Theor Appl Genet (2001) 102:1082–1087

P. Hernández · R. de la Rosa · L. Rallo · A. Martín
G. Dorado

First evidence of a retrotransposon-like element in olive (*Olea europaea*): implications in plant variety identification by SCAR-marker development

Received: 31 July 2000 / Accepted: 22 September 2000

Abstract When developing SCARs by sequencing RAPD markers useful for olive variety identification, one RAPD sequence of 407 bp has been identified that shows significant DNA homology with more than 160 retrotransposon-like sequences. A generally coherent phylogenetic tree has been constructed based on the homologous retrotransposon-like sequences, reflecting genetic distances in 35 species belonging to eight kingdoms. The implications of this finding in the development of SCAR markers are discussed. In addition, specific oligonucleotide primers have been developed to amplify the DNA region and have a practical application as an internal amplification control in SCAR-based multiplexed PCRs. To our knowledge, this is the first report of a retrotransposon-like element in olive.

Keywords Kingdom and biological evolution · Phylogenetic tree · RAPD-PCR · Retrotransposon-like sequence · SCAR

Introduction

The olive tree, *Olea europaea*, is an important oil crop that has been traditionally cultivated throughout the Mediterranean basin (Bartolini et al. 1998). From there its cultivation spread to North and South America and to Australia. The current consideration of olive oil as a healthy source of fats has led to its incorporation into the diet in many countries outside the Mediterranean region. Since the beginning of its cultivation, olive improvement has only been carried out by the empirical breeding of growers (Rallo 1994). Efforts for the development of new varieties by systematic breeding have produced very little results (Lavee 1978; Lavee et al. 1986), mainly because it can only be approached in the long term.

The development of new varieties in fruit-tree species has always been a long-term task, due to the length of the juvenile period. Molecular markers have been recently proposed as good early selection techniques (marker-assisted selection) by their association with agronomic traits in trees (Arús and Olarte 1994). The identification of olive-tree cultivars has been traditionally carried out by morphological and agronomic traits. However, these are dependent on environmental or cultivation factors. Identification based on the analysis of gene products, such as isozymes, has also been used (Trujillo et al. 1995). Nevertheless, this is still limited by the relatively small number of polymorphisms detected, and by the possibility that isozyme expression can also be altered by environmental conditions. Nowadays, the practical use of genetic markers is predominantly performed through PCR-based techniques. As there is little or no sequence information of the olive genome available, random amplified polymorphic DNA (RAPD; Welsh and McClelland 1990; Williams et al. 1990) is a simple and reliable source of polymorphism that has already been used for olive-cultivar identification (Fabbri et al. 1995). The conversion of such markers in sequence-characterized amplified regions (SCARS, Paran and Michelmore 1993) by the development of longer, more specific primers from the RAPD sequences, significantly improves the reproducibility and reliability of PCR assays, and therefore their utility for marker-assisted breeding. However, depending on the specific oligonucleotide sequence, a certain percentage of monomorphic markers may be obtained when primers are extended based on cloned sequences (Paran and Michelmore 1993; Bodenes et al. 1997; Hernández et al. 1999a). We have used this technology to clone and sequence RAPD markers useful for olive-cultivar identification.

Communicated by H.F. Linskens

P. Hernández (✉) · A. Martin
Instituto de Agricultura Sostenible (CSIC), Apdo. 4084,
14080 Córdoba, Spain
e-mail: ge1hemop@uco.es
Tel.: +34-957-49910, Fax: +34-957-499252

R. de la Rosa · L. Rallo
Departamento de Agronomía, ETSIAM, 14080 Córdoba, Spain

G. Dorado
Departamento de Bioquímica y Biología Molecular,
Campus Universitario Rabanales, 14071 Córdoba, Spain

Material and methods

Plant material

Six *O. europaea* L. cultivars from the Germplasm Bank of Córdoba 'Arbequina', 'Frantoio', 'Gordal Sevillana', 'Manzanilla de Sevilla', 'Lechín de Sevilla' and 'Picual', were employed.

