

Development, evaluation, and validation of new EST-SSR markers in olive (*Olea europaea* L.)

R. Mariotti¹ · N. G. M. Cultrera¹ · S. Mousavi² · F. Baglivo¹ · M. Rossi¹ · E. Albertini³ · F. Alagna⁴ · F. Carbone⁵ · G. Perrotta⁶ · Luciana Baldoni¹

Received: 20 November 2015 / Revised: 11 August 2016 / Accepted: 7 November 2016 / Published online: 23 November 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract Rapid and effective genotyping is an important goal to discriminate among the numerous olive cultivars and their wild related forms. The largely used di-nucleotide simple sequence repeat (SSR) markers show a high level of polymorphism and have strongly contributed to solve many inconsistencies in varietal identity, but many problems related to difficult discrimination of neighboring alleles and low comparability of data among different labs severely reduce their applicability for large-scale screening. The availability of numerous transcriptome libraries, which were developed from different tissues of several olive varieties, has allowed their

intensive screening to search for polynucleotide microsatellite regions with long core repeats, potentially polymorphic among varieties. An accurate screening of all these polymorphisms has allowed to select a set of 25 trinucleotide and one tetranucleotide SSRs, showing a good level of discrimination power with a high allele pattern resolution and repeatability. They were preliminarily tested on a group of cultivated varieties then validated on a wider group of cultivated and wild plants, and related species and subspecies, demonstrating a good transferability within the entire *Olea* taxon. Furthermore, an in silico functional prediction has allowed to assign each transcribed sequence to their gene functions and biological process categories, highlighting their potential application of these new EST-SSRs as functional markers.

R. Mariotti and N. G. M. Cultrera are co-first authors, and they have contributed equally to this work.

Communicated by G. G. Vendramin

Electronic supplementary material The online version of this article (doi:10.1007/s11295-016-1077-9) contains supplementary material, which is available to authorized users.

✉ Luciana Baldoni
luciana.baldoni@ibbr.cnr.it

¹ CNR-Institute of Biosciences and Bioresources, Perugia, Italy

² CNR-Institute for Agricultural and Forest Systems in the Mediterranean, Perugia, Italy

³ Department of Agricultural, Food and Environmental Science, University of Perugia, Perugia, Italy

⁴ CREA-Research Council for Agriculture and Agricultural Economy, Research Unit for Viticulture and Enology in Southern Italy, Turi, BA, Italy

⁵ CREA-Research Council for Agriculture and Agricultural Economy, Olive Growing and Olive Oil Industry Research Centre, Rende, CS, Italy

⁶ ENEA-Technical Unit for Technologies Development, Trisaia Research Centre, Rotondella, MT, Italy

Keywords *Olea* genus · EST-SSRs · Genotyping · Functional annotation · Transferability

Introduction

Olive (*Olea europaea* L. subsp. *europaea*, var. *europaea*) is one of the most important oil crops, and its cultivation is mainly concentrated along the Mediterranean basin although now is spreading in many other new areas (Baldoni and Belaj 2010). Olive is a diploid species ($2n = 46$), highly heterozygous, and with a medium-big size genome (haploid = 1500 Mb) (Loureiro et al. 2007).

Currently, a growing interest has been directed towards the genetic diversity of olive germplasm, which was surveyed in different geographical regions, also characterized by severe climatic and ecological conditions (Hosseini-Mazinani et al. 2014; Mousavi et al. 2014). In addition, the relationships between cultivated varieties and related taxa, as different wild trees (*O. europaea* subsp. *europaea* var. *sylvestris*) and related

subspecies, have also been deeply analyzed (Besnard and El Bakkali 2014; Díez et al. 2014), not only to understand the phylogenetic relationships within the genus but also to discover important traits of great agronomical and environmental interest (Kaya et al. 2016).

Cultivar identification is a primary concern for olive growers, breeders, and scientists. Up to now, the high genetic variability of cultivated olive and the little morphological differences among varieties, coupled with the use of low-effective markers and the confusion in genotype assignment within olive collections, have often led to conflicting information about varietal identity, hiding the level of variability and misinterpreting relationships among cultivars (Atienza et al. 2013; Díez et al. 2012). Lately, several efforts were made in order to solve this issue, by applying molecular genotyping to the accessions held by main olive collections and developing more effective markers (Belaj et al. 2012; Biton et al. 2015; Haouane et al. 2011; Trujillo et al. 2014).

Fingerprinting based on microsatellite markers represents a powerful and reliable tool for characterizing plant varieties, which are exploitable for several molecular investigations such as cultivar identification, paternity test, kinship, or population structure analyses (Chankaew et al. 2014; Cipriani et al. 2010; Ellegren 2004; Shen et al. 2014; Zhang et al. 2015). According to their location on the genome, microsatellite regions may be distinguished into genomic neutral simple sequence repeat (SSR) markers and transcript-tagged SSRs, located on expressed sequence tags (ESTs), with a potential functional value (Bradbury et al. 2013; Hinchliffe et al. 2011). SSR markers show significant advantages, such as reproducibility, locus specificity and low quantity of template required; nevertheless, several disadvantages have also been recorded, which are mainly due to stutter products, allele binning and allele miscalling, as a consequence of the wide use of di-nucleotide microsatellites (Cabezas et al. 2011; Kaur et al. 2015; Targońska et al. 2015; This et al. 2004). The majority of discrepancies among laboratories in scoring di-nucleotide microsatellites are due to the binning process in which raw allele lengths are converted into allele classes whose size is then expressed by an integer (Baldoni et al. 2009; Weeks et al. 2002). These problems may be solved partly by discarding di-nucleotide SSRs in favor of microsatellites with longer motifs (Amos et al. 2007), as made in human fingerprinting (Butler 2007) and for fruit crop genotyping (Cipriani et al. 2008; Dai et al. 2015).

The majority of SSRs currently identified for olive genotyping carry di-nucleotide repeats (Carriero et al. 2002; Cipriani et al. 2002; de la Rosa et al. 2002; Diaz et al. 2006; Rallo et al. 2000; Sabino Gil et al. 2006; Sefc et al. 2000), mostly AG/CT repeats. In order to establish a common set of markers able to produce reliable data for olive cultivar discrimination, Baldoni et al. (2009) provided a consensus list of di-nucleotide SSR markers. But, despite the use of highly

robust fingerprinting protocols, genotyping errors may still occur (Atienza et al. 2013; Díez et al. 2012; This et al. 2004) and the standardization of protocols and the exchange of information concerning the genetic profile of reference varieties are still required.

