

# Development of EST-derived SSR Markers with Long-core Repeat in Olive and Their Use for Paternity Testing

Raúl De la Rosa<sup>1</sup> and Angelina Belaj

*Instituto de Investigación y Formación Agraria y Pesquera (IFAPA)–Centro “Alameda del Obispo,”  
Avda. Menéndez Pidal s/n, 14004 Córdoba, Spain*

Antonio Muñoz-Mérida and Oswaldo Trelles

*Dpto. Arquitectura de Computadores, ETSI Informática, Universidad de Málaga, Campus de  
Teatinos, Málaga E-29071, Spain*

Inmaculada Ortiz-Martín, Juan José González-Plaza, Victoriano Valpuesta,  
and Carmen R. Beuzón

*Instituto de Hortofruticultura Subtropical y Mediterránea, Universidad de Málaga-Consejo Superior  
de Investigaciones Científicas (IHSM-UMA-CSIC), Campus de Teatinos, Málaga E-29071, Spain*

**ADDITIONAL INDEX WORDS.** *Olea europaea*, microsatellites, olive breeding, crosses, pollination, cultivar

**ABSTRACT.** In the present work, a set of eight new hexa-nucleotide simple sequence repeats (SSRs) is reported in olive (*Olea europaea* L.). These SSRs loci were generated on the basis of expressed sequence tag (EST) sequences in the frame of an olive genomic project. The markers showed a high level of polymorphism when tested on a set of cultivars used as genitors in the olive breeding program of Córdoba, Spain. The long-core repeat motif of these markers allows a wider separation among alleles, thus permitting an accurate genotyping. Besides, these markers showed comparable levels of polymorphism to di-nucleotide SSRs, the only ones so far reported in olive. Selected on the basis of their discrimination capacity, four of the eight SSRs were used to test their ability for paternity testing in a total of 81 seedlings coming from 12 crosses. The paternity testing showed that seven crosses matched the alleged paternity and the remaining five were products of illicit pollinations. These results exactly matched with previous paternity testing performed with di-nucleotide SSR markers. These results demonstrate the usefulness of the developed hexa-nucleotide repeated motifs for checking the paternity of breeding progenies and suggest their use on variability studies.

Compatibility among genitors has been reported to be one of the main factors for success of crosses in olive breeding programs (Díaz et al., 2007). Olive compatibility has been studied through different systems including fruit set in controlled pollinations and pollen tube germination (Bradley and Griggs, 1963; Breton and Bervillé, 2012). Current information indicates that the compatibility system in olive might be of the sporophytic type (Breton and Bervillé, 2012; Collani et al., 2012), although further investigations should be performed to identify the genetic variability associated with this trait.

Until now, di-nucleotide SSR markers have been the only tool successfully used in paternity testing in controlled olive breeding crosses (De la Rosa et al., 2004; Díaz et al., 2007). They have also been used in pollination studies, identifying cases of full self-incompatibility (Díaz et al., 2006a; Mookerjee et al., 2005) and cross-incompatibility (Díaz et al., 2007; Seifi et al., 2012). Additionally, SSRs have been used to demonstrate that

well-known cultivars such as Picual and Arbequina are self-incompatible (Díaz et al., 2006b). As a consequence, paternity testing with SSRs is considered a very useful tool in olive breeding programs to demonstrate compatibility relationships.

Currently, only  $\approx 100$  SSR markers are available in olive, and all of them were obtained by sequencing enriched genomic libraries and are based on di-nucleotide repeats (Carriero et al., 2002; Cipriani et al., 2002; De la Rosa et al., 2002; Díaz et al., 2006a; Rallo et al., 2000; Sabino-Gil et al., 2006; Sefc et al., 2000). Doubts about the reproducibility of these types of markers among different laboratories as well as among experiments within the same laboratory have been reported as a result of the small bp differences among alleles (Baldoni et al., 2009; Faria et al., 2011; Vélez and Ibáñez, 2012). Additionally, they could be far from coding sequences (Poncet et al., 2006), which may limit their use in marker-assisted selection.

Expressed sequence tags have become a popular source for SSR markers in recent years (Kumpatla and Mukhopadhyay, 2005). EST-derived SSR markers have several advantages compared with SSR markers originated from genomic libraries, including simpler marker development and higher levels of transferability among species (Rahemi et al., 2012; Wöhrmann and Weising, 2011). Additionally, EST-SSRs are of particular interest for marker-assisted selection because they are located inside expressed genes (Cova et al., 2012). In general, they represent a source of variability complementary to genomic SSRs (Huang et al., 2011). In several tree species, EST sequences have

Received for publication 17 Dec. 2012. Accepted for publication 23 Apr. 2013. The present work was partly supported by OLEAGEN (Fundación Genoma España, Junta de Andalucía through Instituto de Investigación y Formación Agraria y Pesquera and Corporación Tecnológica de Andalucía) and P09-AGR-4516 (Junta de Andalucía, Spain) projects. Angelina Belaj has received a postdoctoral fellowship Subprograma DOC-INIA (National Institute of Agricultural Research contract), Ministry of Education and Culture, Spain. We are thankful to Dr. Daniel James Sargent for his valuable comments and review of the manuscript.

