Blanqueta'. The dissimilarity values between the four cultivars suggest that 'Arbequina' and 'Blanqueta' retain more wild type traits than others, and not that non-Andalusian origin is as far scored as the wild type: pairwise distance between 'Arbequina' to 'Blanqueta' is the same that between 'Blanqueta' and wild types. Although this could seem strange, Angiolillo and co-workers (1999) have found that 'Arbequina' is very close to other wild *Olea* species and it belongs to a different cluster than other cultivars considered in this work. Moreover, phylogenetic tree also gathers these 4 cultivars (Figure 3). Group II includes three native cultivars that phenotypically are very different: 'Verdial' is a big tree that gives large-sized olives which never become black on maturity, while 'Nevadillo' corresponds to small trees with small, numerous olives of clear-green colour that become black on maturity. On the basis of their morphology, these trees were considered unrelated, even though they produce indistinguishable olive oils. The three cultivars share the climate and soil of their natural environments, making them so location-dependent that they are not profitable outside their traditional orchards. Group III is heterogeneous and includes trees of different portage, of distinct olive size and form, and originating in a number of Andalusian regions. The common feature of these cultivars is that they are cultivated under different soils and climates without dramatic changes, in contrast to group II. Two main subgroups can be identified here, namely the cultivars used for pickling olive, and the so named 'Lechín' cultivars. These groups are phenetically and phylogenetically localised

in the intermediate position between groups II and III (Figures 2 and 3) although maximum-likelihood method suggests that group II is not intermediate but evolved from group III, while the intermediate position is occupied by the 'Lechín' subgroup (data not shown). This could account for the low reliability (58% of certainty) of the main branch for the third group, but further analysis should be performed to clarify this aspect. Perhaps one can think that group III should be divided into other groups according to some morphological traits to increase the level of certainty.

In conclusion, we have shown that RAPD and AP-PCR can be used to distinguish olive-tree cultivars since we have confirmed identities among cultivars with high reliability. As far as we know, this is the first time that maximum-likelihood has been applied to analyse a fingerprinting, reinforcing the consistency of the proposed classification. In addition, the same conclusion was arisen with other two independent methods to calculate distances, and two phenological and phylogenetic criteria for clustering.

Influence of the geographic origin in olive-tree varieties

The classification presented in this work suggests that groups I, II and III can be discriminated by their geographic origin. Morphological traits can be used to subcluster the most heterogeneous group. The fact that wild type (group I) and varieties cultivated in physical proximity (groups II and III) belong to distinct clusters should rule out the hypothesis that the *sylvestris* variety was obtained by naturalisation of *sativa* varieties (Chevalier, 1948). Moreover, native or local cultivars seem to cluster together, providing evidence that the soil and climate (indirectly geographic origin) have significant influences on their differentiation along the years, while recently introduced cultivars ('Arbequina' and 'Blanqueta') seems to be very distinct from other cultivars. The same behaviour was observed for wild and cultivated olive-trees when Wiesman and co-workers (1998) studied olive-tree cultivars in Israel, or Angiolillo and co-workers (1999) studied genetic variation within and among the *Olea* genus. As mentioned, morphological traits can be the discriminant for the subgroups made from the third branch of the tree (Figures 2 and 3), or even to justify the clustering of two native cultivars in group III. Because geographic origin and morphological traits contribute to cultivar differentiation, our results are in partial agreement with classifications explained exclusively by morphological traits (Ouazzani et al., 1993; Fabbri et al., 1995) or the ones that consider geographical origin only (Trujillo et al., 1990). However, our results strengthen the conclusions obtained in studies that use olive-tree cultivars under environmental rationale (Wiesman et al., 1998; Angiolillo et al., 1999). In conclusion, the present study is consistent with the hypothesis of autochthonic origin of most cultivars (Zohary & Hopf, 1994): although the olive-tree was domesticated early, only cultivars without geographic dependence were moved widely from region to region (in our case, the third group), while those that were strongly dependent on their geographic location reflect some kind of affinity (in our case, group II). This is also consistent with the find that likelihood-mapping proposes a 55.2% tendency for a tree-like phenogram (a common ancestor for the authochthonic cultivars) and a 33.6% tendency for a star-like phenogram (cultivars that do not share a recent ancestor, group III in Figure 3).