Frequency of microsatellites with long core repeats is lower than those carrying shorter motifs (Fungtammasan et al. 2015), thus their identification procedure requires additional efforts. Previous studies highlighted that in plants, microsatellite regions occur more frequently in transcribed regions than in genomic DNA (Morgante et al. 2002). A high frequency of trinucleotide repeats is reported in exons that contain almost no tetranucleotide repeats (Toth et al. 2000; Varshney et al. 2005), whereas mutability tests showed that the allelic variability of exon SSRs is lower than that of intronic repeats (Li et al. 2011). Massive EST and genomic sequence data have made available information on repeat region abundance and position, and bioinformatic analyses have significantly accelerated the process of identification and selection, reducing the time for their application (Acuna et al. 2012; Li et al. 2012; Shiferaw et al. 2012; Yang et al. 2012; Duran et al. 2013; Cubry et al. 2014).

SSRs located on gene exons may potentially control important agronomic traits (Zeng et al. 2010; Zhang et al. 2012; Boccacci et al. 2015; Dutta et al. 2011). Furthermore, since EST-SSR markers stand on expressed sequences, they show a high transferability across taxa (Scott et al. 2000; Zhang et al. 2005; Aggarwal et al. 2007; Luro et al. 2008), representing a valuable resource for comparative genomics, biodiversity and evolutionary studies (De Keyser et al. 2009).

SSR markers with repeat motifs longer than a di-nucleotide have been developed from genome sequences in several crop species, such as grape, cranberry, citrus, celery and common bean (Biswas et al. 2012; Chen et al. 2014; Cipriani et al. 2008; Fu et al. 2013; Zhu et al. 2012), from gene sequences, as in sesame (Zhang et al. 2012), or from EST data, as made for sunflower, castor bean, grape, tea, pea and alfalfa (Heesacker et al. 2008; Huang et al. 2011; Qiu et al. 2010; Wang et al. 2013; Yao et al. 2012; Xu et al. 2012). In fact, EST-derived SSRs have been well documented in several plant species (Lima et al. 2008; Feng et al. 2009; Du et al. 2013; Yao et al. 2012; Xu et al. 2012; Zhou et al. 2014; Ferrao et al. 2015). Lately, also, in olive, some EST-SSRs have been identified and used for paternity testing and mapping purposes (Essalouh et al. 2014; de la Rosa et al. 2013; Khadari et al. 2014).

The present study has taken its cue by the availability of several EST collections of *O. europaea*, in order to address the following issues: i) identifying new polynucleotidic repeats through the screening of these sequences, ii) testing the polymorphism and applicability of these regions as new markers for cultivated olive fingerprinting, and iii) validating their transferability on a wide set of related *Olea* taxa.

Materials and methods

Plant material and DNA extraction

A set of 32 olive (*O. europaea* subsp. *europaea* var. *europaea*) cultivars was primarily selected to test amplificability and applicability of the EST repeats as potential markers (Table S1). Then, to assay their ability to keep the widest range of variability within cultivated olives, other 47 varieties were added to the first group, considering their distribution and genetic variability (Sarri et al. 2006; Baldoni et al. 2009). Moreover, in order to test transferability of markers to other olive-related forms, four samples of wild olives (*O. europaea* subsp. *europaea* var. *sylvestris*), two *O. europaea* subsp. *cuspidata*, two subsp. *laperrinei*, samples of polyploid subspecies *cerasiformis* and *maroccana* (Besnard et al. 2008), and one *Olea paniculata* genotype were also included in the study, for a total of 90 samples (Table S1). DNA samples of this plant material were derived from the CNR-IBBR Olive Collection (Perugia, Italy) and from the World Olive Germplasm Bank (WOGB), IFAPA (Cordoba, Spain). Two clones of cv. *frantoio*, collected from two different growing sites and previously verified as a unique genotype by di-nucleotide well-characterized SSRs, were included in the analysis in order to test the reliability of new markers.

Identification of repeated polymorphic motifs in the EST collections

In order to identify the most effective SSR markers from the EST sequences, a specific pipeline has been established, as reported in Fig. 1.

SSR motifs were searched within the EST collections from flower and fruit tissues of different olive varieties (Alagna et al. 2009; Alagna et al. 2016; Corrado et al. 2012). The adaptor-trimmed 454 read sequence data were assembled by using the GS De Novo Assembler software (Roche Diagnostics Corporation, Basel, Switzerland). Repeated motifs were searched using the Perl script program MISA (<http://pgrc.ipk-gatersleben.de/misa>) within each EST collection, by applying the following parameters: mononucleotide repeats (MNRs) ≥ 10 , di-nucleotide (DNR) ≥ 6 , from trinucleotide to hexanucleotide (TNR, TTNR, PNR, and HNR) ≥ 5 , and distance between two SSRs ≤ 100 bp. Only polymorphic repeated motifs between varieties were selected for further analyses.

Selection of most valuable polynucleotide repeats as potential markers and functional annotation of corresponding putative genes

Transcripts containing polynucleotide repeats were aligned by BLASTN to the olive genome scaffolds of the cv. *leccino*

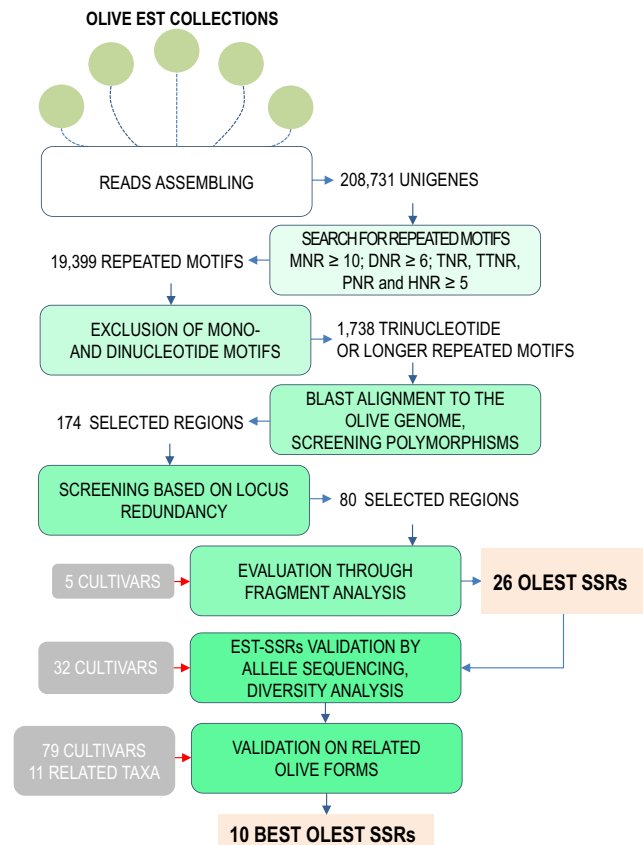


Fig. 1 Pipeline used for the selection of new EST-SSR markers

made available to the OLEA project partners (oleagenome.org) in order to identify the genomic regions containing the target SSRs. A first screening was carried out on the selected repeats in order to discard regions characterized by multiple calls, showing flanking regions inadequate for primer construction or any other serious drawback.

