RESEARCH ARTICLE

Identification of an olive (*Olea europaea* L.) core collection with a new set of SSR markers

María Victoria Gómez-Rodríguez · Carmen Beuzon · Juan J. González-Plaza · Ana María Fernández-Ocaña

Abstract Simple Sequence Repeat (SSR) or microsatellite markers have become extremely useful tools in genetic identification and variability studies. To date around one hundred of these markers have been developed in the olive tree (*Olea europaea* L.subsp. *europaea*) and the majority of them are di- or tri-nucleotide. The analysis of the numerous articles recently published on plant microsatellite markers, shows however a higher relevance of markers with long core repeat motifs than those with di- or tri-nucleotide repeat motifs. This work presents a new set of highly informative and polymorphic tetra- and hexa-nucleotide SSR markers. The newly designed SSR markers have been employed to genotype the 36 olive cultivars of the core collection from Worldwide Olive Germplasm Bank of Cordoba, Spain (WOGBC). This article describes the allelic profiles of the 36 olive cultivars for a set of ten SSR markers and also a protocol for a multiplex polymerase chain reaction (PCR). The designed multiplex PCR can lead to an optimal amplification of up to five different SSR markers in the same reaction tube. Finally, the use of only four of the set of ten SSR markers employed, allows the differentiation of any cultivar of the WOGBC core collection from the remaining 35 cultivars.

M. V. Gómez-Rodríguez · A. M. Fernández-Ocaña (✉)
Departamento de Biología Animal, Biología Vegetal y
Ecología, Universidad de Jaén, Campus de Las Lagunillas
S/N, 23071 Jaén, Spain
e-mail: amocana@ujaen.es

C. Beuzon
Departamento de Biología Celular, Genética y Fisiología,
Universidad de Málaga, Avda. Cervantes, 2.,
29071 Málaga, Spain

J. J. González-Plaza
Faculty of Forestry and Wood Sciences, Česká
Zemědělská Univerzita V Praze Fakulta Lesnická a
Dřevařská, Praha, Czech Republic

Introduction

The olive tree (*Olea europaea* L. subsp. *europaea*), the most representative tree in Mediterranean countries, has been cultivated in this area for approximately 6000 years. It is estimated that there are more than 1200 local and old olive cultivars distributed in 54 countries worldwide and they can be found in a variety of environmental conditions and growing systems (Bartolini 2008). The world catalogue of olive cultivars collects information of 139 cultivars from 23

olive-growing countries that cover almost 85% of the olive crop area. Being a clonally propagated crop, olive cultivars have been traditionally maintained in ex situ field collections (Belaj et al. 2001, 2004b). Evaluation and conservation of olive germplasm are being performed in approximately 100 institutions with a regional, national or international scope (Bartolini 2008).

There is a controversy about the true origin of Mediterranean olive trees. The work of Besnard et al. (2013) supports the existence of three long-term shelters during the Quaternary glaciations in the Mediterranean that could have played a key role for the preservation of the genetic diversity of this plant species. These three shelters would be the Middle East (including *Cyprus)*, the Aegean area and the Strait of Gibraltar. However, the genome of *Olea* is not only related to the existence of areas of glacial retreat since it is also strongly influenced by both the biogeographical conditions of the Mediterranean basin and the human influence. The comparison of the geographical distribution of plastid genome diversity between wild and cultivated olive trees indicates the cradle of the first domestication in the northern Levant that was followed by dispersal across the Mediterranean basin, in parallel with the expansion of civilization and human exchanges in this part of the world (Besnard et al. 2013). According to these authors, humans have widely dispersed the chlorotype present in the East, where 90% of the olive cultivars share the same "eastern-like" chlorotype (Besnard et al. 2013).

The relationships among cultivated olive trees, wild forms and related subspecies need to be extensively explored to gain a better understanding of the genomic profile of wild populations and related subspecies (Green et al. 1989; Bartolini 2008; Belaj et al. 2010; Díez et al. 2011; Besnard et al. 2013; Barazani et al. 2014). A better understanding of their genetic structure would be the first step to clarify these issues (Green et al. 1989; Angiolillo et al. 1999; Belaj et al. 2011). The presence of synonyms (the same cultivar with different names) and homonyms (different cultivars with the same name) in olive cultivars, together with the recent and extensive diffusion of some cultivars (Arbequina, Frantoio, Koroneiki, etc.) out of their areas of origin, imposes the need for reliable and efficient tools for olive cultivar identification. Morphological and biological characters have been used for long for descriptive purposes to distinguish

olive cultivars (Cantini et al. 1999; Barranco et al. 2000, 2005; León et al. 2004). Agronomic characterisation also allowed the classification of different olive cultivars (Barranco and Rallo 2000). However, the use of the morphological characterisation is questionable because the expression of most morpho-biological traits is strongly affected by environmental conditions, the age of the trees, the training systems, and plant-phenological stage of the plants. Nevertheless, the morphological approach is still the initial step for the description and classification of olive germplasm (Rotondi et al. 2003).

A core collection consists of a limited number of selected accessions representing the genetic spectrum of the whole olive cultivar community. According to Brown (1995), it should include as much genetic diversity as possible. Belaj et al. defined in 2012 a core collection of 36 olive cultivars that includes the genetic spectrum of the WOGBC located at IFAPA (Instituto de Investigación y Formación Agraria y Pesquera de Andalucía) "Alameda del Obispo" Córdoba, Spain. It was designed through molecular markers (DArTs, SSRs and SNPs) and agronomic traits. Given its high average genetic distance and good representation of the different regions of the Mediterranean area in a relatively small number of varieties, the core collection from Belaj et al. (2012) is recognised as the most suitable for studies of olive genetic improvement.

In the last decades, the new molecular techniques have allowed the design of different genetic markers, employed, for example, to elucidate the variability of some crops (Trujillo et al. 1995, 2014; Belaj et al. 2001, 2004b; Fendri et al. 2010; Muzzalupo et al. 2010; Haouane et al. 2011; Atienza et al. 2013; Beghè et al. 2015) or for phylogenetic studies (Baldoni et al. 2006; Rubio de Casas et al. 2006; Belaj et al. 2007; Erre et al. 2010; Besnard et al. 2013). Other authors have combined the use of both morphological traits and molecular markers for identification purposes (Fendri et al. 2010; D'Imperio et al. 2011; Trujillo et al. 2014).

As pointed out by Belaj et al. (2018), most of the current identification efforts in plant germplasm collections are based on DNA markers. These authors compiled from several studies (Bracci et al. 2011; De Lorenzis et al. 2015; Mason et al. 2015; Marrano et al. 2017) the desirable properties that a molecular marker should fulfil for olive identification: availability of

many polymorphisms, co-dominant inheritance, high frequent occurrence, easy accessibility, low cost, quick and high throughput, high reproducibility and transferability among different laboratories and detection platforms. At present, SNPs and SSR markers have been the molecular markers more commonly employed to identify olive cultivars. Single nucleotide polymorphism (SNP) is a variation in a single nucleotide that occurs at a specific position in the DNA chain. SNPs are sequence-based and distinguished according to the nucleotide present at each given position, which confers them high reproducibility among laboratories and detection techniques (Bracci et al. 2011; Bevan et al. 2017). During the last years the information about SNPs in olive tree has strongly increased (Reale et al. 2006; Consolandi et al. 2007; Muleo et al. 2009; Hakim et al. 2010; Belaj et al. 2012; Dominguez-Garcia et al. 2012; Kaya et al. 2013; Biton et al. 2015; Ipek et al. 2016; Belaj et al. (2018).

Microsatellites or SSR markers are regions of DNA consisting of tandemly repeated units of mono-, di-, tri-, tetra-, penta- or hexa-nucleotides arranged throughout the genomes of most eukaryotic species (Powell et al. 1996). Over the past 20 years, SSRs have been the most widely used markers for genotyping plants as they are highly informative, co-dominant, multi-allele genetic markers that are experimentally reproducible and exhibit relatively high transferability among related species (Mason 2015). The advent of the genomic period has resulted in the production of vast amounts of publicly available DNA sequence data, including large collections of Expressed Sequence Tags (ESTs), a rich source of SSRs, with many advantages over SSR markers from genomic DNA, and a large number of applications, since they reveal polymorphisms not only within the source taxon but in related taxa as well (Ellis and Burke 2007). SSRs with core repeats from 3 to 6 nucleotides long, have been designed in some woody species such as *Prunus* (Aranzana et al. 2003; Dettori et al. 2015), *Vitis vinifera* (Riaz et al. 2004; Cipriani et al. 2008), *Malus domestica* (Silfverberg-Dilworth et al. 2006) among others, and they are quickly increasing to the detriment of the classical di-nucleotide SSR.