DNA extraction

DNA was extracted from young leaf tissue using the CTAB method (Murray and Thompson 1980) with some modifications. Tissue (0.7–1 g) was ground in liquid N_2. Before the tissue thawed, 3 ml of CTAB buffer containing 0.5% Na bisulphite and 25 mM DTT were added. The solution was incubated for 40 min at 65°C with occasional mixing. After the tissue was extracted once with the same volume of 24:1 chloroform/octanol, the aqueous phase was isopropanol-precipitated and spooled out using a glass hook, washed in 70% ethanol and air-dried. The DNA was dissolved in approximately 500 µl of TE buffer. RNase was added (0.03 mg) and the DNA was dissolved overnight at 4°C. The obtained DNA stock solution was diluted 1:20 in water for PCR-amplification.

DNA amplification

Approximately 20–40 ng of genomic DNA were used in amplification reactions, following the protocol described by Hernández et al. (1996). Random 10-mer primers were purchased from Operon Technologies (Alameda, Calif., USA) and the *AmpliTaq* DNA Polymerase Stoffel Fragment from P-E Biosystems (Foster City, Calif., USA). Amplification was performed in a System 9600 cycler from the same manufacturer. The amplified products were resolved by electrophoresis on gels consisting of 1% (w/v) SeaKem agarose : 1% (w/v) NuSieve agarose from FMC (Rockland, Me., USA), and TBE buffer. Gels were stained with 0.5 µg of ethidium bromide/ml, and photographed with a GDS 5000 system CCD camera from UVP (Cambridge, UK).

Cloning and sequencing of the RAPD product

The RAPD band was excised from agarose gels and soaked in 20 µl of sterile distilled milliQ water. A 1-µl aliquot was re-amplified using the same PCR reaction mixture that was originally employed to generate the RAPD, but the number of cycles was reduced to 25. The amplified product was checked on an agarose gel, purified by selective precipitation and end-polished with *Pfu* DNA polymerase from Stratagene (La Jolla, Calif., USA). The blunt-ended PCR products were inserted into the *p*CR-*Script* vector following the procedures of the *p*CR-*Script* Amp SK (+) Cloning Kit from the same manufacturer. Two different clones were cycle-sequenced using the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit with *AmpliTaq* DNA Polymerase, FS from P-E Biosystems, using the automated ABI 373 Stretch DNA sequencer from the same manufacturer and optimized electrophoresis conditions as described by Lario et al. (1997). Double-stranded DNA sequencing was carried out for both clones using both the –21M13 and M13 forward/reverse primers respectively, to obtain the consensus sequence.

Designing STS primers and subsequent specific-amplification of genomic regions

Based on the sequences of the cloned RAPD products, pairs of oligonucleotide primers were designed using the programs Oligo 5.0 for MacOS from Molecular Biology Insights (Plymouth, Minn., USA) and PrimerSelect 3.03 for MacOS from DNAStar (Madison, Wis., USA) for specific amplification of the loci identified by RAPD markers. Primers were synthesized by Operon Technologies.

Amplification reactions were carried out in 20-µl solutions containing 20–40 ng of DNA, 0.5 units of *AmpliTaq* Gold DNA Polymerase from P-E Biosystems, 100 µM of each dNTP from Roche (Basel, Switzerland), 0.16 µM of each primer, 2.5 mM $MgCl_2$, 50 mM KCl and 10 mM Tris-HCl, pH 8.3. Amplification was carried out under the following conditions: 94°C for 10 min; 35 amplification cycles (15-s denaturation at 94°C 20-s annealing at 69°C, 1 min of extension at 72°C); and a final extension of 6 min at 72°C. Amplified products were resolved by electrophoresis in 2% (w/v) agarose gels as previously described.