A second screening was performed on the remaining 80 SSRs: primers were designed by using Primer3 v. 0.4.0 (<http://frodo.wi.mit.edu>), with a GC content higher than 20% and with a common melting temperature (T_m) of 60 °C so as to obtain fragments with an expected length ranging from 100 to 350 bp, approximately, in order to facilitate their discrimination when multiplexing. Analyses were carried out on a subset of five cultivars (*leccino*, *chemlali*, *izmir sofralik*, *koroneiki*, *picual*), in order to discard non-amplifiable loci and markers showing multi-band PCR products or unexpected amplicon lengths. Polymerase chain reactions were performed in a volume of 25 μ l containing 25 ng of DNA, 10 \times PCR buffer, 200 μ M of each dNTP, 10 pmol of primers forward (with a 18 bp tail at the 5' end) and reverse, and 2 U of PerfectTaq DNA Polymerase (Q5 High-Fidelity DNA Polymerase, New England Biolabs). Fluorescent tail (10 pmol) was annealed with the forward primer using a double-step PCR: the first step consisting of an initial denaturation at 95 °C for 5 min, followed by 35 cycles of 95 °C for 30 s, 60 °C

for 30 s, and 72 °C for 25 s, followed by a final elongation at 72 °C for 30 min; the second one (for tail annealing) made up by 17 cycles, with the same conditions of the first step except for annealing temperature ($T_m = 52$ °C). Negative (no template DNA) and positive (*cv. leccino* DNA) controls were included in each amplification run, in order to detect non-specific products and verify the success of PCR reaction, respectively. All amplifications were performed with the PCR System 9600 (Applied Biosystems, Foster City, CA). PCR products were initially electrophoresed on 2% agarose gels in order to check the amplicons and then loaded on an ABI 3130 genetic analyzer (Applied Biosystems-Hitachi, Foster City, CA), by using the internal GeneScan™ 500 LIZ Size Standard (Applied Biosystems). All amplifications and runs were replicated two times in order to test their repeatability. Output data were analyzed by GeneMapper 3.7 (Applied Biosystems-Hitachi). Non-amplifiable loci, multi-band PCR products (tri or more alleles or smearing when visualized on agarose gel), or amplicons, with unexpected lengths, were discarded.

The final selection of best markers was performed on a set of 32 varieties (Table S1) on 26 selected loci, using the aforementioned amplification conditions and applying the following criteria: (1) signal level, which was ranked as strong, medium, and weak; (2) stuttering level, which was scaled as low (or no stuttering), medium, and high; and (3) number of amplified alleles detected through fragment analysis.

All alleles at each locus were sequenced, in order to verify repeat motif features and base composition and confirm fragment length. Homozygous fragments were directly sequenced through the BigDye Terminator technique (Applied Biosystems). Heterozygous alleles showing different lengths were cloned into *Escherichia coli* XL1 Blue strain by using pGEM-T Easy Vector System I (Promega). All amplifications and cloning products were sequenced on an ABI 310 Genetic Analyzer (Applied Biosystems-Hitachi, Foster City, CA).

To verify the transferability of these loci to other forms related to cultivated olive, only 10 EST-SSRs, showing the best diversity scores, were further applied to the wider sample set of 90 genotypes.

In order to compare the discrimination power of the new developed SSR markers with di-nucleotide SSRs, the same diversity indices were applied on best-ranked di-nucleotide loci for the same set of 79 cultivars (Baldoni et al. 2009; Mariotti unpublished data).

To predict the entire open reading frame (ORF) of each original transcript, genomic scaffolds containing the target SSR regions were analyzed by Softberry FGESH (<http://linux1.softberry.com>) and were then functionally annotated by using the Blast2GO software (Conesa et al. 2005, <http://www.blast2go.com>). The ExpASy translate tool (Gasteiger et al. 2003) was used to predict the protein sequence, starting from the genomic scaffold containing the

validated 26 EST-SSRs. The coding regions were aligned in the NCBI database (BLASTX, non-redundant protein sequences), and those that did not show significant similarity, versus known proteins, were aligned to the nucleotide collection using BLASTN (Standard Nucleotide BLAST). The BLASTX hits were searched on Gene Ontology (GO) terms.

Data analysis

GenAlEx 6.501 (Peakall and Smouse 2012) was used to calculate the number of alleles (N_a); number of effective alleles (N_e); Shannon's information index (I); observed, expected, and unbiased expected heterozygosity (H_o , H_e , and uH_e , respectively); fixation index (F); and genetic differentiation among populations (F_{st}). To evaluate the ability of new SSRs to assess molecular diversity and their potential use in fingerprinting analyses, the polymorphism information content (PIC) was computed at each locus by using Cervus 3.0.3 (Kalinowski et al. 2007), while FreeNa was applied to estimate the presence of null alleles (Dempster et al. 1977). Hardy-Weinberg equilibrium was estimated with GENEPOP 4.2 (Raymond and Rousset 1995; Rousset 2008) through chi-squared tests. Two-dimensional principal component analysis (PCA) was carried out with MultiVariate Statistical Package (MVSP) version 3.22 (Kovach Computing Services, Anglesey, Wales, UK), starting from a square matrix obtained by GenAlEx. Neighbor-joining (NJ) dendrogram was calculated with the Darwin software version 6 (Perrier and Jacquemoud-Collet 2006) using 10,000 bootstrap replications and 50% cutoff and visualized with FigTree v.1.4.2 (<http://tree.bio.ed.ac.uk/software/figtree/>).

