Euphytica **130:** 185–190, 2003.
© 2003 *Kluwer Academic Publishers. Printed in the Netherlands.*

Genomic evidence for a repetitive nature of the RAPD polymorphisms in *Olea europaea* (olive-tree)

Rocío Bautista, Francisco M. Cánovas & M. Gonzalo Claros*

*Departamento de Biología Molecular y Bioquímica, Facultad de Ciencias e Instituto Andaluz de Biotecnología, Unidad Asociada UMA-CSIC, Universidad de Málaga, Málaga, Spain; (*author for correspondence; e-mail: claros@uma.es)*

Received 21 May 2002; accepted 29 September 2002

Key words: olive tree, *Olea europaea*, RAPD, repetitive DNA, SCAR

Summary

The sequence of 12 SCAR markers obtained for olive-tree was analysed in the seek of information about the nature of the RAPD polymorphisms in olive-tree. The absence of internal repetitions of the RAPD primer corroborates that the polymorphisms were not obtained by alternative priming nor by rearrangements. We have found a high degree repetition within each sequence which is a reflect of the repetitive nature of the RAPD polymorphisms. Most the sequence corresponded to unknown DNA sequences, although some of the SCAR markers like GH1, PH2 or PF84 present clear homology or identity to other repetitive sequences (mainly microsatellites) from other plants or animals. This suggests that repetitive sequences could be more conserved among species than initially thought. As expected, none of the markers presented any open reading frame, although PJ65 revealed a significant homology to the gene and pseudogene of a lectin receptor in *Arabidopsis*. Taken together, the results described in this paper demonstrate that the polymorphisms detected by SCAR/RAPD methods are mainly repetitive in olive-tree, and that most of these repetitions corresponded to microsatellite DNA.

Abbreviations: AFLP – Amplification Fragment Length Polymorphism; ORF – Open Reading Frame; PCR – Polymerase Chain Reaction; RAPD – Random-Amplified Polymorphic DNA; SCAR – Sequence-Characterized Amplified Region; SSR – Simple Sequence Repeats

Introduction

The olive-tree (*Olea europaea* L.) is one of the most characteristic species of the Mediterranean area and nowadays it is the only cultivated representative of the genus *Olea*. Two varieties can be distinguished corresponding to cultivated *sativa*, which is propagated vegetatively, and the wild type *sylvestris* (syn. *Olea oleaster* Hoffm. et Link), which only reproduces by sexual means and is used for grafting some *sativa* cultivars. The cultivated olive-tree is a perennial evergreen bearing hermaphroditic flowers that show self-incompatibility. The self-incompatibility, mainly due to male-sterility, is determined by cytoplasmic inheritance and seems to be maintained by cultivar selection (Besnard et al., 2000). The origin and evolution of the cultivated variety is uncertain, although it is believed that it evolved in the Eastern Mediterranean region. In fact, carbon-dating has shown that there are eight thousand years old olive seeds, and Archeological evidence suggests that olive-tree was grown in Crete as long ago as 2,500 B.C. (Liphschitz et al., 1991). Phylogenetic analyses carried out on *Olea* species mainly obtained by RAPD strongly suggest an autochthonic origin of olive-tree cultivars when geographically related varieties are considered (Wiesman et al., 1998; Angiolillo et al., 1999; Claros et al., 2000; Sanz-Cortés et al., 2001), displaying a gradient between the east and west of the Mediterranean Basin. However, molecular markers using mitochondrial or chloroplastic sequences suggest that olive-tree cultivars were originated independently (Besnard & Berville, 2000). The combination of both

kind of results suggests that cultivar selection has occurred in different genetic pools and in different areas (Besnard et al., 2001). Until recently, breeding and improvement of olive-trees was limited due to their long juvenile phase and the lack of phylogenetic relationships among *Olea* species. Efforts for the development of new varieties by systematic breeding have produced very little results because they are only helped by empirical knowledge (Lavee et al., 1986).