Sequence similarity searches

These were carried out for both nucleotide and peptide (i.e. all six translated frames) sequences, using the Basic Local Alignment Search Tool (BLAST) against the RAPD marker (GenBank accession number AF130732). Analyses were performed with the ungapped BlastN 1.4.11 (Altschul et al. 1990). The searches were carried out at The National Center for Biotechnology Information via the internet at <http://www.ncbi.nlm.nih.gov>, both directly, as well as via the GeneQuest 4.00/99 module of LaserGene for Mac OS from DNAStar. Sequences were compared against the "nr" database: all non-redundant GenBank+EMBL+DDBJ+PDB Nucleotide Sequence Databases (NSD); and all non-redundant GenBank CDS translations+PDB+SwissProt+PIR+PRF Peptide Sequence Databases (PSD). The database was posted on August 30, 1999. The query sequence was filtered for low-complexity regions by default. The parameters for BlastN 1.4.11 were the default settings corresponding to: V=100; B=50; H=0; Lambda=0.192; K=0.174 or 0.175; H=0.358 or 0.359; E=10.0 or 120; S=125; T=0; X=73. Sequence comparisons and cluster diagrams for the RAPD marker OPJ5-J6–2, produced by the primer mixture OPJ5 + OPJ6 were obtained using the MegAlign program for Mac OS from DNAStar.

Results and discussion

The identification of the six olive cultivars used in this study has been successfully performed by RAPD analysis. As an example, amplification patterns obtained using

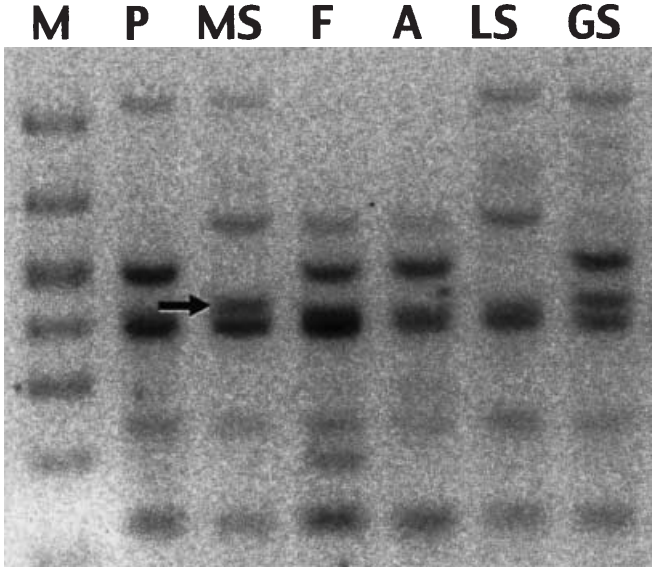


Fig. 1 RAPD amplification pattern. RAPD amplification of five olive cultivars using the pairwise primer combination OPJ5+OPJ6. *Lanes* are as follows: *M* molecular-weight marker; *P* cv 'Picual'; *MS* cv 'Manzanilla de Sevilla'; *F* cv 'Frantoio'; *A* cv 'Arbequina'; *LS* cv 'Lechín de Sevilla', *GS* cv 'Gordal Sevillana'. The 'Manzanilla de Sevilla'-specific RAPD OPJ5-J6–2 is indicated by an *arrow*