The certainty in differentiation of group II by means of phenetic and phylogenetic analyses has been used to promote a Label of Origin (D.O.Ca. Axarquía) for the olive oil produced by the cultivars included in this group. Increasing efforts in developing this kind of molecular marker can also help in olive-tree breeding since, until recently, improvement of olive-trees was rather limited owing to their long juvenility phase and the lack of knowledge of phylogenetic relationships among *Olea* species. These markers could always be exploited to reduce the cost of the conservation of olive-tree germplasm collections (del Río & Caballero, 1994), since the establishment of a molecular marker database will provide important information to assign unknown genotypes to previously known cultivars, and select a number of accessions that can be redundant germplasm and then, be eliminated from the collection. Studies like the one presented here or the study of the genomic organisation of repetitive sequences (Karsiotis et al., 1998) provide new insights that can (i) help in selecting improved olive-tree genotypes, and (ii) guide which varieties may be successfully introduced in new environments.

Acknowledgements

We are indebted to Luis Martín González for his invaluable help in olive-tree cultivar identification and harvesting, and to A. de Grey for the English improvement. We are also grateful to the Research Services

of Málaga University for providing facilities at the Molecular Biology Laboratory. This work was supported by Grants from Consejería de Agricultura y Pesca (C97-082) and Consejería de Educación (Research group CVI-0114) from the Junta de Andalucía. MGC is a recipient of a Contrato de Reincorporación de Doctores y Tecnólogos from the Ministerio de Educación y Cultura (Spain).

References

Adams, R.P. & L.H. Rieseberg, 1998. The effects of non-homology in RAPD bands on similarity and multivariate statistical ordination in *Brassica* and *Helianthus*. Theor Appl Genet 97: 323–326.

Amstrong, J., A. Gibbs, R. Peakall & G. Weiller, 1994. The RAPDistance package. http://life.anu.edu.au/molecular/sofware/rapd.html,

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

Apostol, B.L., W.C. Black, B.R. Miller, P. Reiter & B.J. Beaty, 1993. Estimation of the number of full sibling families at an oviposition site using RAPD-PCR markers: applicationss to the mosquito *Aedes aegypti*. Theor Appl Genet 86: 991–1000.

Asturias, J.A., M.C. Arilla, N. Gomez-Bayón, J. Martínez, A. Martínez & R. Palacios, 1997. Cloning and expression of the panallergen profilin and the major allergen (Ole e 1) from olive tree pollen. J Allerg Clin Immunol 100: 365–372.

Barker, J.H.A., M. Matthes, G.M. Arnold, K.J. Edwuards, I. Åhman, S. Larsson & A. Karp, 1999. Characterisation of genetic diversity in potential biomass willows (*Salix* spp.) by RAPD and AFLP analyses. Genome 42: 173–183.

Barranco, D. & L. Rallo, 1984. Las variedades de olivo cultivadas en Andalucía. Consejería de Agricultura y Pesca de la Junta de Andalucía – MAPA, Madrid.

Batanero, E., A. Ledesma, M. Villalba & R. Rodríguez, 1997. Purification, amino acid sequence and immunological characterization of Ole e 6, a cysteine enriched allergen from olive tree pollen. FEBS Lett 410: 293–296.

Belaj, A., I. Trujillo, R. de la Rosa & L. Rallo, 1998. Selection of RAPD markers for olive (Olea europaea L.) cultivars identification. Acta Hort in press.

Bogani, P., D. Cavalieri, R. Petruccelli, P.L. & G. Roselli, 1994. Identification of olive tree cultivars by using random amplified polymorfic DNA. Acta Hort 356: 98–101.

Brousse, G., 1987. Olive. In: G. Robbelen, R.K. Downey & A. Ashri (Eds.), Oil crops of the world, their breeding and utilization, pp. 462–474. McGraw Hill Publising Company, New York.

Chevalier, A., 1948. L'origine de l'olivier cultivé et ses variations. Rev Internat Bot Appl Agric Trop 28: 1–25.