Information about evolution and speciation can be obtained by analysing the molecular and genomic organization of DNA sequences within the genome. A very large proportion of most eukaryotic genomes is composed of repetitive DNA sequences, which can be organised either in tandem arrays of up to many thousands of base pairs, or in dispersed fashion originated by mobile genetic elements (Smyth, 1991). It is striking that although olive-tree is one of the oldest cultivated crops, very little is known about the molecular, genomic and chromosomal organisation of the *Olea* species: nothing is known about the structure of olive-tree chromosomes, and the fraction of the total DNA content that corresponds to highly-, mid-, and low-repetitive sequences remains also unknown. Moreover, no physical or genetic map have been performed on olive-trees, although the recent techniques like RAPD (Claros et al., 2000; Sanz-Cortés et al., 2001), AFLP (Angiolillo et al., 1999), SSR (Rallo et al., 2000) or SCAR (Bautista et al, accompanying paper) as well as cloned genes (Villalba et al., 1994; Haralampidis et al., 1998; Ledesma et al., 1998; Ledesma et al., 2000) provide enough information to design specific markers whose DNA sequence is (or can be) known since they can be amplified specifically by PCR.

Recently, two tandemly repeated DNA sequences with subtelomeric location have been isolated in *Olea europaea* ssp. *sativa*, one of 81 bp and the other of 218 bp, showing a 79% homology the one to the other (Katsiotis et al., 1998). These repetitive elements seems to be widespread among the *Olea* species. Two other repetitive sequences have been isolated from satellite DNA (T. Palomeque, unpublished results) that seem to be similar to repetitive DNA in other plants. By the other side, when developing SCARs markers to evaluate the olive germplasm, a sequence of 407 bp has been identified to show significant DNA homology with retrotransposon-like sequences (Hernández et al., 2001).

In our laboratory, we have prepared 12 SCAR markers (Bautista et al, see the accompanying paper) that have been used to identify 22 olive-tree cultivars in a reproducible and easy way. Now, we decided to study the sequence of the DNA fragments amplified by these SCAR markers in the seek of information about the nature of the RAPD polymorphisms in olive-tree.

Materials and methods

Sequencing RAPD fragments

The 12 SCAR markers based in *Olea europaea* genomic DNA developed in our laboratory (see the accompanying paper of Bautista et al.) 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. Double-strand sequencing of the cloned fragments was carried from two the ends using the T7, T3 or SP6 primers to obtain the consensus sequence.

Sequence analysis

The searches for sequence identities were carried out for nucleotide sequences using the BLAST tool (Basic Local Alignment Search Tool, Altschul et al., 1990) using the default settings. The searches for widespread sequence similarities were performed with the Pearson and Lipman algorithm (FASTA) since, for nucleotide searches, it provides more sensitive results than BLAST. In both cases, sequences were compared against the non-redundant GenBank-EMBL nucleotide sequence databases. The self comparison of each SCAR sequence was performed with the COMPARE program with a window of 11 residues and a stringency of 10 identities. The results were displayed as a dot matrix with the program DOTPLOT. Similarity searches as well as self-comparisons were carried out locally at the 'Centro Informático Científico de Andalucía (CICA)' as a part of the GCG Wisconsin Package® 10.1 (Devereux et al., 1984).

Results and discussion

The DNA bands used to obtain the SCAR markers (accompanying paper) were sequenced for at least 219 bp from each end, which enabled us to complete the sequence of the fragment in most cases. They were obtained from RAPD polymorphisms, which are thought