<sup>1</sup>Corresponding author. E-mail: raul.rosa@juntadeandalucia.es.

been used to develop SSRs from di- to hexa-nucleotide repeats (Faria et al., 2011; Huang et al., 2011; Poncet et al., 2006). Long-core repeat SSRs have successfully been used in identification and paternity studies in tree species (Cipriani et al., 2008; Faria et al., 2011), because they display higher allelic separation than di-nucleotide SSRs and, therefore, more precise allele calling, although they have also been shown to be less polymorphic. Long-core SSRs could be of interest in olive genetics research as a result of the high heterozygosity of this crop (Baldoni and Belaj, 2009).

In a recent Olive Genomic Project in Spain (OLEAGEN), aimed at developing a genomic tool to assist olive breeding, EST sequences coming from different olive tissues have been identified using Sanger technology. The purpose of the present work is to develop EST-derived hexa-nucleotide SSR markers and to evaluate their polymorphism and use for paternity testing.

### Materials and Methods

**EST-SSR MINING AND PRIMER DESIGN.** In the framework of the OLEAGEN project (Muñoz-Mérida et al., 2013), data from cDNA libraries sequenced using Sanger technology were obtained. The first library was generated from cultivar Lechín de Sevilla fruit mesocarp at three maturation stages (green with lignified endocarp, turning, and purple). The second came from buds from young and adult branches of 10 seedlings of the cross of cultivars Picual and Arbequina. The last was generated from young leaves and stems of ‘Lechín de Sevilla’ plus seeds from fruit at two different maturation stages (turning and purple) from a progeny of ‘Picual’ and ‘Arbequina’. A cleaning step was performed using Seqclean (2005–06; Dana-Farber Cancer Institute, Boston, MA) to discard contaminant sequences from the vectors and primers used in generating the libraries before SSR search.

Identification of SSRs in the three libraries was carried out using MISA software [Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany (Thiel et al., 2003)] (Fig. 1). The output from MISA included the repeated motif, the representative read that contains the motif and both forward and reverse primers for sequence amplification. Sometimes, primers could not be provided by MISA as a result of the SSR being too close to the end of the read.

To assess the level of polymorphism of the detected SSRs, primer sequences were used as queries against all reads available in the OLEAGEN project (Muñoz-Mérida et al., 2013), generated both by Sanger and 454 pyrosequencing technologies, using BLASTN [National Library of Medicine, Bethesda, MD (Altschul et al., 1990)]. These included nine libraries generated from ‘Picual’ fruit mesocarp, seeds, and buds; ‘Arbequina’ fruit mesocarp and buds; and roots from a mix of cultivars. Hits for each primer were compared and only those results for which a hit could be obtained for both primers were selected. For those SSRs for which MISA could not provide primers, the repeat motif was directly used as query to perform a BLASTN search against the entire sequence database. This process provided information about the number of motif repeats found in each library, thus providing useful information about their level of polymorphism within the studied olive cultivars. Among all the sequences containing hexa-nucleotide repeats, those showing polymorphism among the cultivars used for library construction (Tables 1 and 2) were selected for genotyping. As mentioned before, for some SSR loci, primer design could not be provided