In olive tree, microsatellites have been used for many purposes (Bracci et al. 2011) including paternity analysis (De la Rosa et al. 2004; Díaz et al. 2007a, b), construction of linkage maps (De la Rosa et al. 2003; Wu et al. 2004), cultivar traceability in olive oil

(Martins-Lopes et al. 2008; Alba et al. 2009), DNA fingerprinting of cultivars (Sefc et al. 2000; Sarri et al. 2006; Baldoni and Belaj 2009), phylogenetic studies (Belaj et al. 2007; Erre et al. 2010), phylogeography and population genetics (Belaj et al. 2007; Khadari et al. 2008) and admixture events detection (Besnard et al. 2007; Díez et al. 2015). The high level of SSR transferability among olive tree subspecies (Rallo et al. 2003), combined with the level of polymorphism, makes SSR the current markers of choice for identification and variability studies (Trujillo et al. 2014; Belaj et al. 2012; Fendri et al. 2010; Muzzalupo et al. 2010; Haouane et al. 2011).

Despite their utility in olive tree less than 100 good and polymorphic SSR markers, have been developed to date (Sefc et al. 2000; Rallo et al 2000; Carriero et al. 2002; Cipriani et al. 2002; De la Rosa et al. 2002; Gil et al. 2006) and they have been extensively used by many researchers (e.g., De la Rosa et al. 2004; Belaj et al. 2004a; Sarri et al. 2006; Díaz et al. 2006 and D'Imperio et al. 2011). Haouane et al. (2011) used ten of the SSRs previously reported by different authors (Sefc et al. 2000; Carriero et al. 2002; Cipriani et al. 2002; De la Rosa et al. 2002) to study the genetic structure of the core collection from the World Olive Germplasm Bank of Marrakech. El Bakkali et al. (2013) employed a set of 17 markers from the previous authors to construct new core collections. De la Rosa et al. (2002) and Belaj et al. (2011) used 8 SSRs from Cipriani et al. (2002), together with agro-morphological traits, to analyse the variability of wild olive trees. Some more recent papers (Adawy et al. 2015; Mariotti et al. 2016; Arbeiter et al. 2017) have shown new sets of SSR markers, mainly di- or tri-nucleotides.

Even though the advances in SNP technology, the use of microsatellites remains as the predominant molecular tool for identification and characterization of olive cultivars. In its review, Sebastiani and Busconi (2017) highlighting the article from Dominguez-Garcia et al. (2012), concluded that SNPs are less polymorphic than microsatellites, although they showed an interesting level of polymorphism to study some cultivars from Algeria. These authors also recognized the necessity of developing more SNPs to make them as discriminative as SSRs. Belaj et al. (2018) have identified a new set of 1043 EST-SNP markers but according to the authors they display lower levels of genetic diversity than SSRs.

The main objectives of this work were to design and develop a highly efficient and reproducible set of long core repeat EST-SSRs, its use in multiplex PCR, and its validation in the identification of the 36 olive cultivars of the Belaj et al. (2012) core collection. The markers were designed to make the multiplexing easier. In particular, they allowed the design of primers that generated a wide range of allele sizes, labelled with four different fluorophores, in a very standard and unique PCR conditions. This set represents a potent tool to discriminate any other olive cultivar in the world, and is useful for studies of population genetic structure, genetic mapping and evolutionary processes. SSR markers with core repeat 4 to 6 nucleotides long are the election tool in the current analyses with SSR markers, especially if they permit their multiplexing.

Materials and methods

ESTs-SSR markers retrieving and primers design

The initial data for this work were taken from three cDNA libraries, sequenced through Sanger technology, in the framework of the OLEAGEN project (Muñoz-Mérida et al. 2013). The first library proceeded from buds taken from young and adult branches of 10 seedlings from the cross of cultivars "Picual" and"Arbequina". The second library came from "Lechín de Sevilla" fruit mesocarp at different maturation stages (green with lignified endocarp, turning and purple). 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".

Identification of SSRs in the three libraries was carried out using MIcroSAtellite software (MISA, https://www.pgrc.ipk-gatersleben.de/misa, Institute of Plant Genetics and Crop Plant Research, Gatersleben, Germany) (Thiel et al. 2003), following the methodology previously described by De la Rosa et al. (2013). Primers design was developed by Oligo 7 software (Primer Analysis Software Oligo 7.60. Molecular Biology Insights, Inc.; Cascade, CO 80,809, USA).

Plant material, DNA extraction and PCR amplification conditions

Total genomic DNA derived from 100 mg of fresh leaves of the 36 olive cultivars of the core collection obtained from the WOGBC. DNA extraction was carried out using a commercial kit, Phytopure, following the manufacturer's instructions (GE Healthcare). DNA quality was assessed by electrophoresis on 1% (w/v) agarose gels and quantification was performed using a microplate reader (model BioTek, Synergy HT). The first step of EST-SSR selection included only eight olive cultivars from the core collection: "Abbadi Abou Gabra-842", "Arbequina", "Chemlal de Kabylie", "Frantoio", "Koroneiki", "Manzanilla de Sevilla", "Maari" and "Picual". PCR amplification and reproducibility of the 40 initially designed non-fluorescent primers were tested using the selected cultivars. SSR amplification was carried out in a thermal cycler by Bio-Rad (MyCycler TM), in a final volume of 10 μl containing: 100 ng of genomic DNA, 0.25 U of AmpliTaq Gold® DNA polymerase, 2.5 mM MgCl$_2$ final concentration, 1 mM each dNTP mix (Roche), and 0.5 μM each, forward and reverse, non-fluorescent primers. The program used for PCR amplification was as follows: initial denaturation at 95 °C for 5 min; 35 cycles of denaturation at 95 °C for the 30 s, annealing at 50 °C to 55 °C for 30 s, extension at 72 °C for 30 s, and a final extension at 72 °C for 60 min. Detection of amplification was confirmed by electrophoresis in 2% agarose gel. From the initial set of 40 primer pairs, only 24 were selected on the basis of their polymorphism (Table 1).

The next step was the fluorescent dye-labelling of one of the two primers from each of the 24 selected SSR markers (forward or reverse) with FAM, NED, PET or VIC fluorophores (Applied Biosystems, Foster City, CA, USA) (Table 1). The expected fragment–size range for each locus (Table 1) was established according to the OLEAGEN database information. SSR markers with non-overlapping amplicons were labelled with the same fluorescent dye. In contrast, each of the markers that produced amplicons with the same size, were labelled with one different fluorophore.

Fluorescent-labelled markers were initially tested on the eight selected olive cultivars and amplification was checked by electrophoresis in 2% agarose gel. The

Table 1 Set of 24 SSR markers assayed with 8 olive cultivars of the core collection. The 8 finally selected are highlighted in bold type

Primers code	Type	Motif	5′–3′ Primer sequences	Expected amplicon size (bp)
Olea4.2F-FAM	p6	GCTCCG	TAA ATG GCT CGC TAT ATT CTG ATT C	153
Olea4.2R			AATAAGAGACCGAGATTTCAGGTTC	
Olea9.4F-VIC	**p6**	**GAAGAG**	**GTT AAA TGG CCA TAA GAT TGA TGA T**	**199**
Olea9.4R			**CAT TAC CTG TGT AGT GAT TCT ACC AAC**	
Olea25.5F-FAM	p5	TGGGT	GTG ATA GGG CTG GGC AAA AC	322
Olea25.5R			TGA AAT AAC ATT TCA TTT GTA AAT TAT GTT	
Olea29.1F-NED	p5	CCAAG	AAG CAG TGG TAT CAA CGC AGA GTA	136
Olea29.1R			TTT CTC CAT CAG CTC TCT CCT TTC	
Olea32F	p5	TTCTT	AGA GAC AAT TAG GGC GTA GAA TCG	161–300
Olea32R-NED			CAA GCA GTG GTA TCA ACG CAG A	
Olea33F	p5	TCCCT	AGA GAC AAT TAG GGC GTA GAA TCG	164–400
Olea33R-VIC			ACT GAG CGG GTA TGT GAT TGA AG	
Olea34.4F	p4	TTTC	GAT GGA TTA TAA TAC TCG TCG TTG AAC	277–500
Olea34.4R-FAM			AGG ACT AGA CAT GGA TTT GAA ACA TC	
Olea34.7F-VIC	p4	TTTC	TAT CGG TTT ATT TGA AGG TAA ATG C	103
Olea34.7R			ACT CTA GAA ATA CCA AAC ATA AAC CTT TC	
Olea34.14F-PET	p4	TTTC	GGA TGA ACT GGA CAG AGA ACA ATA C	120
Olea34.14R			ACT ATC TTC CAA ATC TGG AGA TAA AGT	
Olea37.8F-PET	p4	AATA	TTC CCC TTC CTC CAA GGT ATT AC	310–450
Olea37.8R			CTC CCT ATA AAG GTT GGA TGG ATA C	
Olea38.13F-PET	p4	TTTA	CAT AGC TCT CGA ATA CCA TCC TTA C	353
Olea38.13R			CAA ATC CCA CTT ATT CCT CTC TAT G	
Olea38.29F	p4	TTTA	TAT CTG GTG AAG TTT CTT GCA GTA TC	247
Olea38.29R-NED			TGT TGC TAT GAA GAA CTG TGT TTT G	
Olea39F-VIC	**p4**	**CTCG**	**GGA TCG CAC ATA TCA AAG ATT CTA C**	**141**
Olea39R			**GAC TCA GAG GGT GAT CTT AAT TCT TC**	
Olea39.1F-VIC	p4	CTCG	AGC AGT GGT ATC AAC GCA GAG TA	135–290
Olea39.1R			GAA ATT TTA GCT TCC GCA TCT TC	
Olea40.4F	p4	CTTT	AAA GAA AAC CTC CTA TAC ACA ATT AGA G	158
Olea40.4-RFAM			ACT ATC TTC CAA ATC TGG AGA TAA AGT	
Olea40.13F	**p4**	**CTTT**	**TTT TGT TCA TAA TTA AGC TTA TAT TAT TCC**	**93**
Olea40.13R- **NED**			**CAA TGG AAG TAG GGA TCT AAA TTA AAC**	
Olea41.2F	**p4**	**TCAT**	**TTT TCA TCC TCC ATC TCT ATG ATT G**	**119**
Olea41.2R-FAM			**AAT ATG ATG ATC CAG CAG CTT AAC A**	
Olea42.9F	**p4**	**GAAA**	**CTC CAT CTT TAC TCT CTA TAC ATC TCG TA**	**157**
Olea42.9R-NED			**GAA AGA GGC GTC AGA GTC AGA AC**	
Olea42.30F	**p4**	**GAAA**	**ATA CAA AAA GCA GAG TAT CTT CAC TGA**	**149**
Olea42.30R-PET			**AGC TCC TAT CTG TTC ATT TCA TAA AG**	
Olea42.31F	**p4**	**GAAA**	**ACT ATC TTC CAA ATC TGG AGA TAA AGT**	**120**
Olea42.31-NED			**GGA TGA ACT GGA CAG AGA ACA ATA C**	
Olea42.34F-PET	**p4**	**GAAA**	**GTA ACG GAG AAA GAT TGA ACA AGA T**	**301**
Olea42.34R			**TTT TGT TCA TAA TTA AGC TTA TAT TAT TCC**	