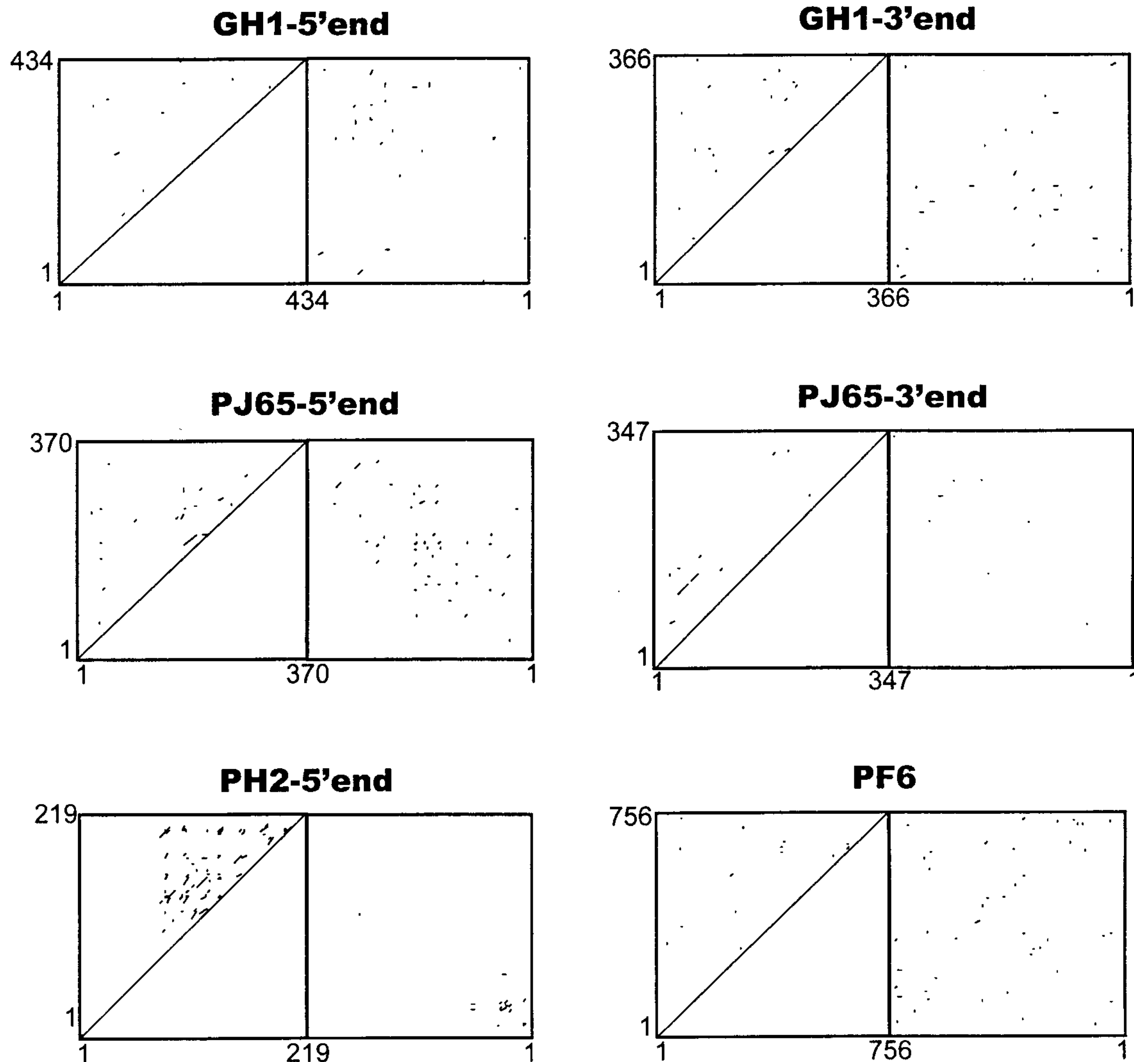

Figure 1. Comparison of the direct (left) and reverse (right) sequence of the more significant SCAR markers with itself. A window of 11 nucleotides allowing one error was used.

to be caused by differences in nucleotide sequence at the priming sites, or by rearrangements within the amplified sequence. The sequences we report do not contain inverted repeats equal to the sequence used as RAPD primer binding site which rules out the possibility that the cloned polymorphisms were obtained by DNA rearrangements or alternative priming. This means that the 12 markers can serve to obtain genomic information about olive-trees by direct analysis of their nucleotide sequence.