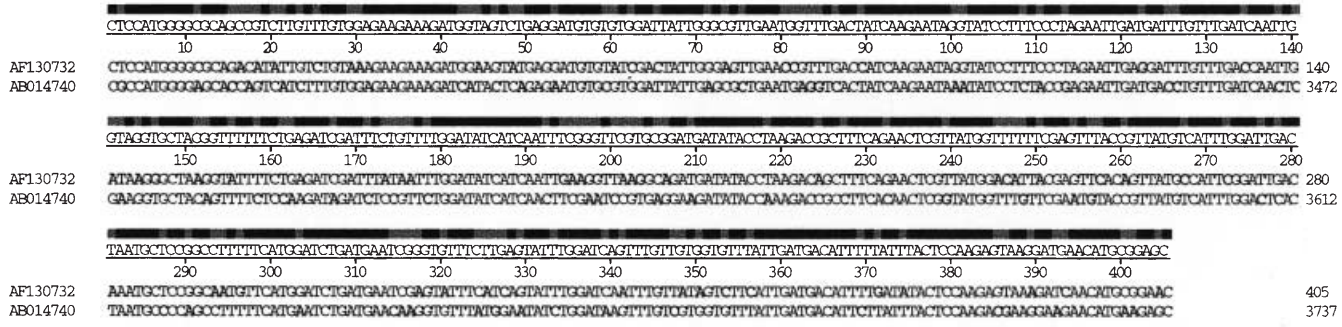


Fig. 2 Wilbur–Lipman DNA alignment. The putative ORF of the RAPD sequence OPJ5-J6–2 (*O. europaea*, GenBank acc. No. AF130732) shows a 68% similarity index over a 405-bp residue consensus length with the *O. sativa* gypsy type retrotransposon *RIRE8A* (GenBank acc. No. AB014740). The alignment settings used were as follows: Ktuple: 3; Gap Penalty: 999; Window: 1

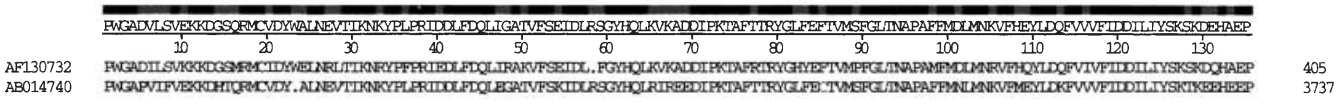


Fig. 3 Lipman-Pearson protein alignment. The translated putative ORF of the RAPD sequence OPJ5-J6–2 (*O. europaea*, GenBank acc. No. AF130732) shows a 68% similarity index over a 135-bp residue consensus length with the retrofit protein of *O. sativa* gypsy type retrotransposon *RIRE8A* (GenBank acc. No. AB014740). Yet the translated sequence is truncated. The alignment settings used were as follows: Ktuple: 2; Gap Penalty: 999; Gap Length Penalty: 999

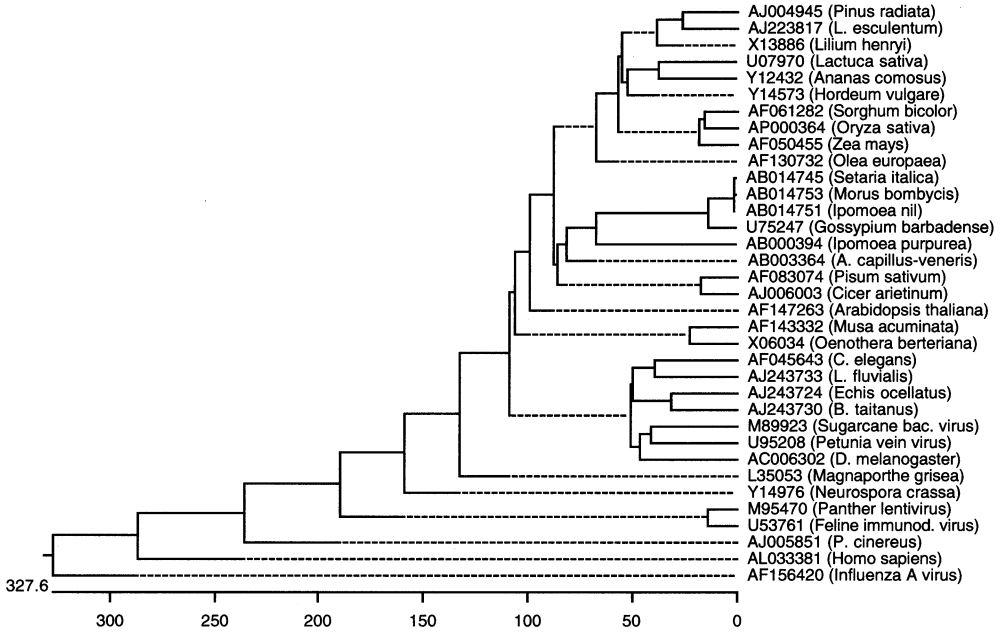


Fig. 4 Phylogenetic tree using the Clustal method with a Weighted residue weight table. Schematic representation obtained using a 405-bp olive RAPD sequence (GenBank acc. No. AF130732) and 34 homologous retrotransposon-like DNA sequences from other plants, virus, fungus, worm, insect, fish, amphibian, reptile and mammal.