Results

Identification of repeated polymorphic motifs in EST collections

From the EST collections, all kinds of repeats from one to n nucleotides have been considered, just to get an idea on the frequency of each type of repeat along the transcript sequences. Fruit EST transcriptomes, represented by a total of 102,133 unigenes, contained 6637 repeats (1 repeated sequence every 3.45 Kbps), while flower EST collections were composed by 106,598 unigenes, with 12,762 total repeats (1 repeat every 3.41 Kbps) that were detected (Table 1). In both cases, a high percentage of mononucleotide motifs was observed. From a total of 19,399 identified repeats, 1738 resulted trinucleotidic or with longer repeated motifs. Only simple sequence repeats with 3 to 6 nucleotide motifs were selected, using the minimal length of SSR repeats ($3 \times 6 = 18$ bp for trinucleotides, $4 \times 5 = 20$ bp for tetranucleotides, $5 \times 4 = 20$ bp for pentanucleotides, and $6 \times 4 = 24$ bp for hexanucleotides).

Table 1 Total repeats derived from fruit and flower EST collections

Motif	Number of repeats												Total
	5	6	7	8	9	10	11	12	13	14	15	>15	
MNR	NA	NA	NA	NA	NA	1518	1321	1271	1396	1470	1440	7024	15,440
DNR	NA	703	407	294	210	161	107	117	59	46	20	97	2221
TNR	852	306	131	60	26	21	20	4	2	8	4	16	1450
TTNR	73	18	3	3	2	–	–	–	–	–	–	–	99
PNR	28	11	4	1	–	–	–	–	–	–	–	–	44
HNR	107	26	5	5	1	1	–	–	–	–	–	–	145
Total													19,399

MNR mononucleotide motif, DNR di-nucleotide, TNR trinucleotide, TTNR tetranucleotide, PNR pentanucleotide, HNR hexanucleotide, NA not available data

Considering the aim of the present work, which intended to validate EST-SSRs with motifs longer than two bases, mononucleotide and di-nucleotide repeats were not examined.

EST-SSRs screening

From the first marker screening, which was carried out through MISA software considering polymorphisms among cultivars, 174 potential EST-SSRs were identified (Fig. 1). Selected SSR regions detected in different EST collections, but related to the same locus, were discarded in order to avoid redundancy. Eighty loci were preselected for primer design and further laboratory analyses, allowing to discard those showing amplification drawbacks.

Technical details of new EST-SSR markers

The 26 EST-SSRs finally selected as best markers were named as Olive EST (OLEST) SSRs (Fig. 1) and kept as the core of new markers. Their primer sequences, repeat patterns, and allele sizes are listed in Table 2. Accession numbers of submitted sequences (one allele for each locus), predicted genes including the microsatellite region, and corresponding biological processes are shown in Table 3. Among the selected EST-SSRs, twenty-five contained trinucleotide repetitions and only one (OLEST11) had a tetramer motif. The most common microsatellite motif was GAA (15.4%) followed by TGG (11.5%).

Sequencing analyses of the alleles detected at all 26 new EST-SSR markers, when amplified on 32 olive cultivars, confirmed the amplicon lengths obtained from fragment analysis and highlighted other sequence polymorphisms than those expected. In OLEST1, the CTT motif was interrupted by an ATT; in OLEST7, OLEST14, OLEST22, and OLEST26, some deletions of 10 to 25 bp were observed; OLEST17 included two 3 bp indels; OLEST10 and OLEST24 presented alleles with 3 bp indels (TGA and TTT, respectively). In

OLEST26, the allele 306 was distinguished by a 10 bp insertion.

All selected loci resulted polymorphic in the 32 analyzed cultivated varieties (Tables 3 and S2). Na and Ne per locus were on average 4.65 and 2.81, respectively, with a minimum Na of 2 for OLEST6 and OLEST11 to a maximum of 10 for OLEST16. Mean values of I, Ho, He, uHe, and F were 1.07, 0.53, 0.56, 0.57, and 0.01, respectively. Ho and He values resulted very similar at all loci, except for OLEST5, for which the Ho value was considerably lower than He. Frequency of null alleles revealed high values only for OLEST5 and OLEST12 while resulted negligible or moderate for the other loci. Chi-squared test, which was applied to detect deviations from Hardy-Weinberg equilibrium, highlighted significant values for seven out of 26 markers, all characterized by a heterozygote deficit. PIC values ranged from 0.18 (OLEST25) to 0.82 (OLEST14), with an average of 0.51.

Transferability of OLEST SSR markers within the *Olea* genus

When based on their allele number and discrimination power, ten EST-SSR loci (OLEST1, OLEST7, OLEST9, OLEST12, OLEST14, OLEST15, OLEST16, OLEST20, OLEST22, and OLEST23) were applied to a larger number of cultivars (79) and to other related forms, such as wild plants, related subspecies, and the *O. paniculata* species, for a total of 90 genotypes (Table S1); new alleles were detected at all observed loci (excepting OLEST23) (Table S3), due to the new variation captured in the larger set of varieties and that specific to related taxa. The number of alleles ranged from two (OLEST6) to ten (OLEST18), with an average of 5.23. The maximum gap between the longest and shortest alleles was 57 bp for OLEST1 to 16 bp for OLEST15. In most cases, the alleles private to the related taxa resulted shorter (OLEST1, OLEST12, and OLEST15) or longer (OLEST9, OLEST14, OLEST20, and OLEST22) than those observed within the cultivated samples.

Table 2 Primer list, repeat patterns, and allele size for the 26 EST-SSR-developed markers