Table 1 continued

Primers code	Type	Motif	5´–3´ Primer sequences	Expected amplicon size (bp)
Olea42.36F	p4	GAAA	TTT GCC TAC CAT AAA AGT GTG CTA	96
Olea42.36R-PET			AAG GTG CAC TCA TGT TTG CAG TA	
Olea42.37F-VIC	p4	GAAA	CAG AGA GAG AGA ACA AAG AAA GGA TT	129
Olea42.37R			GCA TGT ATG TAT ATG TAT GCA TCA AGA A	
Olea42.40F-FAM	p4	GAAA	GCA GCT GTG CTG GTA TGG AAG	267–400
Olea42.40R			CTC CAT TAC CCT TGA ATA GCT TTC T	

evaluation of polymorphism was carried out with an automated sequencer 3500 genetic analyzer (Life Technologies) using the internal standard GeneScan™ 600 LIZ® Size Standard v2.0 (Life Technologies).

The final number of SSR markers selected to identify the 36 olive cultivars was ten: eight specially designed for this work and two previously described by De la Rosa et al. (2013) (Table 1).

PCR multiplex design

Once the set of ten SSR markers had been separately tested, and their allelic profiles described, the possibility of making multiplex PCR was studied. Amplification reactions were performed in a total volume of 10 μl with 100 ng of template DNA, 0.5 μM of primers labelled either with 6-FAM, NED, PET or VIC fluorophores, 0.5 μM of unlabelled primers, 1 mM each dNTP mix (Roche), 2 mM MgCl2, 1× Buffer Gold and 0.25 U of AmpliTaq Gold® DNA polymerase. The PCR program was a touchdown one with the following steps: initial denaturation at 95 °C for 5 min, 20 cycles of denaturation at 95 °C for 30 s and annealing at 65 °C, 10 cycles of denaturation at 95 °C for 30 s and annealing at 55 °C for 30 s, 10 cycles of denaturation at 95 °C for 30 s and annealing at 50 °C, and a final extension at 72 °C for 30 min. Labelled amplification products were resolved onto an automated sequencer 3500 genetic analyzer (Life Technologies) using the internal standard GeneScan™ 600 LIZ® Size Standard v2.0 (Life Technologies) The PCR fragments were detected with the GeneMarker analysis software version 2.00.

As mentioned above, all pairs of primers were designed to allow the indiscriminate mixing of them in multiplexing without giving rise to errors of interpretation (Fig. 1). The optimal number of markers that can be introduced in the same PCR tube without causing problems of lack of amplification is five or less. More than five produces the absence of amplification of some markers.

During the development of PCR assays, successful and repetitive amplifications were usually obtained in all genotypes analysed. In order to exclude errors due to DNA concentration or quality, two additional tests were carried out in genotypes where no amplification was achieved. When one cultivar did not show any allele for a marker (null), the analysis was repeated once again to confirm the result.

Data analysis

The allele profiles and fragment analysis of the olive cultivars were characterised using GeneMarker software. The electropherograms displayed fluorescent signal intensities as a single line trace for each dye colour: FAM, VIC, NED and PET. After having uploaded the raw data files, its processing included the application of a sizing standard, filtering of noisy peaks, and comparison to a known allelic panel. The common size standard used was LIZ600.

The peaks from each of the ten primer pairs employed to identify the 36 olive cultivars were analysed. Those genotypes showing a single peak at a given locus were recorded as homozygous. The absence of any peak in the expected range size was checked twice before it was confirmed as null genotype. The statistical analysis of heterozygosity, number of alleles and their frequency, and values of the polymorphic information content (PIC) was performed using PowerMarker V3.0 software (Liu and

(A)

(B)

Picudo cultivar: 42.31 (118-122), 42,30 (152-152), 42.9 (156-160), 9.4 (195-195), H2 (202-202)

Fig. 1 a Amplicon size range of the 10 SSR markers employed to characterise the 36 olive cultivars of the core collection defined by Belaj et al. (2012). Each amplicon range is represented with the same colour than the flourophore linked to the marker. Multiplexing with no more than 5 different markers at the same time is possible because those markers overlapping in amplicon size are linked to fluorophores that produce a signal of a different colour. Markers named Oleagen were developed by De la Rosa et al. (2013). b Example of an electropherogram resulting from a multiplex PCR with five SSR markers

Muse 2005) (https://www.powermarker.net) and GenAlex V6.5 (Peakall and Smouse 2012) (https://biology-assets.anu.edu.au/GenAlEx/Welcome.html).

Summary statistics (Table 2) were calculated with Power Marker V3 software (Liu and Muse 2005) except for Dj for which GenAlex 6.5 (Peakall and Smouse 2012) was used. Each SSR marker was characterised by the following parameters: *Na* is the number of actual alleles per SSR marker. *Ne* is the number of effective alleles, that is, the number of equally frequent alleles that would take to achieve the same expected heterozygosity than in the studied population. *Ne* allows comparison of populations where the number and distributions of alleles differ

drastically. *Major allele frequency* is the higher frequency of an allele profile, at a particular locus from a population, expressed as a fraction per one unit. *Genotype number* is the number of different allelic combinations found. *Ho* and *He* are the observed and expected heterozygosity respectively. When the heterozygosity is high, the effective number of alleles is also high. *Gene diversity* is defined as the possibility that two randomly chosen alleles from the population are different. *PIC (Polymorphic information content)* is a parameter often used to measure the discriminatory capacity of an SSR marker. PIC takes into consideration the number of alleles present at a marker locus, and the frequency of these alleles.

Table 2 Statistical analysis of the quality of SSR markers tested in this work, their peaks definition and the number of cultivars with null genotype that they produce

Primers code	Na	Ne	Major Allele Frequency	Genotype No	Ho	He	Gene Diversity	PIC	Dj	Peaks definition	Type	Number of cultivars with null genotype
Oleagen-H2	7	5.02	0.31	18.00	0.86	0.80	0.80	0.77	0.82	Perfect	P6	None
Oleagen-H20.1	5	3.84	0.31	10.00	0.75	0.74	0.74	0.69	0.76	Perfect	P6	None
Olea42.31	7	2.17	0.58	8.00	0.64	0.54	0.60	0.49	0.55	Perfect	P4	None
Olea9.4	6	4.05	0.28	12.00	0.35	0.75	0.79	0.76	0.78	Perfect	P6	10
Olea39	5	3.75	0.35	12.00	0.71	0.73	0.75	0.71	0.75	Several double peaks	P4	1
Olea40.13	5	2.44	0.29	8.00	0.56	0.59	0.79	0.76	0.61	Several double peaks	P4	2
Olea41.2	4	2.65	0.54	7.00	0.67	0.62	0.61	0.55	0.64	Good but with some double peaks	P4	None
Olea42.34	5	2.52	0.54	7.00	0.64	0.60	0.60	0.54	0.62	Double peaks but clear	P4	None
Olea42.9	4	2.29	0.49	6.00	0.64	0.56	0.56	0.47	0.58	Double peaks but clear	P4	None
Olea42.30	7	2.47	0.64	10.00	0.67	0.60	0.54	0.55	0.61	Stuttered peaks	P4	None
Average	5.25	3.17	0.43	9.80	.64	0.66	0.68	0.63	0.68			