the 10-mer primer pairwise combination OPJ5+OPJ6 is shown in Fig. 1. RAPD-PCR is a widely used technique in plant genome analysis but, still, little is known about the nature of these markers. Sequencing of RAPD products with the aim of converting them into more specific assays (Paran and Michelmore 1993) is an additional source of information. A database search of the OPJ5-J6–2 (407-bp) sequence detected significant homology to more than 160 gypsy type retrotransposon-related sequences. These correspond to 20 different plant genera, and include, as well, identities with virus, fungus, worm, insect, fish, amphibian, reptile and mammal DNA sequences. Twenty seven of the homologies show DNA identity levels ranging from 65 to 70% over a DNA stretch of 405 bp with gypsy type retrotransposon sequences from the genera *Oryza*, *Sorghum*, *Pisum*, *Arabidopsis*, *Ananas*, *Lilium* and *Cicer*. As an example, sequence identity with the *Oryza sativa* gypsy type retrotransposon *RIRE8A* (68% identity) is shown in Fig. 2. These are the first data suggesting the presence of retrotransposon-like elements in the olive genome. More significantly, comparisons at the amino-acid sequence level

also indicate similarities to reverse transcriptase in other plant retrotransposons. These data suggest that either this retrotransposon-like sequence has not experienced much divergence during evolution or, alternatively, that the original mobile element was introduced into such different hosts via some unknown horizontal transfer mechanism. Figure 3 shows the amino-acid alignment corresponding to the DNA sequences of Fig. 2. Interestingly, the identity (68%) is maintained at the amino-acid level as well, with most divergences found at the third base of each triplet codon. This is clear evidence of the biological significance of such similarities. The potential ORF in OPJ5-J6–2 is interrupted by one stop codon (Fig. 3). Stop codons are frequent in plant polyprotein-like sequences. An extreme case of 21 stop codons in the polyprotein reading frame of a plant retroelement has been reported for the *del* retroelement of *Lilium henryi* (Sayth et al. 1989). This retroelement shows a 66% DNA identity with the OPJ5-J6–2 sequence over a 399-bp DNA stretch. It is very likely that these defective copies of retroelements are inactive. In other words, they may be considered as remains of retrotransposon DNA sequences, and thus could be useful to determine evolutionary distances between different plant and other species containing such genetic evidence. The divergence between such DNA sequences in different species should reflect the basic general DNA genetic drift not subjected to the protein selection filter. As mobile genetic elements, retrotransposons are thought to be important forces in genome evolution (Kidwell and Lish 1997). Despite the sequence variation, conserved areas of retrotransposon DNA have been described (Lewin 1997).

The striking widespread distribution of homologous elements could indicate the existence of related sequences even before kingdom separation. Figure 4 shows the cluster diagram obtained by the Clustal method reflecting the genetic distances of the retroposon-like element in 35 species belonging to five kingdoms (plus virus). The sequence alignment is shown in the Appendix to Fig. 4. It is remarkable that such a generally coherent phylogenetic tree should be generated using just a short 405-bp DNA stretch. This is further evidence supporting the biological significance of the retrotransposon-like similarities described in this work. Additionally, it sheds new light on the cellular origin of viruses from both prokaryotic and eukaryotic organisms. Homology of RAPD products to retrotransposon-like elements has already been reported in grapevine (Böhm and Zyprian 1998) and has been the basis for the isolation and characterization of the *Tnd*–1 retrotransposon from *Nicotiana debneyi* (Kenward et al. 1999).