Marker	Forward	Reverse	(SSR motif) _n	Allele size (bp)
OLEST1	TGCTTTGCTTCTCCTAATACTG	CCAATTCCCTCTCCGGTTAACTATA	(CTT) ₃₋₆ ATT(CTT) ₃₋₇	[200]-(218-224)-227-230-233-236-239-242-245-248-(251)-[257]
OLEST2	AGTACTACTACGGTGGAAAFAGGG	TGGAGGTAGAGGGAGGGTIT	(TGG) ₂₋₄	198-201-204
OLEST3	TACAGCTTGTGGTTTGCAG	ACGGCTGCTTCTTTTCTTGA	(CAA) ₅₋₉	255-258-267
OLEST4	TTGAAGAGAGAATGGGAGCG	TGCTCCCTTTATCGTCTGA	(TGA) ₅₋₁₁	233-236-239-242-245-251
OLEST5	CAAGGAGAGGAGAACCCITAAAT	CCACCTAAGCCATTACCAATTC	(TGG) ₄₋₈	325-328-331-334-337
OLEST6	TGGAGATTACAGAGCTGGAGAAG	GTGATAACTGGTGAATTCACAAA	(ATC) ₅₋₇	336-342
OLEST7	CAGGTGGAGGATGATCGTA	TGAGTATCCAGAAAGCGAAAAG	(TGG) ₅₋₁₀ .ACTTCTGTCTTGTGTGTTG GGTTA/-	(242)-260-263-266-269-(272)-275-(278-281)
OLEST8	CTGTACTGAATGGGCACAA	GGAGCAGAGAGAAGGGTGACTG	(CAG) ₄₋₇	331-337-340
OLEST9	GACCCCTCAACTTTGGTGTGG	ATTCACCCGCAACCATCACTC	(GTG) ₃₋₇	(343)-346-349-352-355-358-[361-370]
OLEST10	GAGCCAAATCAAGATTATCAGC	TTAGGGATTCCAAAGCTCGTC	(AGA) ₅₋₉ .TGA/-	230-233-245
OLEST11	GTGCAGGAGCATAACAAGCA	ACATAAGATCTTACACATAGCCATT	(AATT) ₄₋₅	253-257
OLEST12	TCTCAAATTCACATACCAACTCC	CTCTGTGAAAATTCACCCAGATA	(TCT) ₈₋₁₇	[144-159]-162-165-168-174-177-(180)-183-(186)-189
OLEST13	CTTGTACAGACTTGCAGAAITAAAGG	GAACCTCAATAGCCGCCCTTATCA	(GAA) ₆₋₉	245-248-254
OLEST14	TTTTTCGAACGAAITGGAAGG	CTCCATAAAGGAAATGGACGAA	(TTC) ₄₋₁₂ .AGATCACTATTTCATCT/-	272-(275)-[278]-284-287-290-296-298-(302)-[305]-308-311-[314]-(317)
OLEST15	GGAGTTCCTGGTGAACAGA	CGGCAGGAGAATCCACTTT	(GGA) ₃₋₅	[161]-171-174-177
OLEST16	GAATATGCACGGCTTGTCT	TTCTCTCCATTTCCACCACCG	(GAA) ₅₋₂₀	161-164-167-170-173-176-179-182-185-(188)-206-(209)
OLEST17	TTCATCATTTGTTTCACATCCAA	GGTGTATTTGGGACATAATCTCCA	(CGA) ₃₋₁₀ .GAT/-CAA/-	194-203-206-209-215-218
OLEST18	CCTGCAATTTAATTTCTGAGTTGA	TCCATCTCAGGATTCCTTC	(TCT) ₆₋₉	326-329-332-335
OLEST19	GCCTCACATCATGCTCTTCAG	AACACTTCTGACATTCCTCCAGA	(AGC) ₃₋₆	213-216-219-222
OLEST20	CCAGTTATTGCTTGCCTTGCCAAAT	GAAAGGCATTTCTTGCAATCAG	(GCT) ₄₋₈	229-232-235-238-241-(244)-[250-256-259]
OLEST21	GAAAGCCTGGCAGAAAAGC	GCCCTTCTTCTCAGATTTCTCA	(GAA) ₅₋₇	335-338-341
OLEST22	GAGGGAGCAACAGCAACATC	GCGCTAGCAGTCCCGAAT	(AGC) ₅₋₉ .CAACAGGCACAA/- -261-270-273]	231-(237)-243-246-249-252-255-258- -261-270-273]
OLEST23	AGAAGGAAACTTTTGGCAGAGG	TTGACAAGGGAGGGCTGATTC	(GAA) ₄₋₁₀	204-207-210-213-216-219-222
OLEST24	TAGGGTTGTCGAAATGTCAG	TGGACATTTCAAATATAICATACCA	(ATT) ₄₋₇ .TTT/-	199-208-211
OLEST25	TGCTGTGAATGGTATCAGTTT	CCATGGCAGGTGGTAGAATAA	(GGC) ₃ .T ₄₋₈	242-245
OLEST26	CCTAAAATCCTTGTGCCACTG	AAAAAATAGCARGGAAGATTCGA	(CTT) ₅ .CTTTTCTTTT/-T ₄₋₅	295-298-306

In bold are those loci tested on 90 samples. Allele lengths were detected on 32 cultivars; in the parentheses are those on 79 cultivars; in the square brackets are those specific to wild olives, subspecies, or related species; underlined alleles refer to those characterized by indels. In bold are the alleles with sequences published in GenBank

Table 3 Diversity indices for each locus resulting from the analysis of 32 cultivars

Locus	Na	Ne	I	Ho	He	uHe	F	r	HW	PIC
OLEST1	8	5.361	1.844	0.719	0.813	0.826	0.116	0.065	ns	0.790
OLEST2	3	1.247	0.390	0.219	0.198	0.201	-0.106	0	ns	0.183
OLEST3	3	1.380	0.539	0.281	0.275	0.280	-0.021	0	ns	0.257
OLEST4	6	2.616	1.211	0.719	0.618	0.627	-0.164	0	ns	0.566
OLEST5	5	2.616	1.170	0.250	0.618	0.627	0.595	0.227	***	0.559
OLEST6	2	1.640	0.579	0.531	0.390	0.396	-0.362	0	ns	0.314
OLEST7	5	3.266	1.330	0.563	0.694	0.705	0.189	0.067	*	0.640
OLEST8	3	1.680	0.675	0.531	0.405	0.411	-0.312	0	ns	0.344
OLEST9	5	2.619	1.251	0.594	0.618	0.628	0.039	0.005	ns	0.585
OLEST10	3	2.024	0.786	0.375	0.506	0.514	0.259	0.112	*	0.406
OLEST11	2	1.909	0.669	0.594	0.476	0.484	-0.247	0	ns	0.363
OLEST12	7	4.433	1.634	0.406	0.774	0.787	0.475	0.207	***	0.743
OLEST13	3	2.256	0.913	0.563	0.557	0.565	-0.011	0	ns	0.474
OLEST14	8	6.206	1.883	0.625	0.839	0.852	0.255	0.120	**	0.818
OLEST15	3	2.930	1.086	0.688	0.659	0.669	-0.044	0	ns	0.584
OLEST16	10	5.885	1.923	0.750	0.830	0.843	0.096	0.041	ns	0.808
OLEST17	6	2.296	1.089	0.625	0.564	0.573	-0.107	0	ns	0.515
OLEST18	4	1.870	0.863	0.281	0.465	0.473	0.396	0.138	***	0.425
OLEST19	4	2.195	1.039	0.594	0.544	0.553	-0.091	0	ns	0.506
OLEST20	5	2.889	1.203	0.750	0.654	0.664	-0.147	0	ns	0.595
OLEST21	3	1.824	0.779	0.563	0.452	0.459	-0.245	0	ns	0.398
OLEST22	7	4.623	1.669	0.813	0.784	0.796	-0.037	0	ns	0.752
OLEST23	7	4.462	1.641	0.781	0.776	0.788	-0.007	0	ns	0.743
OLEST24	3	1.577	0.602	0.469	0.366	0.372	-0.282	0	ns	0.309
OLEST25	2	1.242	0.345	0.219	0.195	0.198	-0.123	0	ns	0.176
OLEST26	3	2.042	0.859	0.375	0.510	0.518	0.265	0.089	*	0.444
Mean	4.65	2.811	1.076	0.534	0.561	0.570	0.015	-	-	0.511
	(±0.429)	(±0.284)	(±0.092)	(±0.036)	(±0.037)	(±0.037)	(±0.048)	-	-	0.038