Na = number of actual alleles per SSR marker; Ne (number of effective alleles) = number of equally frequent alleles that would take to achieve the same expected heterozygosity than in the studied population; *Major allele frequency* = the higher frequency of an allele profile, at a particular locus from a population, expressed as a fraction per one unit; *Genotype number* = number of different allelic combinations found; *Ho* = observed heterozygosity; *He* = expected heterozygosity; *Gene Diversity* = the possibility that two randomly chosen alleles from the population are different; *PIC* (polymorphic information content) takes into consideration the number of alleles present at a marker locus, and the frequency of these alleles; *Dj* (Discrimination power) = the way to evaluate the efficiency of a primer for the purpose of identifying varieties; *Type* = number of base pairs in the repetitive unit of the microsatellite

Consequently, loci with a large number of alleles usually have higher PICs, although the PIC is also influenced by the frequency of these alleles. It is possible to have a large number of alleles and a relatively small PIC if one or two of the alleles predominate. PIC values range from 0 to 1. According to Botstein et al. (1980) three categories are defined: high (PIC > 0.5), moderate (0.5 > PIC > 0.25), and low (PIC < 0.25). From the PIC parameter, Tessier et al. (1999) defined *Dj* as a way to evaluate the efficiency of a primer for the purpose of identifying varieties.

Results and discussion

Selection and checking of the SSR markers

As it has already been described in material and methods, the 40 initial non-fluorescent hexa, penta and tetra-nucleotide SSRs were reduced to a set of 24 selected primer pairs after PCR amplification assays (Table 1) that were linked to a fluorescent dye. The selection relied on their amplification capacity and the high polymorphism they showed in agarose gels. From the set of 24 primer pairs that at first sight seemed to be polymorphic, 16 generated the same allelic pattern in all the olive cultivars assayed and therefore, they were rejected. In summary, from the 40 initial SSRs, only 8

were chosen. The final set of SSR markers selected to characterise the cultivars of the olive core collection WOGBC, was composed of ten: 8 the newly described ones above mentioned, plus 2 hexa-nucleotide (Oleagen-H20.1 and Oleagen-H2) previously described by De la Rosa et al. (2013) and tested here for the first time on all the 36 cultivars of the core collection.

Table 3 shows the amplification results of the set of ten SSR markers on the 36 olive cultivars studied, and the allele size range of every SSR marker employed. Differences of 1 bp between alleles from different cultivars were checked by re-amplification to establish whether a coding error had occurred. A total of 98 different genotypes were produced. The null genotypes can be the result of mutations in the flanking region at primer binding sites (Guichoux et al. 2011). Nulls have not been taken into account in the general record of alleles. Nevertheless, the absence of alleles in a cultivar has been considered as one genotype (null).

Multiplex SSR set coupled with fluorescent detection systems have already shown to be relevant and successfully applied in plant genetic studies. Nevertheless, most of these studies are based on SSR markers not originally developed to be multiplexed and thus, in general, these markers are available for their use in multiplexing in very low number (Merdinoglu et al. 2005). In contrast, our approach is based on the specific design of SSR markers for their multiplexing (pre-PCR) and multiloading (post-PCR). The optimisation of this multiplex-PCR was one of the more time-consuming tasks in this work and had several critical steps: (i) the new touchdown-PCR conditions to save costs and time of work, (ii) the determination of the best combinations of grouped fluorescent labelled markers, and (iii) the optimal number of markers in each multiplex PCR that permits a good amplification of all of them.

The development of multiplex PCR has been considered especially important because it represents a clear benefit by reducing laboratory work and consumption of expensive reagents without compromising test accuracy (Guichoux et al. 2011). On the one hand, the method is faster, because in every PCR tube up to five different markers can be amplified with very good results. On the other hand, the multiplex PCR allows a considerable costs save in reagents and sequencing processes. The touchdown PCR is the technique of choice in the majority of multiplex PCR experiments (Hill et al. 2009; Guichoux et al. 2011) as it allows to amplify heterogeneous SSR primers, with different annealing temperature by reducing this parameter in successive annealing cycles.

Figure 1a shows the amplicon size range of each of the 10 SSR markers. Each amplicon range is represented with the same colour that the fluorophore linked to its marker. As seen in this figure, those markers producing amplicons with the same size, are labelled with fluorophores of different colour. Figure 1b) shows an example of an electropherogram resulting from a multiplex PCR with five different SSR markers

Evaluation of SSR markers polymorphism and discrimination power

A total of 36 cultivars, the core collection, that have been chosen from the WOGBC and whose origins are the main Mediterranean olive-cultivating countries, were genotyped with the set of ten selected SSR markers.

SSR markers were classified (Table 2) on the basis of the allelic profile they produced, using as selection criteria: a) the presence of sharp peaks, b) the number of different alleles and genotypes they revealed and c) the number of cultivars with a null genotype they produce. ESM_1 shows the allele combinations obtained with the best markers (Oleagen-H2, Oleagen-H20.1 and Olea42.31). They are considered the best because they provide electropherograms in which alleles are clearly defined and they do not give null genotype in any of the cultivars assayed. ESM_1 also includes the allele combinations obtained with Olea9.4. This last marker produces also very well defined peaks, but it gives null genotype in ten of the cultivars. The second group in order of quality (ESM_2) consists of Olea39, Olea40.13, Olea41.2, Olea42.34 and Olea42.9. These five markers provide clearly distinguishable alleles, although sometimes appear double peaks for a single allele. According to Clark (1988) and Esselink et al. (2003), double peaks are caused by the non-template addition of a nucleotide (generally an adenine) to PCR fragments by the *Taq* polymerase. When the adenylation is incomplete, in the resulting electropherogram appears one peak from the original fragment and an additional peak 1 bp longer corresponding to the adenylated fragment (Guichoux et al. 2011). The marker that gave worse results in terms of peaks definition was Olea42.30

Table 3 Genotype of the 36 cultivars studied at each of the 10 SSR markers tested. SSR markers are grouped according to their quality. Oleagen H2 and Oleagen-H20.1 were previously designed by de la Rosa et al. (2013)

Cultivar	Peaks definition of SSR markers									Stuttered peaks
	Perfect				Good but with some double peaks					
	Oleagen-H2	Oleagen-H20.1	Olea 42.31	Olea 9.4	Olea 39	Olea 40.13	Olea 41.2	Olea 42.34	Olea 42.9	Olea 42.30
Allele size range	196–226	230–251	93–126	165–195	123–139	75–95	104–119	277–298	148–164	142–160
Abbadi Abou Gabra-842	214–226	233–239	114–118	195–195	131–139	83–83	115–115	288–288	156–160	152–160
Abou Satl Mohazam	205–220	239–251	118–118	195–195	131–139	83–91	115–119	288–296	156–160	152–152
Aboukanani (Abou kanani)	202–226	233–239	118–122	189–189	Null	83–91	115–115	288–296	156–160	152–152
Arbequina	202–214	233–245	118–118	177–177	135–135	75–91	111–119	280–296	156–156	152–160
Barnea	196–214	245–251	118–118	Null	135–139	75–83	111–115	280–288	148–160	142–152
Barri	202–214	233–245	113–118	171–189	127–131	83–95	119–119	288–300	160–160	152–160
Chemlal de Kabylie	202–214	239–245	118–118	171–171	127–139	83–83	111–119	280–288	156–160	138–152
Chengue (Shengeh	205–220	233–233	114–118	189–189	123–135	75–75	115–119	280–296	156–160	152–160
Clon-14 (Klon-14–1812)	196–214	245–245	118–122	189–195	131–131	Null	115–115	288–288	156–156	152–152
Dokkar	202–220	239–245	118–118	Null	127–127	75–83	111–119	280–288	148–160	142–148
Forastera de Tortosa	202–202	233–245	113–118	189–189	127–131	83–83	115–115	288–288	156–156	152–160
Frantoio	202–208	245–245	118–122	165–165	135–135	Null	111–115	280–288	156–156	152–160
Grappolo	208–226	245–251	118–122	165–189	135–139	83–83	115–119	288–288	156–160	152–160
Jabali	214–214	251–251	110–118	171–171	123–139	75–91	104–104	280–296	160–164	152–160
Kalamon	214–226	239–245	110–118	195–195	131–139	75–83	111–115	280–288	156–160	152–152
Koroneiki	202–214	239–245	118–122	171–171	130–135	75–83	111–119	280–288	148–156	142–152
Leccino	205–208	245–251	118–122	Null	135–139	83–83	115–119	288–288	156–160	152–160
Llumeta	202–205	233–245	114–118	Null	131–135	83–83	115–115	288–288	156–156	142–152
Maarri	214–220	233–251	93–93	195–195	131–139	75–83	111–111	280–288	160–160	152–154
Manzanilla de Sevilla	202–205	239–251	113–118	189–195	131–139	75–83	111–115	280–288	156–160	152–152
Manzanillera de Huércal Overa	205–214	233–239	118–118	Null	127–139	83–95	115–115	288–300	156–160	152–152
Maari	205–214	233–233	118–122	Null	131–135	75–87	115–115	280–292	156–156	152–160
Mastoidis	202–214	245–245	113–118	Null	135–135	75–75	115–119	280–280	156–160	152–160
Maureya (Mavreya)	202–208	233–239	118–122	189–189	131–135	75–83	111–115	280–288	160–160	148–152
Mechjul-1013 (Majhol-1013)	205–205	239–251	118–122	171–195	131–131	83–91	115–115	288–296	156–156	152–152
Medjhoul (Majhol-152)	196–208	233–245	114–118	177–183	131–131	75–83	115–115	280–288	156–156	152–152
Megaritiki	202–214	239–245	118–118	Null	127–135	83–83	111–111	288–288	156–160	148–152
Menya (Menya-669)	196–214	233–239	113–118	Null	135–139	75–91	111–115	280–296	156–160	142–160
Morrut	202–205	239–245	118–122	195–195	135–135	75–75	115–115	280–280	156–160	152–160
Myrtolia	202–214	245–245	93–102	177–177	131–135	75–75	111–115	280–280	156–160	152–160
Picual	205–226	239–251	118–122	189–195	131–131	83–83	111–115	288–288	156–160	148–152