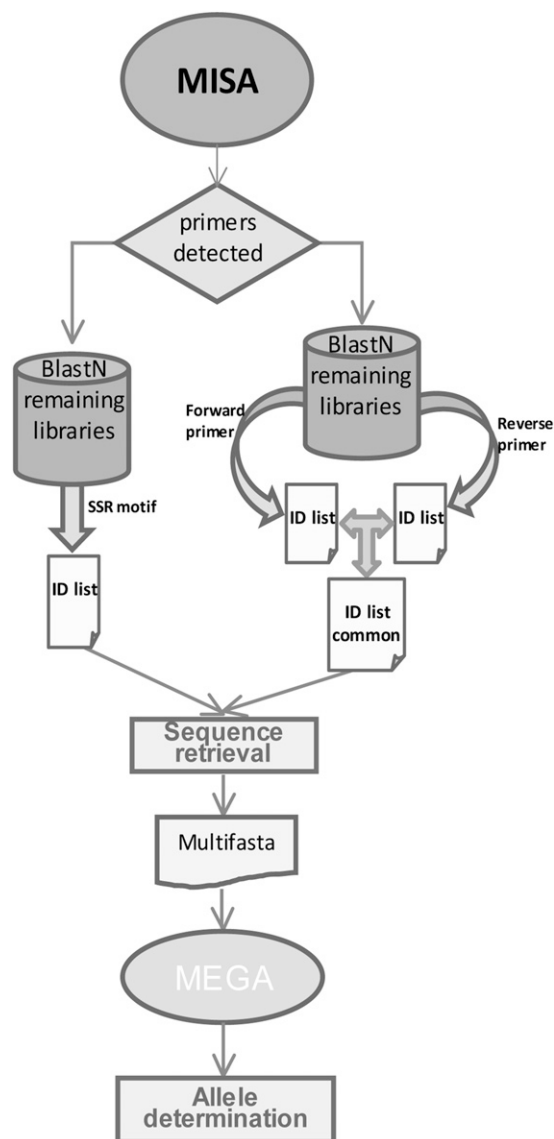


Fig. 1. Workflow of the express sequence tag–simple sequence repeat (SSR) mining process on olive cDNA genomic libraries. Sequences containing SSR are identified using MISA software [Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany (Thiel et al., 2003)]. Polymorphism on the SSR sequences is detected using BLASTN [National Library of Medicine, Bethesda, MD (Altschul et al., 1990)].

by MISA software. In those cases, the Primer3 program (Rozen and Skaletsky, 2000) was used to design amplification primers. To have at least one SSR per each repeated motif, OLEAGEN-H21 (Table 1) was also selected for primer design, although no polymorphism was initially detected.

**PLANT MATERIAL AND DNA EXTRACTION.** Thirteen olive cultivars from the World Olive Germplasm Bank of Cordoba, Instituto de Investigación y Formación Agraria y Pesquera, Junta de Andalucía, Spain (Table 3), the new olive breeding cultivar Sikitita, also named as Chiquitita in the United States (Rallo et al., 2008), and one wild olive genotype (Belaj et al., 2010) were used to test the amplification quality and polymorphism of the eight primer pairs designed. These cultivars were used as genitors of 12 breeding progenies obtained in the olive breeding program of Córdoba, Spain (Table 4). The progenies

Table 1. Hexa-nucleotide simple sequence repeat (SSR) loci found in the olive express sequence tag mining.<sup>z</sup>

SSR	Motif	Motif repeats (no.)			
		Picual	Arbequina	Lechín de Sevilla	Others
OLEAGEN-H2 <sup>y</sup>	ACCAGC	7	6	4	7
OLEAGEN-H2.1 <sup>y</sup>	ACCAGC			4, 3	
OLEAGEN-H3 <sup>y</sup>	ATGAGG	7, 6	6, 4		7, 6, 5
OLEAGEN-H3.1	ATGAGG				7
OLEAGEN-H3.2	ATGAGG		5		
OLEAGEN-H3.4	ATGAGG		5		
OLEAGEN-H4 <sup>y</sup>	GCTCCG	7			3, 2
OLEAGEN-H4.1 <sup>y</sup>	GCTCCG	5, 4			5
OLEAGEN-H4.2 <sup>y</sup>	GCTCCG		4	3	
OLEAGEN-H5	CGCAA			6	
OLEAGEN-H6	CTCTTC		6		
OLEAGEN-H6.1 <sup>y</sup>	CTCTTC			5	2
OLEAGEN-H6.2 <sup>y</sup>	CTCTTC	3			5, 4
OLEAGEN-H6.3	CTCTTC				4
OLEAGEN-H6.4	CTCTTC		5		
OLEAGEN-H6.5	CTCTTC		5		
OLEAGEN-H6.6	CTCTTC	5			
OLEAGEN-H6.7	CTCTTC	5		5	5
OLEAGEN-H6.8	CTCTTC	5	5		
OLEAGEN-H7	TTGTAT			8	
OLEAGEN-H8 <sup>y</sup>	ATTCC	3	5	5	
OLEAGEN-H8.1 <sup>y</sup>	ATTCC	4	2	2	2
OLEAGEN-H9 <sup>y</sup>	GAAGAG	5	4	5	5
OLEAGEN-H9.1	GAAGAG		6		
OLEAGEN-H9.2	GAAGAG	5	5		
OLEAGEN-H9.3	GAAGAG			5	
OLEAGEN-H9.4 <sup>y</sup>	GAAGAG				5, 4
OLEAGEN-H9.5	GAAGAG		5		
OLEAGEN-H9.6	GAAGAG		5		
OLEAGEN-H9.7	GAAGAG	5			
OLEAGEN-H9.8	GAAGAG			4	
OLEAGEN-H9.9	GAAGAG	4			
OLEAGEN-H9.10	GAAGAG	4			4
OLEAGEN-H14 <sup>y</sup>	CGTGAT	6, 5	6		9, 5
OLEAGEN-H17 <sup>y</sup>	GAGGAA	6, 5	6		6
OLEAGEN-H17.1	GAGGAA			5	
OLEAGEN-H17.2	GAGGAA		5	5	
OLEAGEN-H17.3	GAGGAA		5		
OLEAGEN-H17.4	GAGGAA	5			
OLEAGEN-H17.5	GAGGAA			4	
OLEAGEN-H17.6	GAGGAA	4			4
OLEAGEN-H17.7	GAGGAA	4		4	4
OLEAGEN-H17.8	GAGGAA	4	4		
OLEAGEN-H17.9	GAGGAA	4			
OLEAGEN-H20	GGTGAA		5		5
OLEAGEN-H20.1 <sup>y</sup>	GGTGAA	3		5	4
OLEAGEN-H20.2	GGTGAA	5			
OLEAGEN-H21 <sup>x</sup>	CCTCCA			6	
OLEAGEN-H21.1	CCTCCA				6
OLEAGEN-H21.2	CCTCCA				6
OLEAGEN-H21.3	CCTCCA	5			
OLEAGEN-H21.4	CCTCCA	5			