In the last row are mean values and standard errors (in the parentheses)

Na number of alleles, Ne number of effective alleles, I Shannon's information index, Ho observed heterozygosity, He expected heterozygosity, uHe unbiased expected heterozygosity, F fixation index, r null allele frequency, HW chi-squared tests for Hardy-Weinberg equilibrium, PIC polymorphism information content

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Most of EST-SSR markers were amplified correctly, except for OLEST16 and OLEST17. In particular, OLEST16 did not amplify in the *O. europaea* subspecies (except for subsp. *cerasiformis*) and in two out of four wild genotypes, while OLEST17 was detectable in the subsp. *cuspidata* and wild samples. The *Olea* subsp. *maroccana* (known as hexaploid) showed problems during the amplification for five out of 26 SSRs (OLEST10, OLEST14, OLEST16, OLEST23, and OLEST24). For each EST-SSR, a maximum of two alleles was detected in all accessions, including the subspecies known as polyploids (*maroccana* and *cerasiformis*).

The information index increased in most cases excepting for OLEST15, OLEST22, and OLEST23. Values of Ho and He remained quite similar to those found in the restricted set of varieties, with Ho generally lower than expected, especially

for OLEST9 and OLEST12, whereas it was higher for OLEST22. On the contrary, fixation index values varied considerably, decreasing for OLEST7 and increasing for OLEST9 and OLEST15. Most of the selected markers showed zero or low null allele frequencies, four loci showed moderate values, and none showed high frequencies (higher than 0.2), as indicated by Chapuis and Estoup (2007). GENEPOP highlighted a deviation from Hardy-Weinberg equilibrium (due to heterozygosity deficit) for six out of ten loci and PIC values remained high, comparable to what was previously obtained for 32 cultivars. When verifying the ability of new SSRs to differentiate between the group of cultivars and the other related taxa considered as a separate population, an Fst value of 0.064 was obtained, indicating a moderate genetic differentiation.

By analyzing the entire data set, PCA described about 37% of variance (21.58 and 15.02% for first and second axes, respectively), revealing that wild olives, subspecies, and the *O. paniculata* samples were grouped apart from all varieties and, in particular, wild plants appeared separated, which were up to the extreme part of the graph, far from all other genotypes (Fig. 2).

Moreover, the ten EST-SSRs were also able to fully distinguish the 79 cultivars, unless for the two clones of *cv. frantoio*, as expected. Neighbor-joining analysis grouped them into three main clusters (Fig. 3). Cultivars *oueslati*, *elmacik*, and *zalmati*, which in the multivariate analysis were positioned close to wilds and subspecies, in neighbor joining were clustered apart from the other cultivars. Finally, it has been observed that three EST-SSR markers, which were characterized by the highest allele number and PIC values (OLEST1, OLEST14, and OLEST16), well distinguished all the 79 cultivars.

When, for the same cultivar sample set, the ten best dinucleotide SSR marker data were considered, they showed a higher mean number of alleles and slight higher values for I, Ho, He, uHe, and PIC, with a lower frequency of null alleles (Table S4).

Functional annotation

BLAST search in GenBank, performed for the 26 OLEST SSRs, revealed high identities for many EST-SSRs. Considering the low level of information available for these EST-SSRs, these markers have been also annotated. From the 26 unigenes, 18 showed similarity to known functional proteins and eight resulted unknown (Table 4). In particular, the predicted gene related to OLEST21 showed 73% of identity with *OesDHN* (dehydrin gene isolated and characterized from

an oleaster, Chiappetta et al. 2015) and in silico annotation showed an insertion of 107 bp. Moreover, OLEST11 and OLEST13 were related to cysteine-type endopeptidase inhibitor activity and defense responses, respectively; OLEST18 is involved in the biological process of proteolysis, and OLEST25 works as signaling pathway related to ethylene activation.

Discussion

In the last decades, the development of more effective molecular markers has been promoted for genotyping purposes in several fruit crop species (Guo et al. 2014; Jiao et al. 2014; Sun et al. 2015), and especially for the olive (Biton et al. 2015; Dominguez-Garcia et al. 2012; Kaya et al. 2013; Torkzaban et al. 2015). Despite the availability of a large set of molecular tools, olive fingerprinting still remains a difficult task, mainly due to the several weaknesses of most widely used markers (Bracci et al. 2011) that include the lack of sequence information, the difficulties in distinguishing among alleles, and the impossibility to share data among labs.

The development of highly reliable markers, which are potentially linked to traits of interest, may improve theoretical and applied research on genotyping, mapping, and marker-assisted breeding (Cubry et al. 2014; El-Rodeny et al. 2014; Liu et al. 2014; Kalia et al. 2011; Ramchiary et al. 2011; Shirasawa et al. 2011; Zhang et al. 2014; Zhao et al. 2012).

In order to identify new and effective SSR markers for olive, a deep screening of fruit and flower EST collections has been performed. In silico analysis and laboratory procedures made it possible to validate 26 promising OLEST SSRs, discarding markers characterized by inadequate flanking regions, multiple calls, and uncertain amplifications. Twenty-