We have compared each sequence (or fragment of sequence) with itself and its complementary sequence to detect short repeats within a window of 11 nt allowing a mismatch of 1 nt. A RAPD amplification means that the same sequence is inversely repeated at no more than 2 kbp, which usually occurs in the repeated sequences. Accordingly, we have found a high degree of direct and inverted repeats within each sequence. All sequences contain several repeats longer than 10 bp which are a reflect of the repetitive nature of RAPD polymorphisms. The most representative results are shown in Figure 1 where it should be noted that the fragments are the longest ones described in Table 1. This could be because larger repeats

Table 1. Name of each SCAR locus, sequence of the oligonucleotide primers used to amplify it, and length of the fragment that should be amplified by PCR. When the complete sequence was not obtained, the size was calculated by band mobility in agarose gel electrophoresis, and the parentheses contain the size of the end-sequenced fragments from the 5′ and 3′ ends respectively

SCAR marker	5′ primer	3′ primer	Size (pb)
PF 83	CTCTGCCGATGGTCTAGGC	CTCTGCCTGAAACAGAATT	310
PF80	CTCTGCCTGAGTTCTAGTGA	CTCTGCCTGAACCAACAGAA	303
PF84	CTCTGCCTGAACATATTGCT	CTCTGCCTGAGAGATGCCTA	264
PF6	GGGAATTCCGCAGAACTAAA	GGGAATTCCGGCACATCCTT	756
GF10	GTGCCAATTTCTGACATTTT	ACCTGATAGGGAAGATTTAT	453
PJ11	CCCGGCATAAATTTCGGGCTC	CCCGGCATAAAACCAAATATT	447
PJ65	TCGTTCCGCAAGTTAAGACC	TCGTTCCGCATGTATTGTGG	905 (370+347)
GJ1 8	TGGTCGCAGACGGTACGCGT	TGGTCGCAGAACACGATTAC	225
GX4B	CACTTCAACAACTTTATCGA	TTCATCTTCTTCCCACAAAG	408
GH1	TCCACATCATAGCCAACCTT	CATGATTATATAGTCCAGGC	978 (366+434)
PH2	CACTTCCGCTCATTGAAGTA	CACTTCCGCTGAAACTTAAT	720 (219+486)

would be able to assume more effective conformations and thus, repetitions should become more abundant (Heringa, 1998). In addition, less structurally stable repeats are shortened during the processes of recombination, replication and repair, which could explain their low abundance. Except in the case of the 5′ end of the marker PF83, the last 10 bases exactly matched the primer sequences. It should be noted the case of marker PH2 that initiates with a sequence enriched in inverted repeats and follows with a sequence enriched in direct repeats (Figure 1). The frequency of direct and inverted repeats observed in the 5′ end of the marker PJ65 is also noteworthy. The GC content of the sequenced bands should not be taken into account for a general calculation of the overall GC content of olive DNA since the sequences obtained in this work are mainly repetitive and it is well known that the eukaryotic inverted repeats contained a very biased GC content (Cox & Mirkin, 1997).

BLAST algorithm was used to detect long sequence identities between markers described in this work and sequences in the databases (Table 2). Although most of the sequences did not show similarity to any sequence of known function, some striking identities were detected: GH1 shares sequences of up to 23 nt with human, *Drosophila*, *Caenorhabditis*, bacteria, etc., both within the 5′ and the 3′ end. We have to note that the GH1-3′end matches with L1 family of repeated sequences in *Mus musculus*. PH2 also presents identities with DNA from a wide range of organisms. Due to its high CA content, most of these sequences belong to microsatellite sequences. The most surprising is the 5′ end of PJ65 that shares 54 nt with the gene and the pseudogene of an *Arabidopsis* lectin receptor. In the search of generalised homology, the algorithm FASTA was used, resulting that PF84 is 75.4% identical to a 57 nt length sequence of tomato (accession number AJ223822) amplified by RAPD too. PF84 is also 59.4% identical to 217 nt of a microsatellite sequence contained in the pXP106 clone of *Gossypium barbadense* (accession number AF060614) which has been proved to be widely distributed among genomes. Most of RAPD markers are grouped in clusters located around centromeric, telomeric or heterochromatic regions (Zhao et al., 1998; Saliba-Colombani et al., 2000), making them excellent landmarks for studying chromosome structure, function and evolution. These loci are enriched in repetitive DNA which are hypervariable in plants since the polymorphism might originate from unequal crossing-over of from slipped-strand mispairing during replication (Morgante & Olivieri, 1993). This can explain why the PJ65 marker presents the same 5′ and 3′ end-sequence than the RAPD band OPJ6-1278 (Bautista et al., accompanying paper) while the fragment length is different. This can account for the fact that PH2 and PF84 revealed homology with microsatellite sequences, and could reflect that these repetitive sequences are more conserved among species than initially thought (Jurka, 1998). This is supported by the fact that GH1-3′end is significantly similar to the L1 family of repeated sequences in *Mus musculus*.