<sup>z</sup>The number of motif repeats in the three cultivars and seedlings of 'Picual' × 'Arbequina' cross are indicated.

<sup>y</sup>Polymorphic loci; i.e., those where two numbers of motif repeats have been found (within cultivars, between cultivars, or both).

<sup>x</sup>SSR 21 was also selected to have at least one representative of each repeated motif.

were derived from controlled crosses following the routine protocol of the breeding program (De la Rosa et al., 2004); i.e., male pollen is applied to flowers of the female parent located in bag-covered branches. The mentioned progenies were included to test the ability of the new SSRs for paternity testing.

Total DNA was extracted from fresh leaves following the protocol described by De la Rosa et al. (2002).

**SSR GENOTYPING.** Polymerase chain reactions (PCRs) were carried out in 12.5 µL volume containing 25 ng of DNA (extracted as described previously), 10 mM Tris-HCl pH 8.0,

Table 2. Variability parameters of the olive simple sequence repeat loci when tested in the set of 15 genitors of the olive breeding program of Córdoba, Spain.<sup>z</sup>

Primer	Motif	Forward (5' → 3')	Reverse (5' → 3')	Alleles				
				(no.)	Ho	He	Dj	PIC
OLEAGEN-H2	ACCAGC	TCAAAATGATGCAACAACCC	TGCAGGAGCAGAAACATTTG	7	1.000	0.809	0.914	0.755
OLEAGEN-H3	ATGAGG	TGGCCATAAGATTGATGATGA	CGATTACCTCGGCCATTCTA	7	0.867	0.846	0.952	0.793
OLEAGEN-H4	GCTCCG	ACACGGAGGATCCAAGTCTG	TGAGCAGGTGGTAGAGGCTT	5	0.800	0.740	0.914	0.674
OLEAGEN-H6	CTCTTC	CGGTATCATTTGCAGAGCAC	AGGGGAGTGGGTTGTTGTTA	7	0.867	0.846	0.952	0.793
OLEAGEN-H8	ATTTC	AGGTTTCATTAAGCCGCAAGA	TGTCCACACTGAACAGCCAT	3	0.385	0.551	0.819	0.424
OLEAGEN-H9	GAAGAG	TATGTCATCTTTGGGGAGGC	TTACTCCTTGACACCCCTGC	2	0.500	0.495	0.695	0.363
OLEAGEN-H20	GGTGAA	TTCAACCAGTCTCCCAGTC	GTAAGCGAGGGAGAGGCTTT	5	0.667	0.729	0.952	0.655
OLEAGEN-H21	CCTCCA	AAGAGTTGTTTCTGCCGCTC	CGCCAAAGCTACACATGAGA	6	0.933	0.784	0.867	0.726

<sup>z</sup>Number of alleles (Na), observed (Ho) and expected (He) heterozygosity, discriminating power (Dj), and polymorphism information content (PIC) are indicated.

Table 3. Allele pairs at the eight olive simple sequence repeat loci tested from the 15 genitors of the crosses considered in the paternity testing.