Claros, M.G. & F.M. Cánovas, 1998. Rapid high quality RNA preparation from pine seedlings. Plant Mol Biol Rept 16: 9–18.

del Río, C. & J.M. Caballero, 1994. Caracterización agronómica preliminar de las variedades introducidas en el banco de germoplasma de olivo de Córdoba en 1987. Frutic Prof 62: 9–15.

Demeke, T., R.P. Adams & R. Chibbar, 1992. Potential taxonomic use of random amplified polymorphic DNA (RAPD): a case study of *Brassica*. Theor Appl Genet 84: 990–994.

Dice, L.R., 1945. Measure of the amount of ecological association between species. Ecology 26: 297–302.

Doyle, J.J. & J.L. Doyle, 1987. A rapid isolation procedure for small quantities of fresh leaf tissue. Phytochem Bul 19: 11–15.

Dulson, J., L.S. Kott & V.L. Ripley, 1998. Efficacy of bulked DNA samples for RAPD DNA fingerprinting of genetically complex *Brassica napus* cultivars. Euphytica 102: 65–70.

Fabbri, A., J.I. Hormaza & V.S. Polito, 1995. Random amplified polymorphic DNA analysis of olive (*Olea europaea* L) cultivars. J Amer Soc Hort Sci 120: 538–542.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17: 368–376.

Felsenstein, J., 1988. Phylogenies from molecular sequences: inference and reliability. Annu Rev Genet 22: 521–565.

Felsenstein, J., 1993. PHYLIP (Phylogeny Inference Package) version 3.5c. Seattle, Distributed by the author. Department of Genetics, University of Washington.

Ferreira, A.R. & P. Keim, 1997. Genetic mapping of soybean [*Glycine max* (l.) Merr.] using random amplified polymorfic DNA (RAPD). Plant Mol Biol Rept 15: 335–354.

Ghérardi, M., B. Mangin, B. Goffinet, D. Bonnet & T. Huguet, 1998. A method to measure genetic distance between allogamous populations of alfalfa (*Medicago sativa*) using RAPD molecular markers. Theor Appl Genet 96: 406–412.

Gower, J.C., 1985. Measures of similarity, dissimilarity and distance. In: S. Klotz & N.L. Johnson (Eds.), Encyclopedia of Statistical Sciences vol 5, pp. 397–405. Wiley, New York.

Green, P.S. & G.E. Wickens, 1989. The *Olea europaea* complex. In: K. Tan (Ed.), The Davis and Hedge Festschrift, pp. 287–299. Edinburgh University Press, Edinburgh.

Haralampidis, K., D. Milioni, J. Sanchez, M. Baltrusch, E. Heinz & P. Hatzopoulos, 1998. Temporal and transient expression of stearoyl-ACP carrier protein desaturase gene during olive fruit-development. J Exp Bot 49: 1661–1669.

Huelsenbeck, J.P., 1995. Performance of phylogenetic methods in simulation. Syst Biol 44: 17–48.

Jaccard, P., 1908. Nouvelles recherches sur la distribution florale. Bull Soc Vaud Sci Nat 44: 223–270.

Jones, C.J., K.J. Edwards, S. Castaglione, M.O. Winfield, F. Sala, C. van de Wiel, G. Bredemeijer, B. Vosman, M. Matthes, A. Daly, R. Brettschneider, P. Bettini, M. Buiatti, E. Maestri, A. Malcevschi, N. Marmiroli, R. Aert, G. Volckaert, J. Rueda, R. Linacero, A. Vázquez & A. Karp, 1997. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Mol Breeding 3: 381–390.

Karsiotis, A., M. Hagidimitriou, A. Douka & P. Hatzopoulos, 1998. Genomic organization, sequence interrelationship, and physical localization using in situ hybridization of two tandemly repeated DNA sequences in the genus *Olea*. Genome 41: 527–534.

Lavee, S., 1990. Aims, methods and advances in breeding of new olive (*Olea europaea* L.) cultivars. Acta Hort 286: 26–36.

Loukas, M. & C.B. Krimbas, 1983. History of olive cultivars based on their genetic distances. J Hort Sci 58: 121–127.