No one of the 12 markers contained an ORF (result not shown). Taking into account that (i) much more

Table 2. Database sequences identical to fragments of the SCAR markers described in this work. Nucleotides in lowercase are those not conserved with respect to the SCAR sequence

Marker (position)	Homologous sequence origin	Alignment
GH1-5′end		
(300 to 326)		CATTTCATTATTATCATTATTTTGAGA
	Lupinus angustifolius	CATTTCATTATTATCATTATTTTG
	Human DNA 1	TTTCATTATTATCATTATTTTGAGA
	Human DNA 2	ATTATTATCATTATTTTGAGA
	Mus musculus	TCATTATTATCATTATTTTG
	Plasmodium falciparum	TTTCATTATTATCATTATTTT
GH1-3′end		
(90 to 108)		TGGATAATAAGCTGTGAAT
	Human DNA	TGGATAATAAGCTGTGAAT
	Human factor RIP140	TGGATAATAAGCTGTGAAT
GH1-3′end		
(235 to 259)		TCATTTTTCTTTAATTTTTAATTAA
	Caenorhabditis elegans	TTTTTCTTTAATTTTTAATTAA
	Drosophila melanogaster	TCATTTTTCTTTAATTTTTAATTAA
	Methanococcus jannaschii	TTTTTCTTTAATTTTTAATT
	Meloidogyne javanica	TCATTTTTCTTTAATTTTTAATT
PF80		
(247 to 273)		TTTGTTCTTTTGGCTTATGGTTTACTT
	Human DNA	TTTGTTCTTTTGGCTTATGGTTTACTT
PH2-5′ end		
(4 to 27)		TTCCGCTCATTGAAGTAGCTGTGC
	Fasciola hepatica	TTCCGCTCATTGAAGTAGCTGTGC
	Human DNA	TCATTGAAGTAGCTGTGC
PH2-3′end		
(190 to 214)		ATTTTATATATATATGCACACACAC
	Human DNA 1	ATTTTATATATATATGCACACACAC
	Human DNA 2	TTATATATATATGCACACACAC
	Felis catus	TATATATATATGCACACACA
	Rattus norvegicus	TATATATATATGCACACACAC
	Mus musculus	TATATATATATGCACACACAC
PJ65-5′end		
(240 to 265)		AAGAATATCTAGAGAATTCTAGGAAA
	Caenorhabditis elegans	AAGAATATCaAGAGAATTCTAGGAAA
PJ65-3′		
end(25 to 79)		CTTGAGTAGCAAAATCAAGTTTAAGGGGGGATGTTGAGTAGTAAACTTGATTTT
	Arabidopsis thaliana	CTTGAGTAGCAAAATCAAGTTTAAGGGGGGATGTTGAGTAGTAAACTTGATTTT

than half of most eukaryotic genomes is repetitive DNA, and that (ii) RAPD are more easily produced in repetitive DNA, this absence of ORF should not be unexpected. This is consistent with the fact that, in eukaryotes, both inverted and mirror repeats are virtually excluded from coding sequences (Morgante & Olivieri, 1993; Cox & Mirkin, 1997). Mirror repeats have often been associated with inversion endpoints, becoming recombination hot spots and significantly enhance inversed recombination events, although they seem to be somatically stable as no somatic mosaicism was observed (Morgante & Olivieri, 1993). In other cases, direct repeats could be the prompt for a transposition event that might help in evolutionary adaptations.

It has been reported that repetitive DNA comprises a large and rapidly evolving portion of the genomes (Zhao et al., 1998), and that repeats allow to classify the genotypes by the retention of hybridization signals after high-stringency washes. This may be equivalent to presence/absence of a PCR band, or different degrees of intensity provided that the primers are able to recognise the divergent sequence. The importance of genomic repetitions is enhanced as their formation could be the cause of evolutionary genome variability. Markers GJ18, GH1, PH2 and PJ65, produce a small amplification of 50 nt (Bautista et al., accompanying paper) that clearly illustrates the repetitive nature of DNA amplified by RAPD. Perhaps this can explain why RAPD is the easiest and quickest technology to differentiate organisms and/or individuals.