Cultivar	Primer							
	H2	H3	H4.1	H6.2	H8	H9	H20.1	H21
Arbequina	204/216	197/224	137/154	141/168	270/null <sup>z</sup>	240/246	234/246	289/295
Arbosana	204/216	193/224	143/154	138/168	MD <sup>y</sup>	240/240	240/246	278/295
Chalkidiki	204/227	197/242	143/160	141/187	270/270	240/240	246/246	278/301
Changlot Real	207/210	193/197	137/143	138/141	270/270	240/246	240/240	261/278
Dolce Agogia	198/204	193/197	143/160	138/141	270/270	MD	290/290	278/283
Empeltre	198/204	224/236	143/148	168/181	258/258	240/240	234/240	278/289
Frantoio	204/210	197/212	143/143	141/155	258/241	240/240	246/246	278/295
Koroneiki	204/216	197/218	137/137	141/162	270/270	240/246	240/246	278/289
Lechín de Sevilla	198/204	193/218	143/148	138/162	258/258	240/240	240/246	278/301
Manzanilla de Sevilla	204/207	236/242	137/148	181/187	258/270	240/246	240/252	278/283
Meski	204/207	193/197	137/160	138/141	258/270	240/246	240/246	278/295
Ocal	204/207	193/193	143/143	138/138	258/270	246/246	240/246	278/301
Picual	207/227	236/242	143/148	181/187	258/270	246/246	240/252	278/301
Sikitita	207/216	224/236	137/143	168/181	258/null	240/246	234/252	289/301
Wild olive (Je Ac 9)	210/221	224/224	143/148	168/168	270/270	240/246	246/246	283/283

<sup>z</sup>Null allele of 'Arbequina' was identified in H8 by the allele configuration of its descendant 'Sikitita'.

<sup>y</sup>MD = Missing data.

Table 4. Ability the new hexa-nucleotide primers here reported for the detection of olive seedlings with wrong paternity in comparison with the di-nucleotide primers routinely used in the breeding program of Córdoba, Spain.<sup>z</sup>

Progeny	Total seedlings (no.)	Seedlings with wrong paternity (no.)		Avg alleles detecting wrong paternity (no.)	
		Hexa-nucleotide	Di-nucleotide	Hexa-nucleotide	Di-nucleotide
Arbequina × Picual	5	0	0	0.0	0.0
Frantoio × Dolce Agogia	7	0	0	0.0	0.0
Koroneiki × Ocal	6	0	0	0.0	0.0
Lechín de Sevilla × Arbosana	5	0	0	0.0	0.0
Lechín de Sevilla × Picual	5	0	0	0.0	0.0
Picual × Wild (Je Ac 9)	4	0	0	0.0	0.0
Picual × Meski	10	0	0	0.0	0.0
Empeltre × Changlot Real	10	10	10	1.8	3.4
Frantoio × Changlot Real	9	9	9	2.3	3.4
Koroneiki × Sikitita	9	9	9	2.2	2.9
Manzanilla de Sevilla × Picual	5	5	5	2.0	3.0
Picual × Chalkidiki	5	4	5	2.8	3.0

<sup>z</sup>Four loci are used in both cases. Data for 12 progenies of the mentioned breeding program are presented.

1.5 mM MgCl<sub>2</sub>, 50 mM KCl, 0.2 mM of each dNTP, 0.2 mM forward (fluorescent labeled with FAM or HEX), and reverse primer and 0.0252 units of Taq polymerase (Gibco-BRL; Life Technologies, Carlsbad, CA). The PCR conditions comprised an initial denaturation step at 95 °C for 5 min followed by 35 cycles of 95 °C for 20 s, 55 °C for 30 s, and 72 °C for 30 s plus a final extension at 72 °C for 6 min in a thermal cycler (PTC-200; MJ Research, St. Bruno, Quebec, Canada). Electrophoresis was performed in an automated sequencer (PerkinElmer Applied Biosystems 3100; Life Technologies). The analysis of the PCR products was carried out using Genescan 3.7 and Genotyper 3.7 software (Applied Biosystems; Life Technologies). However, the peaks were also checked visually to detect possible errors in the assignation of alleles.

For the eight microsatellite primer pairs giving reproducible and readable peaks, number of alleles, observed ( $H_o$ ) and expected ( $H_e$ ) heterozygosity, polymorphism information content (PIC), and discrimination power ( $D_j$ ) were calculated (Tessier et al., 1999).

Four microsatellite loci were used for paternity testing of seedlings of the 12 breeding crosses. They were selected on the basis of their variability parameters and their polymorphism between the parents of the crosses tested. Paternity testing was performed as reported in previous works (De la Rosa et al., 2004); i.e., a seedling was considered to have a correct paternity if one maternal and one paternal allele were found for all the four loci considered. Results were compared with paternity testing of the same crosses with four di-nucleotide SSRs: DCA-9, DCA-16, DCA-18 (Sefc et al., 2000), and UDO-43 (Cipriani et al., 2002), performed as a routine in the olive breeding program (De la Rosa et al., 2004). Additionally, the different alleles of each hexa-nucleotide locus were sorted by their numbers of bps. The differences in bps among contiguous alleles have been recorded. The values obtained for the set of the four hexa-nucleotide SSRs were compared with the same measurements made for the four di-nucleotide SSRs used.