Table 3 continued

Cultivar	Peaks definition of SSR markers									
	Perfect				Good but with some double peaks					Stuttered peaks
	Oleagen-H2	Oleagen-H20.1	Olea 42.31	Olea 9.4	Olea 39	Olea 40.13	Olea 41.2	Olea 42.34	Olea 42.9	Olea 42.30
Picudo	202–202	239–251	118–122	195–195	131–139	83–83	111–115	288–288	156–160	152–152
Piñonera	202–202	230–239	118–118	183–189	131–135	75–83	115–119	280–288	148–160	142–160
Temprano	202–226	239–251	110–122	Null	131–131	83–83	111–115	288–288	156–160	152–152
Uslu	196–226	233–251	113–118	189–189	131–139	75–83	115–115	280–288	160–160	152–152
Verdial de Velez-Malaga-1	202–205	230–239	114–118	189–195	131–139	83–83	111–115	288–288	160–160	145–152

(Table 2 and ESM_3). It produced electropherograms that were difficult to interpret, at least in some cultivars, because of their stuttered peaks. Reproducible and polymorphic amplification products were obtained, displaying from 4 to 7 different alleles per locus (Table 2).

It has been suggested that a research assessing an array of possible primer pairs, should select those associated with di-nucleotide repeats over more elaborated motif lengths (tri-, tetra-, or penta-nucleotide motifs) to ensure higher levels of genetic variation (Levinson and Gutman 1987; Grist et al. 1993; Chakraborty et al. 1997; Sup Lee et al. 1999; Ellegren 2004). In fact, most (48–67%) microsatellite markers found in many species are di-nucleotide repeats, although they are less frequent in coding regions (Li et al. 2002). Tri-nucleotide and hexa-nucleotide repeats are thought to be more common in coding regions because they do not cause any change in the frameshift (Tóth et al. 2000; Ellegren 2004).

All SSR markers employed in this work have shown a low number of alleles compared to di-nucleotide SSRs although they are in accordance with previous studies on the olive tree (De la Rosa et al. 2013). Even though long core SSR markers have a lower number of alleles than di-nucleotides (Nishio et al. 2011; Poncet et al. 2006; Rahemi et al. 2012), they are more appreciated because they produce wider distances among alleles and less stuttered peaks, contributing to a more reliable scoring of microsatellites (Dettori et al. 2015).

Oleagen-H2 and Oleagen-H20.1 SSR markers have been previously employed with some of the olive tree core collection cultivars ('Arbequina' 'Frantoio', 'Manzanilla de Sevilla', 'Koroneiki' and 'Picual') (De la Rosa et al. 2013). This allows us to confirm the reproducibility of the results obtained with them by both laboratories. All results obtained in this research agreed with those reported in the work above mentioned, which provides evidence of the quality of the markers. Oleagen-H2 and Oleagen-H20.1 showed the same allele profile in the common cultivars assayed except for minor displacements of no more than one or two nucleotides in the position of the peak.

Interestingly, 17 of the 36 cultivars studied could be identified using only one marker, because it reveals a specific and exclusive genotype (ESM_4). Jabali cultivar is a striking case because it shows an exclusive genotype with five different markers. With respect to SSRs, Oleagen-H2 is the one that produces more exclusive genotypes (nine), followed by Olea9.4 with six. Noticeably, Oleagen-H2 and Olea9.4 are hexa-nucleotide markers, which reinforces the idea of higher cultivar discriminating ability of polynucleotide microsatellites compared to di-nucleotide ones. Olea42.31 is the only marker that does not reveal any exclusive genotype.

Through the use of only four SSR markers, one cultivar can be differentiated from the rest of the core collection. Nevertheless, in 34 of the 36 cultivars, three SSR markers (one trio) is just enough for this purpose. There are three possible trios: Oleagen-H2,

Olea41.2 and Olea42.9 (ESM_5), the best one; Oleagen-H2, Oleagen-H20.1 and Olea41.2 (ESM_6) and Oleagen-H2, Olea42.34 and Olea42.9 (ESM_7). From each trio, there are only two cultivars that cannot be distinguished one from another (Chenge (Shengeh) and Abou Satl Mohazam when using the first trio; Chemlal de Kabylie and Koroneiki when using the second trio and Myrtolia and Mastoidis when using the third trio). In these cases, there are several SSR markers that can be used as an additional fourth marker to differentiate between those cultivars. ESM_8 shows the different possibilities of using a new fourth marker to differentiate the only two cultivars with the same genotype. The presence of null alleles can lead to an interpreting mistake when they are in heterozygosis because the single peak they show might be interpreted as homozygous instead of a heterozygous including a null allele (Dakin and Avise 2004). The advantage of the three trios proposed here is that none of them includes a marker that produces null genotype.

Once proved the discriminating power of this set of ten markers, the next step is to use this tool to genotype the hundreds of cultivars that are present in the WOGBC. Evidently, when faced with this task of genotyping new cultivars not included in the core collection, the method of choice is to employ the entire set of ten markers, once the multiplex-PCR has been perfected. Although the ten SSR markers can be mixed indiscriminately, to avoid confusion in the sizes of some alleles, it is advisable neither mix Olea40.13 with Olea42.30 nor 344 Oleagen-H2 with Oleagen-H20.1 due to the proximity between the size of alleles and the fluorophore used with them.

Polymorphism and diversity study. The capability of SSRs to identify the olive cultivars from the core collection.

Statistical analysis of the SSR markers shown in Table 2, reveals an allele number per SSR, Na, varying from 4 to 7, with an average of 5.25. These values could be considered as relatively low, but they are common in long core repeat molecular markers (De la Rosa et al. 2013; Cipriani et al. 2008). Ne values, the effective number of alleles, ranged from 2.17 and 5.02 with an average of 3.17. Ho and He ranged from 0.35 to 0.86, and from 0.54 to 0.80 respectively, with an average of 0.64 and 0.66. As these parameters can reach values from zero (no heterozygosity) to nearly 1.0 (for a system with a large number of equally

frequent alleles) these results are similar to those described by De la Rosa et al. (2013) for *Olea europaea* and Cipriani et al. (2008) for *Vitis vinifera*.

The Major allele frequency parameter ranged from 0.28 for Olea9.4 SSR to 0.64 for Olea42.30 marker. Gene diversity showed a range of high values (from 0.54 for Olea42.31 SSR to 0.80 for Oleagen-H2 marker). Regarding PIC values, all microsatellites employed in this research showed a value higher than 0.5, except for two of them: Olea42.9 (PIC = 0.47) and Olea42.31 (PIC = 0.49) which means that they have a high discriminatory capacity.

As mentioned in Material and Methods, both Dj and PIC values are based on allele frequencies, and therefore, both parameters have similar values. Dj ranged from 0.55 for Olea42.31 to 0.82 for Oleagen-H2. The three markers with higher Dj values were: Oleagen-H2 (0.82); Olea9.4 (0.78) and Oleagen-H20.1 (0.76). In general, the discriminatory power of the SSRs tested in the present research is similar to that described by De la Rosa et al. (2013).

Comparison of data from this work and the results from other core collections

Over the last decade, several core subsets have been proposed for both annual species, e.g. *Arabidopsis thaliana* (McKhann et al. 2004); *Oryza sativa* (Zhao et al. 2010), *Triticum aestivum* (Balfourier et al. 2007) and *Zea mays* (Franco et al. 2005), and perennial species, e.g. *Annona cherimola* (Escribano et al. 2008), *Malus domestica* (Richards et al. 2009), *Prunus armeniaca* (Wang et al. 2011) and *Vitis vinifera* (Le Cunff et al. 2008) using different eco-geographical, agro-morphological, biochemical or molecular data.