In conclusion, this study has confirmed, in olive-trees, the molecular basis of the polymorphisms detected as RAPD/SCAR methods. The 12 SCARs revealed the repetitive nature of RAPD amplifications in olive-tree.

Acknowledgements

We are indebted to R. Crespillo for her technical assistance. We are also grateful to the Research Services of Málaga University for providing facilities at the Molecular Biology Laboratory.

References

Altschul, S.F., W. Gish, W. Miller, E.W. Myers & D.J. Lipman, 1990. Basic local alignment search tool. J Mol Biol 215: 403–410.

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

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

Besnard, G. & A. Berville, 2000. Multiple origins for mediterranean olive (*Olea europaea* L. ssp. *europaea*) based upon mitochondrial DNA polymorphisms. C R Acad Sci Ser III-Vie 323: 173–181.

Besnard, G., B. Khadari, P. Villemur & A. Berville, 2000. Cytoplasmic male sterility in the olive (*Olea europaea* L.). Theor Appl Genet 100: 1018–1024.

Claros, M.G., R. Crespillo, M.L. Aguilar & F.M. Cánovas, 2000. DNA fingerprinting and classification of geographically related genotypes of olive-tree (*Olea europaea* L.). Euphytica 116: 131–142.

Cox, R. & S.M. Mirkin, 1997. Characteristic enrichment of DNA repeats in different genomes. Proc Natl Acad Sci USA 94: 5237–5242.

Devereux, J., P. Haeberli & O. Smithies, 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res 12: 387–395.

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.

Heringa, J., 1998. Detection of internal repeats: how common are they? Curr Opin Struct Biol 8: 338–345.

Hernández, P., R. de la Rosa, L. Rallo, A. Martin & G. Dorado, 2001. First evidence of a retrotransposon-like element in olive (*Olea europaea*): Implications in plant variety identification by SCAR marker development. Theor Appl Genet 102: 1082–1087.

Jurka, J., 1998. Repeats in genomic DNA: mining and meaning. Curr Opin Struct Biol 8: 333–337.

Katsiotis, 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., A. Haskal & M. Wodner, 1986. 'Barnea': A new olive cultivar from first-breeding generation. Olea 17: 95–99.

Ledesma, A., M. Villalba, E. Batanero & R. Rodríguez, 1998. Molecular cloning and expression of active Ole e 3, a major allergen from olive-tree pollen and member of a novel family of Ca^{2+}-binding proteins (polcalcins) involved in allergy. Eur J Biochem 258: 454–459.

Ledesma, A., M. Villalba & R. Rodríguez, 2000. Cloning, expression and characterization of a novel four EF-hand Ca^{2+}-binding protein fromo olive pollen with allergenic activity. FEBS Letters 466: 192–196.

Liphschitz, N., R. Gophna, M. Hartman & G. Biger, 1991. The beginning of the olive (*Olea europaea*) cultivation in the old world: a reassessment. J Archeol Sci 18: 441–453.

Morgante, M. & A.M. Olivieri, 1993. PCR-amplified microsatellites as markers in plant genetics. Plant J 31: 175–182.

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

Saliba-Colombani, V., M. Causse, L. Gervais & J. Philouze, 2000. Efficiency of RFLP, RAPD and AFLP markers for the construction of an intraspecific map of the tomato genome. Genome 43: 29–40.

Sanz-Cortés, F., M.L. Badenes, S. Paz, A. Iñiguez & G. Llacer, 2001. Molecular characterization of olive cultivars using RAPD markers. J Amer Soc Hort Sci 126: 7–12.

Smyth, D.R., 1991. Dispersed repeats in plant genomes. Chromosoma 100: 355–359.

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

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.

Zhao, X., Y. Ji, X. Ding, D.M. Stelly & A.H. Paterson, 1998. Macromolecular organization and genetic mapping of a rapidly evolving chromosome-specific tandem repeat family (B77) in cotton (*Gossypium*). Plant Mol Biol 38: 1031–1042.