## Results and Discussion

EST mining in three EST libraries resulted in a total of 52 hexa-nucleotide SSR loci showing 12 different repeat motifs (Table 1). The comparison of the sequences of these 52 loci with all reads available in the OLEAGEN project (Muñoz-Mérida et al., 2013) revealed that 15 were polymorphic. Primers could be developed, either by MISA or by Primer3, for nine of them. Primers were also designed for OLEAGEN-H21 (Table 1) to have at least one SSR per each repeated motif detected.

Reproducible and polymorphic amplification products were obtained from eight hexa-nucleotide SSR loci (including H21), displaying from two to seven different alleles per locus (Tables 2 and 3). The relatively low number of alleles and PIC values found could be because EST-SSRs may display less polymorphism than genomic-based SSRs (Nishio et al., 2011; Poncet et al., 2006; Rahemi et al., 2012). Our results are also in agreement with previous studies in other species (Cipriani et al., 2008; Durand et al., 2010; Faria et al., 2011) that reported lower levels of polymorphism and variability in long-core repeat SSRs in comparison with di-nucleotide repeats. However, heterozygosity ( $H_o = 0.38-1.00$ ;  $H_e = 0.49-0.85$ ) and  $D_j$  values obtained were almost comparable to previous reports on olive di-nucleotide SSRs (De la Rosa et al., 2002; Díaz et al., 2006a; Noormohammadi et al., 2007; Sabino-Gil et al., 2006; Sefc et al., 2000).

Accordingly, in future studies with a higher number of cultivars and wild genotypes, more precise estimations of number of alleles and diversity parameters could be expected.

It is interesting to note that the SSR locus OLEAGEN-H21, found monomorphic in the BLASTN search, gave a high level of variability on the cultivars tested (Table 2). This indicates that the rest of the EST-SSR loci identified as monomorphic here (Table 1) could be polymorphic when evaluated in genomic DNA of a range of genotypes. Therefore, those loci could provide additional SSR markers that might be of interest for future identification, mapping, and variability studies.

Four SSR loci (OLEAGEN-H2, OLEAGEN-H3, OLEAGEN-H6.2, OLEAGEN-H20.1) were used for paternity testing of 12 crosses of the breeding program of Córdoba. They were selected on the basis of the variability parameters and the polymorphism among the respective parents. Two pairs of the four loci selected (OLEAGEN-H2 + OLEAGEN-H3 and OLEAGEN-H2 + OLEAGEN-H6.2) were sufficient to discriminate among all the 15 genotypes used as parents of the crosses (Table 3). The four SSR loci selected were sufficient to determine the paternity of the seedlings of the 12 crosses analyzed. The results of the paternity test with long-core repeat SSRs fully agree with previous determinations of the same crosses performed with di-nucleotide primers; i.e., seven crosses were considered true and in the remaining five, seedling were found to have the wrong paternity (Table 4). Only one seedling coming from 'Picual' × 'Chalkidiki', which was identified as contaminated using di-nucleotide SSRs, was not identified with the hexa-nucleotide SSRs. The higher polymorphism of the di-nucleotide primers is likely the cause for the higher frequency of alleles detecting wrong paternity (Table 4). In any case, as mentioned, using the hexa-nucleotide SSRs, the crosses analyzed can be categorized as successful or unsuccessful in the same way by both hexa-nucleotide and di-nucleotide loci. In previous reports, olive breeding crosses have been identified as successful or failed without intermediate categories (De la Rosa et al., 2004; Díaz et al., 2007). Therefore, the set of four hexa-nucleotide SSRs used seems to be adequate to control the success of olive breeding crosses as previously reported for long-core repeat EST-SSRs in *Eucalyptus* L'Hér. (Faria et al., 2011) and *Vitis* L. (Cipriani et al., 2008).

The main advantage of the hexa-nucleotide SSRs is that they showed a higher average distance among alleles (6.5 vs. 4.0 bp) than the di-nucleotide, the latter ones having 1 or 2 bp differences in 77% of the cases (Table 5). Therefore, the use of hexa-nucleotide SSRs would increase the precision in allele calling (Faria et al., 2011) because the minimum bp differences among alleles would increase reducing, thus the chance of genotype misscoring. It is also striking that few di-nucleotide SSR alleles showed 1-bp differences and some hexa-nucleotide alleles showed 3- to 5-bp differences. This could be caused by complex mutation patterns as reported for *Coffea* L. (Poncet et al., 2006) or to genotyping errors.