Minet, M., M.E. Dufour & F. Lacroute, 1992. Complementation of *Saccharomyces cerevisiae* auxotrophic mutants by *Arabidopsis thaliana* cDNAs. Plant J 2: 417–422.

Nei, M. & W.H. Li, 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc Natl Acad Sci USA 76: 5269–5273.

Nicese, F.P., J.I. Hormaza & G.H. McGranahan, 1998. Molecular characterization and genetic relatedness among walnut (*Juglans regia* L.) genotypes based on RAPD markers. Euphytica 101: 199–206.

Ochiai, A., 1957. Zoogeographic studies on the soleoid fishes found in Japan and its neighbouring regions. Bull Jap Soc Sci Fish 22: 526–530.

Ouazzani, N., R. Lumaret & P. Villemur, 1996. Genetic variation in the olive tree (*Olea europaea* L.) cultivated in Morocco. Euphytica 91: 9–20.

Ouazzani, N., R. Lumaret, P. Villemur & F. Di Giusto, 1993. Leaf allozyme variation in cultivated and wild olive trees (*Olea eurpaea* L.). J Heredity 84: 34–42.

Page, R.D.M., 1996. TreeView: an application to display phylogenetic trees on personal computers. Comput Applic Biosci 12: 357–358.

Pandey, R.N., R.P. Adams & L.E. Flournoy, 1996. Inhibition of random amplified polymorfic DNAs (RAPDs) by plant polysaccharides. Plant Mol Biol Rept 14: 17–22.

Pontikis, C.A., M. Loukas & G. Kousounis, 1980. The use of biochemical markers to distinguish olive cultivars. J Hort Sci 55: 333–343.

Rallo, L., 1994. Evaluación agronómica y obtención de nuevas variedades de olivo. Frutic Prof 62: 17–27.

Rubio, M.J. & P. Arús, 1997. Un vivero español, 'Agromillora Catalana', aplica en su producción de planta de olivo un control de calidad genético y sanitario basado en las nuevas tecnologías de RAPDs y ELISA-DAS. Frutic Prof 88: 14–18.

Saitou, N. & M. Nei, 1987. The neighbor-joining method: a new method for reconstructing filogenetic trees. Mol Biol Evol 4: 406–425.

Strimmer, K. & A. von Haeseler, 1996. Quartet-puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Mol Biol Evol 13: 964–969.

Strimmer, K. & A. von Haeseler, 1997. Likelihood-mapping: a simple method to visualize phylogenetic content of a sequence alignement. Proc Natl Acad Sci USA 94: 6815–6819.

Trujillo, I., L. Rallo, E.A. Carbonell & M.J. Asins, 1990. Isoenzymatic variability of olive cultivars according to their origin. Acta Hort 286: 137–140.

Vergari, G., M. Patumi & G. Fontanazza, 1996. Utilización de los marcadores RAPDs para la caracterización del germoplasma de olivo. Olivae 60: 19–22.

Villalba, M., E. Batanero, R.I. Monsalve, M.A.G. Delapena, C. Lahoz & R. Rodriguez, 1994. Cloning and expression of Ole e 1, the major allergen from olive tree pollen – Polymorphism analysis and tissue-specificity. J Biol Chem 269: 15217–15222.

Welsh, J. & M. McClelland, 1990. Fingerprint genomes using PCR with arbitrary primers. Nucl Acids Res 18: 7213–7218.

Wiesman, Z., N. Avidan, S. Lavee & B. Quebedeaux, 1998. Molecular characterization of common olive varieties in Israel and the West-Bank using randomly amplified polymorphic DNA (RAPD) markers. J Amer Soc Hort Sci 123: 837–841.

Williams, J.G.K., M.K. Hanafey, J.A. Rafalski & S.V. Tingey, 1993. Genetic analysis using random amplified polymorphic DNA markers. Methods Enzymol 218: 704–740.

Yang, X. & C. Quiros, 1993. Identification and classification of celery cultivars with RAPD markers. Theor Appl Genet 86: 205–212.

Zohary, D. & M. Hopf, 1994. Olive: *Olea europaea*. In: Domestication of plants in the Old World, pp. 137–143. Clarendon Press, Oxford.