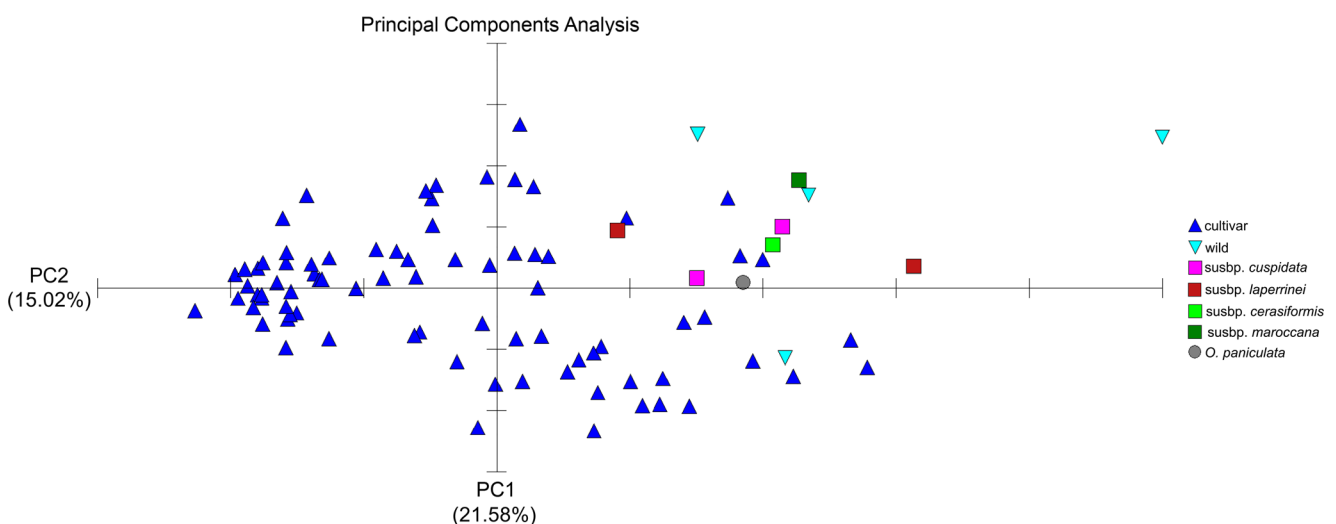


Fig. 2 Principal component analysis (PCA) carried out by using the whole sample set, analyzed through ten EST-SSRs, and amplified in all genotypes, by using MVSP version 3.22 (Kovach Computing Services,

Anglesey, Wales, UK). *Coordinates 1 and 2* represent the 21.58 and 15.02% of the total variance, respectively

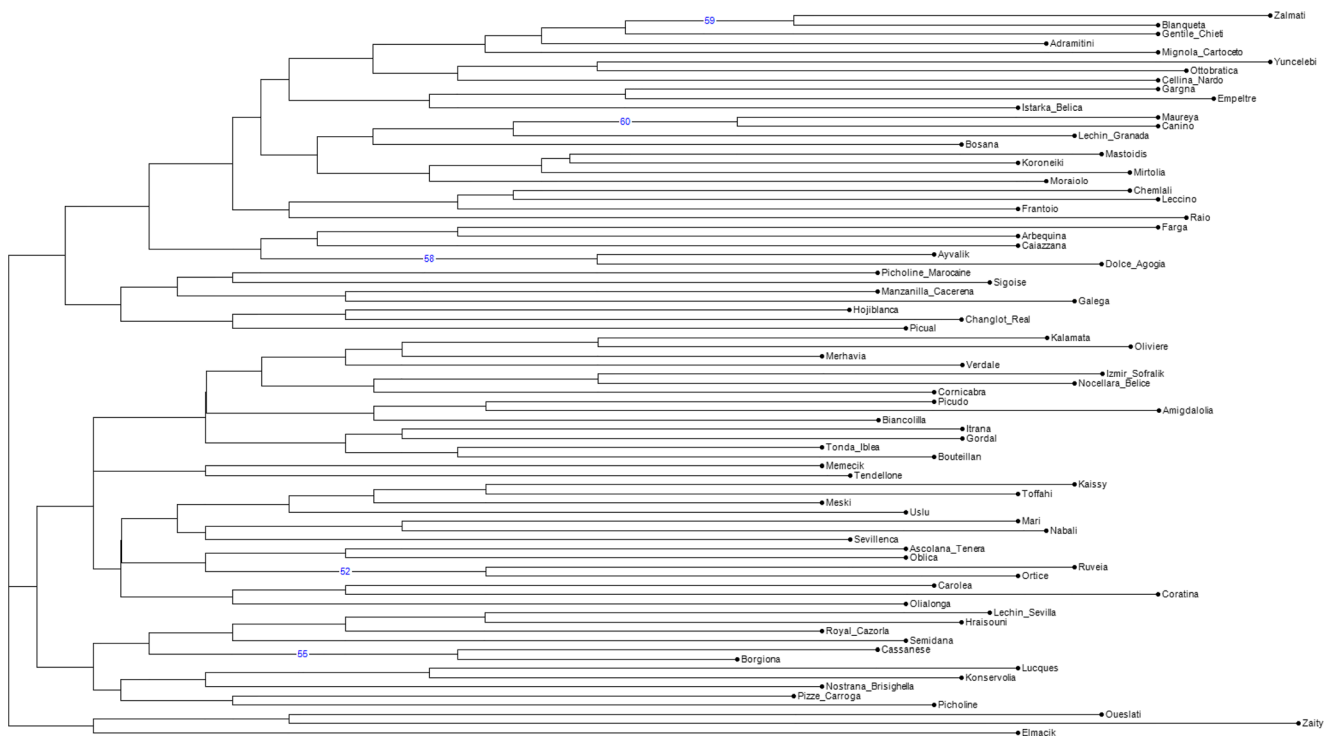


Fig. 3 Relationships among the 79 olive cultivars assessed by constructing a neighbor-joining dendrogram and performed with the Darwin software, applying 1000 replicates and bootstrap cutoff of 50%. Starting data set was represented by the ten best EST-SSRs

five of the selected EST-SSRs showed trinucleotide repeated motifs and only one was a tetranucleotide, likely reflecting the distribution of repeat regions among transcribed and repeated DNAs in many other plants (Morgante et al. 2002; Varshney et al. 2005). The absence of longer repeated patterns within the selected set probably depends on their low frequency (288 long repeats found in the EST collections versus 1450 trinucleotidic patterns) and their low polymorphism level, as already observed in other plant species (Bocacci et al. 2015; Long et al. 2015; You et al. 2015).

Tandem repeats occurring in coding regions can result in the variation of the polynucleotide sequence, thereby causing changes in protein products. In fact, many human diseases have been reported to be associated with trinucleotide repeats (Duran et al. 2013; King 2012). No clear information is yet available about how EST-SSRs in plant genomes can change the function of genes or their expression rate (Asadi and Monfared 2014).

Variations in EST allele sequences are more strictly related to protein function since they are responsible for important changes, leading to the shift in reading frames or unexpected stop codons or different protein lengths and structures, as detected in other species (Qi et al. 2010; Emebiri 2010). The presence of variations in the coding regions increase the functional value of these new EST markers, which are mainly involved in important biological processes, as defense and response to abiotic and biotic stresses. For example, the sequence similarity detected

between the predicted dehydrin from the OLEST21 locus and the *OesDHN* dehydrin gene (Chiappetta et al. 2015), involved in drought tolerance, confirms the potentiality of the OLEST SSRs as functional markers. Further studies will contribute to clarify their potential role as markers linked to traits of main interest.