In conclusion, the new EST sequences developed for olive (Muñoz-Mérida et al., 2013) have allowed the identification of the first set of hexa-nucleotide microsatellites in this crop. EST-mining has been shown to be a convenient strategy to search for long-core repeat SSR markers as previously reported for other species (Kumpatla and Mukhopadhyay, 2005). The new markers developed have shown a high degree of polymorphism and usefulness in paternity testing. However, as a result of the conservative nature of EST-SSR, transferability to other species of

Table 5. Comparison of the frequency of the occurrence of different distances (bps) among contiguous alleles in the hexa-nucleotide primers here reported versus di-nucleotide simple sequence repeat previously used.<sup>z</sup>

Distance between contiguous alleles (bp)	Frequency (no.)	
	Di-nucleotide	Hexa-nucleotide
1	2	
2	22	
3		3
4		1
5		2
6	2	11
7		2
10	2	
12	1	1
13		1
14		1
16	1	
17		1
18	1	

<sup>z</sup>The frequency is calculated on the basis of both the genitors and progenies tested.

*Olea* could be tested in future works, as reported for oak [*Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. (Durand et al., 2010)] and *Prunus* L. (Rahemi et al., 2012). Furthermore, because the reported SSRs have been localized in expressed regions, they could be included in future studies of mapping progenies to find quantitative trait loci, as reported for apple [*Malus domestica* L. (Cova et al., 2012)]. Finally, in addition to the hexa-nucleotide SSRs, the identification of many tri-, tetra-, and penta-nucleotide SSRs is currently in progress in the OLEAGEN EST sequencing project, which could also be of great interest for future studies on identification, variability, and mapping.

### Literature Cited

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

Baldoni, L. and A. Belaj. 2009. Olive, p. 397–421. In: Vollmann, J. and I. Rajean I (eds.). *Oil crops. Handbook of plant breeding. Vol 4.* Springer, New York, NY.

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

Belaj, A., C. Muñoz-Díez, L. Baldoni, Z. Satovic, and D. Barranco. 2010. Genetic diversity and relationships of wild and cultivated olives at regional level in Spain. *Sci. Hort.* 124:323–330.

Bradley, M.V. and W.H. Griggs. 1963. Morphological evidence of incompatibility in *Olea europaea* L. *Phytomorphology* 13:141–156.

Breton, C.M. and A. Bervillé. 2012. New hypothesis elucidates self-incompatibility in the olive tree regarding S-alleles dominance relationships as in the sporophytic model. *C. R. Biol.* 335:563–572.

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

Cipriani, G., M.T. Marrazzo, G. Di Gaspero, A. Pfeiffer, M. Morgante, and R. Testolin. 2008. A set of microsatellite markers with long core repeat optimized for grape (*Vitis* sp.) genotyping. *BMC Plant Biol.* 8:127.

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

Collani, S., G. Galla, A. Ramina, F. Alagna, E.M. Càceres, G. Barcaccia, L. Baldoni, R. Muleo, and G. Perrotta. 2012. Self-incompatibility in olive: A new hypothesis on the s-locus genes controlling pollen–pistil interaction. *Acta Hort.* 967:133–140.

Cova, V., D. Perini, V. Soglio, M. Komjanc, E. van de Weg, C. Gessler, and L. Gianfranceschi. 2012. Exploiting expressed sequence tag databases for mapping markers associated with fruit development and fruit quality in apple. *Mol. Breed.* 29:699–715.

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

De la Rosa, R., C.M. James, and K.R. Tobutt. 2004. Using microsatellites for paternity testing in olive progenies. *HortScience* 39:351–354.

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

Díaz, A., A. Martín, P. Rallo, D. Barranco, and R. De la Rosa. 2006b. Self incompatibility of ‘Arbequina’ and ‘Picual’ olives assessed by SSR markers. *J. Amer. Soc. Hort. Sci.* 131:250–255.

Díaz, A., A. Martín, P. Rallo, and R. De la Rosa. 2007. Cross-compatibility of the genitors as the main factor for successful olive breeding crosses. *J. Amer. Soc. Hort. Sci.* 132:830–835.

Durand, J., C. Bodenes, E. Chancerel, J.M. Frigerio, G. Vendramin, F. Sebastiani, A. Buonamici, O. Gailing, H.P. Koelewijn, F. Villani, C. Mattioni, M. Cherubini, P.G. Goicoechea, A. Herran, Z. Ikarán, C. Cabane, S. Ueno, F. Alberto, P.Y. Dumoulin, E. Guichoux, A. de Daruvar, A. Kremer, and C. Plomion. 2010. A fast and cost-effective approach to develop and map EST-SSR markers: Oak as a case study. *BMC Genomics* 11:570.