Trujillo et al. (2014), used 33 out of 77 (43%) di-nucleotide SSRs previously developed by the same authors to identify, together with morphological markers, the WOGBC. The same 33 sequences were used again by Díez et al. (2015). Some of them were discarded and 72% of the previously chosen remained. Muzzalupo et al. (2014) employed 11 SSRs designed by the same authors to analyse the genetic biodiversity of Italian olive trees. Chalak et al. (2015) worked on the genetic diversity in Lebanese olive trees, using 12 out of 17 SSR previously employed by Haouane et al. in 2011. Also, Ipek et al. (2015) have recently employed 20 of the same microsatellites to study the contents of fatty acids in olive oil and the genetic

diversity of an olive tree core collection. Las Casas et al. (2014) carried out a molecular characterization of olive Sicilian cultivars using 8 new di-nucleotide SSR markers designed by themselves. However, molecular discrimination of all of these core subsets were based on di-nucleotide SSRs

In the last years, De la Rosa et al. (2013) have designed a set of 8 new hexa-nucleotide SSRs based on 54-core sequences with long repeats, which is 15% of the previously chosen. These SSR loci were generated on the basis of ESTs in the frame of an olive genomic project (Muñoz-Mérida et al. 2013). Looking at other woody crops, only the study of Cipriani et al. (2008) have used polynucleotide SSRs. They tested a set of 94 long core repeat SSRs from *Vitis vinifera* selecting 38 of them (40.4%).

The present work started using 40 new long-core repeat sequences but only 8 of them (20%) were finally selected to be employed in order to discriminate the 36 cultivars of the core collection developed by Belaj et al. (2012). The excellent sharpness and discriminatory capability of electropherograms obtained with these SSR markers highlight their quality (ESM_1 to ESM_3). The final set of SSR consisted of 10, 8 of them specifically designed for this work, plus 2 previously designed by De la Rosa et al. (2013. As a probe of the discrimination power of the set, 94.4% of the cultivars can be discriminated from the rest of the core collection, using only three of these markers (ESM_5 to ESM_7).

The sequences of the microsatellite loci developed in this study represent a new and informative set of markers which can be easily combined for multiplexing and multi-loading according to the needs of any user and thus suitable for large-scale genetic analyses in olive. This work has made possible to obtain a set of SSR markers that produce reliable allelic profiles of 36 olive cultivars of the core collection from WOGBC and represents a powerful tool for genetic and plant breeding.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Adawy SS, Mokhtar MM, Alsamman AM, Sakr MM (2015) Development of annotated EST-SSR database in olive (*Olea europaea*). Int J Sci Res 4(9):1063–1073

Alba V, Sabetta W, Blanco A, Pasqualone A, Montemurro C (2009) Microsatellite markers to identify specific alleles in DNA extracted from monovarietal virgin olive oils. Eur Food Res Technol 229:375–382

Angiolillo A, Mencuccini M, Baldoni L (1999) Olive genetic diversity assessed using amplified fragment length polymorphisms. Theor Appl Genet 98(3–4):411–421

Aranzana M, Pineda A, Cosson P, Dirlewanger E, Ascasibar J, Cipriani G, Arús P (2003) A set of simple-sequence repeat (SSR) markers covering the Prunus genome. Theor Appl Genet 106(5):819–825

Arbeiter AB, Hladnik M, Jakše J, Bandelj D (2017) Identification and validation of novel EST-SSR markers in olives. Scientia Agricola. https://doi.org/10.1590/1678-992x-2016-0111

Atienza SG, De la Rosa R, Domínguez-García MC, Martín A, Kilian A, Belaj A (2013) Use of DArT markers as a means of better management of the diversity of olive cultivars. Food Res Int 54(2):2045–2053

Baldoni L, Belaj A (2009) Olive. In: Vollmann J, Rajean I (eds) Oil crops. Handbook of plant breeding 4, Springer Science Business Media, New York, pp 397–421

Baldoni L, Tosti N, Ricciolini C, Belaj A, Arcioni S, Pannelli G, Porceddu A (2006) Genetic structure of wild and cultivated olives in the central Mediterranean basin. Ann Bot 98(5):935–942

Balfourier F, Roussel V, Strelchenko P, Exbrayat-Vinson F, Sourdille P et al (2007) A worldwide bread wheat core collection arrayed in a 384-well plate. Theor Appl Genet 114:1265–1275

Barazani O, Westberg E, Hanin N, Dag A, Kerem Z, Tugendhaft Y, Kadereit JW (2014) A comparative analysis of genetic variation in rootstocks and scions of old olive trees—a window into the history of olive cultivation practices and past genetic variation. BMC Plant Biol 14(1):1

Barranco D, Rallo L (2000) Olive cultivars in Spain. HortTechnology 10(1):107–110

Barranco D, Cimato A, Fiorino P, Rallo L, Touzani A, Castañeda C, Trujillo I (2000) World catalogue of olive varieties. International Olive Oil Council, Madrid

Barranco D, Ruiz N, Gómez-del Campo M (2005) Frost tolerance of eight olive cultivars. HortScience 40(3):558–560

Bartolini G (2008) Olive germplasm (*Olea europaea* L.), cultivars, synonyms, cultivation area, collections, descriptors. https://www.oleadb.it

Beghè D, Molano JFG, Fabbri A, Ganino T (2015) Olive biodiversity in Colombia. a molecular study of local germplasm. Sci Hortic 189:122–131

Belaj A, Trujillo I, De la Rosa R, Rallo L, Gimenez MJ (2001) Polymorphism and discrimination capacity of randomly amplified polymorphic markers in an olive germplasm bank. J Am Soc Hortic Sci 126(1):64–71

Belaj A, Cipriani G, Testolin R, Rallo L, Trujillo I (2004a) Characterization and identification of the main Spanish and Italian olive cultivars by simple-sequence-repeat markers. HortScience 39(7):1557–1561

Belaj A, Rallo L, Trujillo I, Baldoni L (2004b) Using RAPD and AFLP Markers to Distinguish Individuals Obtained by Clonal Selection of Arbequina and Manzanilla de Sevilla Olive. HortScience 39(7):1566–1570

Belaj A, Muñoz-Díez C, Baldoni L, Porceddu A, Barranco D, Satovic Z (2007) Genetic diversity and population structure of wild olives from the north-western Mediterranean assessed by SSR markers. Ann Bot 100(3):449–458

Belaj A, Muñoz-Díez C, Baldoni L, Satovic Z, Barranco D (2010) Genetic diversity and relationships of wild and cultivated olives at regional level in Spain. Sci Hortic 124(3):323–330

Belaj A, León L, Satovic Z, De la Rosa R (2011) Variability of wild olives (*Olea europaea* subsp. europaea var. *sylvestris*) analyzed by agro-morphological traits and SSR markers. Scientia Hortic 129(4):561–569

Belaj A, del Carmen D-G, Atienza SG, Urdíroz NM, De la Rosa R, Satovic Z, Del Río C (2012) Developing a core collection of olive (*Olea europaea* L.) based on molecular markers (DArTs, SSRs, SNPs) and agronomic traits. Tree Genet Genom 8(2):365–378

Belaj A, De La Rosa R, Lorite IJ, Mariotti R, Cultrera NG, Beuzón CR, González-Plaza JJ, Muñoz-Mérida A, Trelles O, Baldoni L (2018) Usefulness of a new large set of high throughput EST-SNP Markers as a tool for olive germplasm collection management. Front Plant Sci 9:1320

Besnard G, Henry P, Wille L, Cooke D, Chapuis E (2007) On the origin of the invasive olives (*Olea europaea* L, Oleaceae). Heredity 99(6):608–619

Besnard G, Khadari B, Navascués M, Fernández-Mazuecos M, El Bakkali A, Arrigo N, Savolainen V (2013) The complex history of the olive tree: from Late Quaternary diversification of Mediterranean lineages to primary domestication in the northern Levant. Proc R Soc Lond B Biol Sci 280(1756):2012–2833

Bevan MW, Uauy C, Wulff BB, Zhou J, Krasileva K, Clark MD (2017) Genomic innovation for crop improvement. Nature 543:346–354. https://doi.org/10.1038/nature22011

Biton I, Doron-Faigenboim A, Jamwal M, Mani Y, Eshed R, Rosen A et al (2015) Development of a large set of SNP markers for assessing phylogenetic relationships between the olive cultivars composing the Israeli olive germplasm

collection. Mol Breed 35:107. https://doi.org/10.1007/s11032-015-0304-7

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32(3):314

Bracci T, Busconi M, Fogher C, Sebastiani L (2011) Molecular studies in olive (*Olea europaea* L.): overview on DNA markers applications and recent advances in genome analysis. Plant Cell Rep 30(4):449–462