On the other hand, the BLAST homology search results explain the loss of specificity when developing a SCAR marker that amplified the six olive varieties used in this study (‘Arbequina’, ‘Frantoio’, ‘Gordal Sevillana’, ‘Manzanilla de Sevilla’, ‘Lechín de Sevilla’ and ‘Picual’), in relation to the polymorphic RAPD OPJ5-J6–2, which discriminated the variety ‘Manzanilla de Sevilla’. The monomorphic product SCOeMS-1, amplified by oligonucleotide primers 5´-CTCCATGGGGCGCAGACATATTGTC-3´ and 5´-TCGTTCCGCATGTTGATCTTTACTCTTG-3´, is produced in every olive variety that we have tested later and for a range of annealing temperatures. Therefore, the SCAR marker SCOeMS-1 is being used as an internal control of PCR amplification in olive SCAR-based multiplexed PCRs (see Fig. 5 for an example). Loss of polymorphism has also been reported in other plants (Paran and Michelmore 1993; Bodenes et al. 1997; Hernández et al. 1999a) when primers were extended based on cloned RAPD sequences. It has been considered that the RAPD polymorphism might be the result of mismatch at the priming sites, rather than DNA sequence divergence or structural rearrangements. Consequently, extending the primers would prevent such a putative polymorphism. Actually, we have found a clear example in which the OPJ5-J6–2 RAPD polymorphism relies on the 10-bp primer sequence. Most probably, it is based on the last two bases of the 3´ end of the RAPD sequence, given the DNA sequence conservation across species for the rest of it. The design of new primer pairs in internal sites has been a successful approach in cases where the sequence divergence is sufficient (Hernández et al. 1999b). Nevertheless, in this work, the BLAST results suggest that any attempt to amplify variety ‘Manzanilla de Sevilla’-specific internal DNA stretches of the RAPD region with longer, more-specific primers is likely to fail. We strongly recommend performing BLAST homology searches prior to developing such primers as an aid to avoid known conserved sequences. The utility of this approach will increase with time, as more plant sequence information becomes available. In addition, the results presented here support the view (Devos and Gale 1992) that RAPD markers include a significant proportion of repetitive DNA amplification. Therefore, they may have limited value for mapping purposes unless a selection of markers is applied, as for example in introgression approaches (Hernández et al. 1999a). It is remarkable that in the very early sequencing of such markers we were able to find a retrotransposon-like, and therefore most probably a repetitive, sequence. Further studies on repetitive and retrotransposon-like sequences will shed new light on their evolutionary implications, as well as their potential use in plant variety identification and breeding.