Faria, D.A., E.M.C. Mamani, G.J. Pappas, and D. Grattapaglia. 2011. Genotyping systems for *Eucalyptus* based on tetra-, penta-, and hexa-nucleotide repeat EST microsatellites and their use for individual fingerprinting and assignment tests. *Tree Genet. Genomes* 7:63–77.

Huang, H., J. Lu, A. Ren, W. Hunter, S.E. Dowd, and P. Dang. 2011. Mining and validating grape (*Vitis* L.) ESTs to develop EST-SSR markers for genotyping and mapping. *Mol. Breed.* 28:241–254.

Kumpatla, S.P. and S. Mukhopadhyay. 2005. Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. *Genome* 48:985–998.

Mookerjee, S., J. Guerin, G. Collins, C. Ford, and M. Sedgley. 2005. Paternity analysis using microsatellite markers to identify pollen donors in an olive grove. *Theor. Appl. Genet.* 111:1174–1182.

Muñoz-Mérida, A., A. Cañada, J.J. González-Plaza, A.M. Blanco, M.C. García-López, J.M. Rodríguez, L. Pedrola, M.D. Sicardo, M.L. Hernández, R. De la Rosa, A. Belaj, M. Gil-Borja, J.M. Martínez-Rivas, F. Luque, D.G. Pisano, O. Trelles, V. Valpuesta, and C.R. Beuzón. 2013. De novo assembly and functional annotation of the olive (*Olea europaea*) transcriptome. *DNA Res.* 20:93–108.

Nishio, S., T. Yamamoto, S. Terakami, Y. Sawamura, N. Takada, C. Nishitani, and T. Saito. 2011. Novel genomic and EST derived SSR markers in Japanese chestnuts. *Sci. Hort.* 130:838–846.

Noormohammadi, Z., M. Hosseini-Mazinani, I. Trujillo, L. Rallo, A. Belaj, and M. Sadeghizadeh. 2007. Identification and classification of main Iranian olive cultivars using microsatellite markers. *HortScience* 42:1545–1550.

Poncet, V., M. Rondeau, C. Tranchant, A. Cayrel, S. Hamon, A. de Kochko, and P. Hamon. 2006. SSR mining in coffee tree EST databases: Potential use of EST-SSRs as markers for the *Coffea* genus. *Mol. Genet. Genomics* 276:436–449.

Rahemi, A., R. Fatahi, A. Ebadi, T. Taghavi, D. Hassani, T. Gradziel, K. Folta, and J. Chaparro. 2012. Genetic diversity of some wild

- almonds and related *Prunus* species revealed by SSR and EST-SSR molecular markers. *Plant Syst. Evol.* 298:173–192.
- Rallo, L., D. Barranco, R. De la Rosa, and L. León. 2008. ‘Chiquitita’ olive. *HortScience* 43:529–531.
- Rallo, P., G. Dorado, and A. Martín. 2000. Development of simple sequence repeats (SSRs) in olive tree (*Olea europaea* L.). *Theor. Appl. Genet.* 101:984–989.
- Rozen, S. and H. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* 132:365–386.
- Sabino-Gil, F., M. Busconi, A. Da Câmara Machado, and C. Fogher. 2006. Development and characterization of microsatellite loci from *Olea europaea*. *Mol. Ecol. Notes* 6:1275–1277.
- Sefc, K.M., M.S. Lopes, D. Mendonça, M. Rodrigues dos Santos, M. Laimer da Camara Machado, and A. Da Camara Machado. 2000. Identification of microsatellite loci in olive (*Olea europaea*) and their characterization in Italian and Iberian olive trees. *Mol. Ecol.* 9:1171–1173.
- Seifi, E., J. Guerin, B. Kaiser, and M. Sedgley. 2012. Sexual compatibility of the olive cultivar ‘Kalamata’ assessed by paternity analysis. *Span. J. Agr. Res.* 10:731–740.
- Tessier, C., J. David, P. This, J.M. Boursiquot, and A. Charrier. 1999. Optimization of the choice of molecular markers for varietal identification in *Vitis vinifera* L. *Theor. Appl. Genet.* 98:171–177.
- Thiel, T., W. Michalek, R. Varshney, and A. Graner. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor. Appl. Genet.* 106:411–422.
- Vélez, M.D. and J. Ibáñez. 2012. Assessment of the uniformity and stability of grapevine cultivars using a set of microsatellite markers. *Euphytica* 186:419–432.
- Wöhrmann, T. and K. Weising. 2011. In silico mining for simple sequence repeat loci in a pineapple expressed sequence tag database and cross-species amplification of EST-SSR markers across Bromeliaceae. *Theor. Appl. Genet.* 123:635–647.