Brown AHD (1995) The core collection at the crossroads. In: Hodgkin T, Brown AHD, van Hintum ThJL, Morales EAV (eds) in Core collections of plant genetic resources. Wiley, New York, pp 3–19

Cantini C, Cimato A, Sani G (1999) Morphological evaluation of olive germplasm present in Tuscany region. Euphytica 109(3):173–181

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

Chakraborty R, Kimmel M, Stivers DN, Davison LJ, Deka R (1997) Relative mutation rates at di-, tri-, and tetranucleotide micro-satellite loci. Proc Natl Acad Sci USA 94:1041–1046

Chalak L, Haouane H, Essalouh L, Santoni S, Besnard G, Khadari B (2015) Extent of the genetic diversity in Lebanese olive (*Olea europaea* L.) trees: a mixture of an ancient germplasm with recently introduced varieties. Genet Resour Crop Evolut 62(4):621–633

Cipriani G, Marrazzo MT, Marconi R, Cimato A, Testolin R (2002) Microsatellite markers isolated in olive (*Olea europaea* L.) are suitable for individual fingerprinting and reveal polymorphism within ancient cultivars. Theoret Appl Genet 104(2–3):223–228

Cipriani G, Marrazzo MT, Di Gaspero G, Pfeiffer A, Morgante M, Testolin R (2008) A set of microsatellite markers with long core repeat optimized for grape (*Vitis* spp.) genotyping. BMC Plant Biol 8(1):127

Clark JM (1988) Novel non-templated nucleotide addition-reactions catalysed by procaryotic and eukaryotic DNA-polymerases. Nucleic Acids Res 16:9677–9686

Consolandi C, Palmieri L, Doveri S, Maestri E, Marmiroli N, Reale S et al (2007) Olive variety identification by ligation detection reaction in a universal array format. J Biotechnol 129:565–574. https://doi.org/10.1016/j.jbiotec.2007.01.025

Dakin EE, Avise JC (2004) Microsatellite null alleles in parentage analysis. Heredity 93(5):504–509

De la Rosa R, James CM, Tobutt KR (2002) Isolation and characterization of polymorphic microsatellites in olive (*Olea europaea* L. and their transferability to other genera in the Oleaceae. Mol Ecol Resour 2(3):265–267

De la Rosa R, Angiolillo A, Guerrero C, Pellegrini M, Rallo L, Besnard G, Baldoni L (2003) A first linkage map of olive (*Olea europaea* L.) cultivars using RAPD, AFLP, RFLP and SSR markers. Theoret Appl Genet 106(7):1273–1282

De la Rosa R, James CM, Tobutt KR (2004) Using microsatellites for paternity testing in olive progenies. HortScience 39(2):351–354

De la Rosa R, Belaj A, Muñoz-Mérida A, Trelles O, Ortiz-Martín I, González-Plaza JJ, Beuzón CR (2013)

Development of EST-derived SSR markers with long-core repeat in olive and their use for paternity testing. J Am Soc Hortic Sci 138(4):290–296

De Lorenzis G, Chipashvili R, Failla O, Maghradze D (2015) Study of genetic variability in *Vitis vinifera* L. germplasm by high-throughput Vitis18kSNP array: the case of Georgian genetic resources. BMC Plant Biol 15:154. https://doi.org/10.1186/s12870-015-0510-9

Dettori MT, Micali S, Giovinazzi J, Scalabrin S, Verde I, Cipriani G (2015) Mining microsatellites in the peach genome: development of new long-core SSR markers for genetic analyses in five *Prunus* species. SpringerPlus 4(1):1

Díaz A, Martín A, Rallo P, Barranco D, De la Rosa R (2006) Self-incompatibility of Arbequina and Picual Olive Assessed by SSR Markers. J Am Soc Hortic Sci 131(2):250–255

Díaz A, de la Rosa R, Rallo P, Muñoz-Díez C, Trujillo I, Barranco D, Martin A, Belaj A (2007a) Selections of an olive breeding program identified by microsatellite markers. Crop Sci 47:2317–2322

Díez CM, Trujillo I, Barrio E, Belaj A, Barranco D, Rallo L (2011) Centennial olive trees as a reservoir of genetic diversity. Ann Bot 108(5):797–807

Díez CM, Trujillo I, Martínez-Urdiroz N, Barranco D, Rallo L, Marfil P, Gaut BS (2015) Olive domestication and diversification in the Mediterranean Basin. New Phytol 206(1):436–447

D'Imperio M, Viscosi V, Scarano MT, D'Andrea M, Zullo BA, Pilla F (2011) Integration between molecular and morphological markers for the exploitation of olive germoplasm (*Olea europaea*). Sci Hortic 130(1):229–240

Dominguez-Garcia MC, Laib M, De la Rosa R, Belaj A (2012) Characterisation and identification of olive cultivars from North-eastern Algeria using molecular markers. J Hortic Sci Biotechnol 87(2):95–100

El Bakkali A, Haouane H, Moukhli A, Costes E, Van Damme P, Khadari B (2013) Construction of core collections suitable for association mapping to optimize use of Mediterranean olive (*Olea europaea* L.) genetic resources. PLoS ONE 8(5):e61265

Ellegren H (2004) Microsatellites: simple sequences with complex evolution. Nat Rev Genet 5:435–445

Ellis JR, Burke JM (2007) EST-SSRs as a resource for population genetic analyses. Heredity 99(2):125–132. https://doi.org/10.1038/sj.hdy.6801001

Erre P, Chessa I, Muñoz-Diez C, Belaj A, Rallo L, Trujillo I (2010) Genetic diversity and relationships between wild and cultivated olives (*Olea europaea* L.) in Sardinia as assessed by SSR markers. Genetic Resour Crop Evolut 57(1):41–54

Escribano P, Viruel MA, Hormaza JI (2008) Comparison of different methods to construct a core germplasm collection in woody perennial species with simple sequence repeat markers. A case study in cherimoya (*Annona cherimola*, Annonaceae): an underutilised subtropical fruit tree species. Ann Appl Biol 153:25–32

Esselink GD, Smulders MJM, Vosman B (2003) Identification of cut rose (Rosa hybrida) and rootstock varieties using robust sequence tagged microsatellite site markers. Theoret Appl Genet 106:277–286

Franco J, Crossa J, Taba S, Shands H (2005) A sampling strategy for conserving genetic diversity when forming core subsets. Crop Sci 45:1035–1044

Fendri M, Trujillo I, Trigui A, Rodríguez-García MI, Ramírez JDDA (2010) Simple sequence repeat identification and endocarp characterization of olive tree accessions in a Tunisian germplasm collection. HortScience 45(10):1429–1436

Gil FS, Busconi M, Da Câmara Machado A, Fogher C (2006) Development and characterization of microsatellite loci from Olea europaea. Mol Ecol Note 6(4):1275–1277

Green PS, Wickens GE, Tan K, Mill RR, Elias TS (1989) The *Olea europaea* complex, The Davis and Hedge Festschrift, Edinburgh. Edinburgh University Press, pp 287–299

Grist SA, Firgaira FA, Morley AA (1993) Dinucleotide repeat polymorphisms isolated by the polymerase chain reaction. Biotechniques 15(2):304–309

Guichoux E, Lagache L, Wagner S, Chaumeil P, Léger P, Lepais O, Petit RJ (2011) Current trends in microsatellite genotyping. Mol Ecol Resourc 11(4):591–611

Hakim IR, Kammoun NG, Makhloufi E, Rebaï A (2010) Discovery and potential of SNP markers in characterization of Tunisian olive germplasm. Diversity 2:17–27. https://doi.org/10.3390/d2010017

Haouane H, El Bakkali A, Moukhli A, Tollon C, Santoni S, Oukabli A, Khadari B (2011) Genetic structure and core collection of the World Olive Germplasm Bank of Marrakech: towards the optimised management and use of Mediterranean olive genetic resources. Genetica 139(9):1083–1094

Hill CR, Butler JM, Vallone PM (2009) A 26plex autosomal STR assay to aid human identity testing. J Forensic Sci 54:1008–1015

Ipek M, Seker M, Ipek A, Gul MK (2015) Identification of molecular markers associated with fruit traits in olive and assessment of olive core collection with AFLP markers and fruit traits. Genet Mol Res 14(1):2762–2774

Ipek A, Yılmaz K, Sıkıcı P, Aktepe Tangu N, Tülin Öz A et al (2016) SNP discovery by GBS in olive and the construction of a high-density genetic linkage map. Biochem Genet 54:313–325. https://doi.org/10.1007/s10528-016-9721-5

Kaya HB, Cetin O, Kaya H, Sahin M, Sefer F, Kahraman A et al (2013) SNP discovery by Illumina-based transcriptome sequencing of the olive and the genetic characterization of Turkish olive genotypes revealed by AFLP, SSR and SNP markers. PLoS ONE 8:e73674. https://doi.org/10.1371/journal.pone.0073674