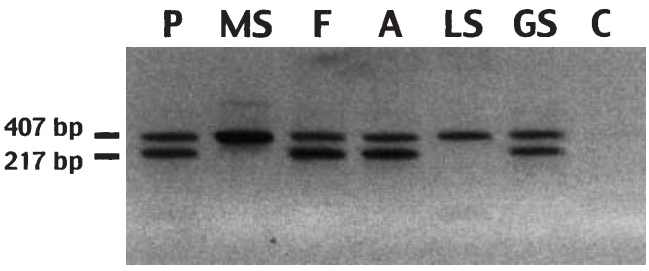
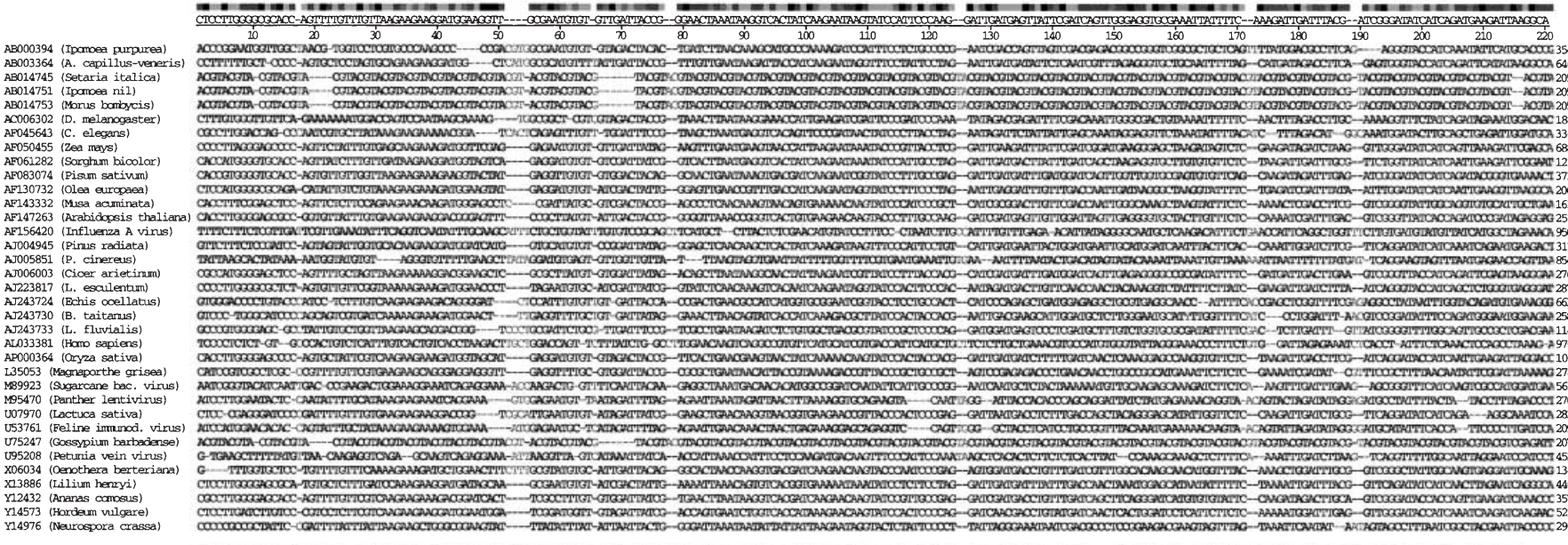


Fig. 5 Olive variety identification by multiplexed PCR. Coamplification of the SCAR SCOeMS-1 (407-bp fragment) as an internal amplification control and SCAR SCOeP-1 (217-bp fragment) is shown. *Lanes* are as follows: *P* cv ‘Picual’; *MS* cv ‘Manzanilla de Sevilla’; *F* cv ‘Frantoio’; *A* cv ‘Arbequina’; *LS* cv ‘Lechín de Sevilla’, *GS* cv ‘Gordal Sevillana’, *C* multiplexed reaction without template DNA (negative control of amplification)


AB000394 (Ipomoea purpurea)
AB003364 (A. capillus-veneris)
AB014745 (Setaria italica)
AB014751 (Ipomoea nil)
AB014753 (Morus bombycis)
AC006302 (D. melanogaster)
AF045643 (C. elegans)
AF050455 (Zea mays)
AF061282 (Sorghum bicolor)
AF083074 (Pisum sativum)
AF130732 (Olea europaea)
AF143332 (Musa acuminata)
AF147263 (Arabidopsis thaliana)
AF156420 (Influenza A virus)
AJ004945 (Pinus radiata)
AJ005851 (P. cinereus)
AJ006003 (Cicer arietinum)
AJ223817 (L. esculentum)
AJ243724 (Echis ocellatus)
AJ243730 (B. taitanus)
AJ243733 (L. fluvialis)
AL033381 (Homo sapiens)
AP000364 (Oryza sativa)
L35053 (Magnaporthe grisea)
M89923 (Sugarcane bac. virus)
M95470 (Panther lentivirus)
U07970 (Lactuca sativa)
U53761 (Feline immunod. virus)
U75247 (Gossypium barbadense)
U95208 (Petunia vein virus)
X06034 (Oenothera berteriana)
X13886 (Lilium henryi)
Y12432 (Ananas comosus)
Y14573 (Hordeum vulgare)
Y14976 (Neurospora crassa)

◀ **Appendix.** Alignment workspace using the Clustal method with a Weighted residue weight table. Schematic representation obtained of a 405-bp olive RAPD sequence (GenBank acc. No. AF130732) and 34 homologous retrotransposon-like DNA sequences from other plants, virus, fungus, worm, insect, fish, amphibian, reptile and mammal (GenBank acc. No. and spp. indicated).