Khadari B, Charafi J, Moukhli A, Ater M (2008) Substantial genetic diversity in cultivated Moroccan olive despite a single major cultivar: a paradoxical situation evidenced by the use of SSR loci. Tree Genetics Genomes 4(2):213–221

Las Casas G, Scollo F, Distefano G, Continella A, Gentile A, La Malfa S (2014) Molecular characterization of olive (*Olea europaea* L.) Sicilian cultivars using SSR markers. Biochem Syst Ecol 57:15–19

Le Cunff L, Fournier-Level A, Laucou V, Vezzulli S, Lacombe T et al (2008) Construction of nested genetic core collections to optimize the exploitation of natural diversity in *Vitis vinifera* L. subsp. Sativa. BMC Plant Biol 8:31

Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequences evolution. Mol Biol Evol 4:203–221

León L, Martín LM, Rallo L (2004) Phenotypic correlations among agronomic traits in olive progenies. J Am Soc Hortic Sci 129(2):271–276

Li YC, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol Ecol 11:2453–2465. https://doi.org/10.1046/j.1365-294X.2002.01643.x

Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21(9):2128–2129

Marrano A, Birolo G, Prazzoli ML, Lorenzi S, Valle G, Grando MS (2017) SNP-discovery by RAD-sequencing in a germplasm collection of wild and cultivated grapevines (V. vinifera L.). PLoS ONE 12:e0170655. https://doi.org/10.1371/journal.pone.0170655

Mason A (2015) SSR genotyping. In: Batley J (ed) Plant genotyping. Springer, New York NY, pp 77–89

Mason AS, Zhang J, Tollenaere R, Vasquez Teuber P, Dalton-Morgan J, Hu L et al (2015) High-throughput genotyping for species identification and diversity assessment in germplasm collections. Mol Ecol Res 15:1091–1101. https://doi.org/10.1111/1755-0998.12379

Mariotti R, Cultrera NGM, Mousavi S, Baglivo F, Rossi M, Albertini E, Baldoni L (2016) Development, evaluation, and validation of new EST-SSR markers in olive (Olea europaea L.). Tree Genetics Genomes 6(12):1–14

Martins-Lopes P, Gomes S, Santos E, Guedes-Pinto H (2008) DNA markers for Portuguese olive oil fingerprinting. J Agric Food Chem 56:11786–11791

McKhann HI, Camilleri C, Berard A, Bataillon T, David JL et al (2004) Nested core collections maximizing genetic diversity in Arabidopsis thaliana. Plant J 38:193–202

Merdinoglu D, Butterlin G, Bevilacqua L, Chiquet V, Adam-Blondon AF, Decroocq S (2005) Development and characterization of a large set of microsatellite markers in grapevine (Vitis vinifera L.) suitable for multiplex PCR. Mol Breed 15(4):349–366

Muleo R, Colao MC, Miano D, Cirilli M, Intrieri MC, Baldoni L et al (2009) Mutation scanning and genotyping by high resolution DNA melting analysis in olive germplasm. Genome 52:252–260. https://doi.org/10.1139/G09-002

Muñoz-Mérida A, González-Plaza JJ, Cañada A, Blanco AM, García-López MDC, Rodríguez JM, Pedrola L, Sicardo MD, Hernández ML, De la Rosa R, Belaj A, Gil-Borja M, Luque F, Martínez-Rivas JM, Pisano DG, Trelles O, Valpuesta V, Beuzón CR (2013) De novo assembly and functional annotation of the olive (Olea europaea) transcriptome. DNA Res 20(1):93–108

Muzzalupo I, Chiappetta A, Benincasa C, Perri E (2010) Intracultivar variability of three major olive cultivars grown in different areas of central-southern Italy and studied using microsatellite markers. Sci Hortic 126(3):324–329

Muzzalupo I, Vendramin GG, Chiappetta A (2014) Genetic biodiversity of Italian olives (Olea europaea) germplasm analyzed by SSR markers. Sci World J. https://doi.org/10.1155/2014/296590

Nishio S, Yamamoto T, Terakami S, Sawamura Y, Takada N, Nishitani C, Saito T (2011) Novel genomic and EST-derived SSR markers in Japanese chestnuts. Sci Hortic 130(4):838–846

Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. Bioinformatics, 28: 2537–2539. Freely available as an open access article from: http://bioinformatics.oxfordjournals.org/content/28/19/2537

Poncet V, Rondeau M, Tranchant C, Cayrel A, Hamon S, De Kochko A, Hamon P (2006) SSR mining in coffee tree EST databases: potential use of EST–SSRs as markers for the Coffea genus. Mol Genet Genomics 276(5):436–449

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1:215–222

Rahemi A, Fatahi R, Ebadi A, Taghavi T, Hassani D, Gradziel T, Chaparro J (2012) Genetic diversity of some wild almonds and related Prunus species revealed by SSR and EST-SSR molecular markers. Plant Syst Evol 298(1):173–192

Rallo P, Dorado G, Martín A (2000) Development of simple sequence repeats (SSRs) in olive tree (Olea europaea L.). Theoret Appl Genet 101(5–6):984–989

Rallo P, Tenzer I, Gessler C, Baldoni L, Dorado G, Martin A (2003) Transferability of olive microsatellite loci across the genus Olea. Theoret Appl Genet 107(5):940–946

Reale S, Doveri S, Díaz A, Angiolillo A, Lucentini L, Pilla F et al (2006) SNP-based markers for discriminating olive (Olea europaea L.) cultivars. Genome 49:1193–1205. https://doi.org/10.1139/g06-068

Riaz S, Dangl GS, Edwards KJ, Meredith CP (2004) A microsatellite marker based framework linkage map of Vitis vinifera L. Theoret Appl Genet 108(5):864–872

Richards CM, Volk GM, Reeves PA, Reilley AA, Henk AD (2009) Selection of stratified core sets representing wild Apple (Malus sieversii). J Am Soc Hortic Sci 134:228–235

Rotondi A, Magli M, Ricciolini C, Baldoni L (2003) Morphological and molecular analyses for the characterization of a group of Italian olive cultivars. Euphytica 132(2):129–137

Rubio de Casas RG, Besnard P, Schönswetter L, Balaguer, Vargas P (2006) Extensive gene flow blurs phylogeographic but not phylogenetic signal in Olea europaea L. Theoret Appl Genet 113(4):575–583

Sarri V, Baldoni L, Porceddu A, Cultrera NGM, Contento A, Frediani M, Cionini PG (2006) Microsatellite markers are powerful tools for discriminating among olive cultivars and assigning them to geographically defined populations. Genome 49(12):1606–1615

Sebastiani L, Busconi M (2017) Recent developments in olive (Olea europaea L.) genetics and genomics: applications in taxonomy, varietal identification, traceability and breeding. Plant Cell Rep 36:1–16

Sefc KM, Lopes MS, Mendonça D, Santos MRD, Machado LM, Machado ADC (2000) Identification of microsatellite loci in olive (Olea europaea) and their characterization in Italian and Iberian olive trees. Mol Ecol 9(8):1171–1173

Silfverberg-Dilworth E, Matasci CL, Van de Weg WE, Van Kaauwen MPW, Walser M, Kodde LP, Patocchi A (2006) Microsatellite markers spanning the apple (Malus x domestica Borkh.) genome. Tree Genetics Genomes 2(4):202–224

Sup Lee J, Hanford MG, Genova JL, Farber R (1999) Relative stabilities of dinucleotide and tetranucleotide repeats in cultured mammalian cells. Hum Mol Genet 8(13):2567–2572

Tessier C, David J, This P, Boursiquot JM, Charrier A (1999) Optimization of the choice of molecular markers for varietal identification in *Vitis vinifera* L. Theoret Appl Genet 98(1):171–177

Thiel T, Michalek W, Varshney R, Graner A (2003) Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theoreti Appl Genet 106(3):411–422

Tóth G, Gáspári Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: survey and analysis. Genome Res 10(7):967–981

Trujillo I, Rallo L, Arús P (1995) Identifying olive cultivars by isozyme analysis. J Am Soc Hortic Sci 120(2):318–324

Trujillo I, Ojeda MA, Urdiroz NM, Potter D, Barranco D, Rallo L, Díez CM (2014) Identification of the Worldwide Olive Germplasm Bank of Córdoba (Spain) using SSR and morphological markers. Tree Genetics Genomes 10(1):141–155

Wang Y, Zhang J, Sun H, Ning N, Yang L (2011) Construction and evaluation of a primary core collection of apricot germplasm in China. Sci Hortic Amster 128:311–319

Wu SB, Collins G, Sedgley M (2004) A molecular linkage map of olive (*Olea europaea* L.) based on RAPD, microsatellite, and SCAR markers. Genome 47(1):26–35

Zhao W, Cho GT, Ma KH, Chung JW, Gwag JG et al (2010) Development of an allele-mining set in rice using a heuristic algorithm and SSR genotype data with least redundancy for the post-genomic era. Mol Breed 26:639–651