Acknowledgements Financial support from the CICYT and INIA, Spain (Projects OLI96–2184 and CAO98–001-C3–2), is also acknowledged. The experiments carried out in this work comply with the current laws of Spain.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Arús P, Olarte C (1994) Linkage analysis of ten isozyme genes in F1 segregating almond progenies. J Am Soc Hort Sci 119: 339–344

Bartolini G, Prevost G, Messeri C, Carignani G (1998) Olive germplasm: cultivars and world-wide collections. FAO, Roma

Bodenes C, Joandet S, Laigret F, Kremer A (1997) Detection of genomic regions differentiating two closely-related oak species *Quercus-Petraea* (Matt) Liebl and *Quercus-Robur* L. Heredity 78:433–444

Böhm A, Zyprian E (1998) RAPD marker in grapevine (*Vitis* spp.) similar to plant retrotransposons. Plant Cell Rep 17:415–421

Devos KM, Gale MD (1992) The use of random amplified polymorphic DNA markers in wheat. Theor Appl Genet 84:567–572

Fabbri J, Ormaza I, Polito VS (1995) Random amplified polymorphic DNA analysis of olive (*Olea europaea* L.) cultivars. J Am Soc Hort Sci 120:538–542

Hernández P, Rubio MJ, Martin A (1996) Development of RAPD markers in tritordeum and addition lines of *Hordeum chilense* in *Triticum aestivum*. Plant Breed 115:52–56

Hernández P, Hemmat M, Weeden NF, Dorado G, Martín A (1999a) Development and characterization of *Hordeum chilense* chromosome-specific STS markers suitable for wheat introgression and marker-assisted selection. Theor Appl Genet 98:721–727

Hernández P, Martín A, Dorado G (1999b) Development of SCARs by direct sequencing of RAPD products: a practical tool for the introgression and marker-assisted selection of wheat. Mol Breed 5:245–253

Kenward KD, Bai D, Ban MR, Brandle JE (1999) Isolation and characterization of Tnd-1, a retrotransposon marker linked to black root rot resistance in tobacco. Theor Appl Genet 98: 387–395

Kidwell MG, Lish D (1997) Transposable elements as sources of variation in animals and plants. Proc Natl Acad Sci USA 94: 7704–7711

Lario A, González A, Dorado G (1997) Automated laser-induced fluorescence DNA sequencing: equalizing signal-to-noise ratios significantly enhances overall performance. Anal Biochem 247:30–33

Lavee S (1978) 'Kadesh' table olive. HortScience 13:62–63

Lavee S, Haskal A, Wodner M (1986) 'Barnea': a new olive cultivar from first-breeding generation. Olea 17:95–99

Lewin B (1997) Genes VI. Oxford University Press, Incorporated, New York

Murray YHG, Thompson WF (1980) Rapid isolation of high-molecular-weight plant DNA. Nucleic Acids Res 8:4321–4326

Paran I, Michelmore R (1993) Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. Theor Appl Genet 85:985–993

Rallo L (1994) Evaluación agronómica y obtención de nuevas variedades de olivo. In: Rallo L, Troncoso A (eds) Avances en olivicultura. Fundación La Caixa-Fruticultura Profesional-SECH, Barcelona, pp 17–27

Sayth DR, Kalitsis P, Joseph JL, Sentry JW (1989) Plant retrotransposon from *Lilium henryi* related to Ty3 of yeast and the gypsy group of *Drosophila*. Proc Natl Acad Sci USA 96: 5013–5019

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

Welsh J, McClelland M (1990) Fingerprinting genomes using PCR with arbitrary primers. Nucleic Acids Res 18:7213–7218

Williams JGK, Rubelik AR, Livak KJ, Rafalski A, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res 18: 6531–6535