Their potential capacity to keep the olive variability has been tested on a restricted set of olive cultivars, which is a representative of the wide variability of cultivated olive (Belaj et al. 2012; El Bakkali et al. 2013), then extended to a larger set of cultivars and other related forms, in order to test their capacity to capture molecular variation at a wider scale. Their mean values of polymorphism information content fit in an intermediate class (Xie et al. 2010), highlighting promising levels of variability, as also confirmed by their observed heterozygosity.

Based on main genetic diversity indices (Na, Ho, and PIC), ten OLEST markers have been further selected within the initial set, as best candidates to assess the genetic variability over a high number of varieties and wild-related forms. In fact, it has been observed that only three new SSR markers (OLEST1, OLEST14, and OLEST16) may allow to discriminate all the analyzed genotypes. On the contrary, also the less polymorphic EST-derived markers, highlighting the presence of private or rare alleles, could find useful application and represent a valuable resource for genotyping, for the detection of specific cultivars or for DNA testing of olive oils obtained by blending different varieties.

Table 4 GenBank accession numbers (AN) of new EST-SSRs, related predicted genes, and putative biological process

EST-SSR	EST-SSR-sequenced allele (GenBank AN)	Predicted gene (GenBank AN)	Predicted gene	Biological process
OLEST1	KT281452	–	Unknown	–
OLEST2	KT281453	KT290269	RNA-binding protein	RNA processing
OLEST3	KT281454	KT340584	Mediator of RNA polymerase II transcription subunit 26b	Regulation of transcription
OLEST4	KT281455	KT340585	Stress response protein <i>nst1</i> -like	Stress response
OLEST5	KT281456	KT340586	Glycine-rich cell wall structural protein 1-like	Structural cell wall protein
OLEST6	KT281457	KT340587	Putative homeobox protein	Regulating the expression of targeted genes
OLEST7	KT281458	KT340588	Eukaryotic translation initiation factor 3 subunit A	Component of the eukaryotic translation initiation factor 3 (eIF-3) complex
OLEST8	KT281459	KT340589	Topoisomerase II-associated protein PAT1	RNA-binding protein involved in deadenylation-dependent decapping of mRNAs
OLEST9	KT281460	–	Unknown	–
OLEST10	KT281461	KT340590	LisH domain-containing protein C1711.05-like	Nucleocytoplasmic transport
OLEST11	KT281462	KT340591	Cysteine proteinase inhibitor 1-like	Defense response
OLEST12	KT281463	–	Unknown	–
OLEST13	KT281464	KT340592	Ethylene-responsive transcription factor ERF109-like	Defense response
OLEST14	KT281465	–	Unknown	–
OLEST15	KT281466	KT340593	N(6)-adenine-specific DNA methyltransferase 2	Methyltransferase activity
OLEST16	KT281467	KT340594	<i>PsbD</i> messenger RNA (mRNA) maturation factor Nac2, chloroplastic	Involved, directly or indirectly, in the processing of chloroplast-encoded <i>psbD</i> mRNA to its mature form
OLEST17	KT281468	KT340595	FAS-associated factor 2-B	Endoplasmic reticulum-associated degradation
OLEST18	KT281469	KT340596	Serine carboxypeptidase-like 7	Proteolysis
OLEST19	KT281470	–	Unknown	–
OLEST20	KT281471	–	Unknown	–
OLEST21	KT281472	KT340597	Dehydrin, <i>DHN</i>	Response to water stress, temperature stimulus
OLEST22	KT281473	KT340598	Transcriptional corepressor LEUNIG	Cell differentiation
OLEST23	KT281474	KT340599	MYB2	Dehydration at the transcriptional level
OLEST24	KT281475	–	Unknown	–
OLEST25	KT281476	KT340600	Ethylene-responsive transcription factor 4	Ethylene-responsive element-binding proteins
OLEST26	KT281477	–	Unknown	–

The transferability of the 26 new OLEST SSR markers developed on cultivated olives was assessed on related taxa within the *Olea* genus, including wild plants (*O. europaea* subsp. *europaea* var. *sylvestris*), *O. europaea* subspecies, and related *Olea* species, as previously tested for other markers (Besnard et al. 2011; Besnard and El Bakkali 2014; Rallo et al. 2003). Most of them resulted to be easily amplifiable and detectable across all genotypes, and just a few showed problems of amplification or scoring, probably due to polymorphisms (insertions/deletions or base mutations) in primer regions or, regarding subsp. *maroccana* and *cerasiformis*, to ploidy level. However, the majority of the OLEST SSR markers resulted to be highly suitable to discriminate among related *Olea* taxa, clearly distinguishing olive varieties from related taxa (wilds, subspecies, and subgenera),

and could be applied as useful tools to investigate comparative genomics, genetic differentiation, and evolutionary dynamics within the *Olea* genus (Ma et al. 2010; Qiu et al. 2010; Varshney et al. 2005).

Compared to the ten best-ranked di-nucleotide SSRs (Baldoni et al. 2009), the ten most informative OLEST-SSRs showed a lower variability and a higher frequency of null alleles, probably because EST-derived SSRs are associated with transcribed regions, thus reflecting a lower genetic variability than genomic, neutral, and randomly selected markers as di-nucleotide SSRs (Hu et al. 2011; Leonarduzzi et al. 2016). However, both kinds of markers are well suitable to characterize the olive variability and their combined application could efficiently contribute to explore the olive germplasm and resolve variety identification problems arising from

the use of di-nucleotide markers. Best primer sequences and amplification protocols for genotyping application of these new markers have been released, in order to allow for a clear discrimination among genetic profiles and the use of multiplexing strategies.

The 26 EST-SSR markers developed in this work represent new particularly useful tools for the olive genotyping. In particular, ten OLEST markers resulted to be highly polymorphic and effective to discriminate olive cultivars and are proposed as a new set of markers relevant to discriminate genotypes within the *Olea* genus.

Acknowledgements This research was supported by the Project “OLEA—Genomics and Breeding of Olive,” funded by the MIPAF, D.M. 27011/7643/10, and by the Project “BeFOre—Bioresources for Oliviculture,” 2015–2019, H2020-MSCA-RISE Marie Skłodowska-Curie Research and Innovation Staff Exchange, Grant Agreement No. 645595. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions LB, RM, NC, and EA conceived and designed the experiments. RM, SM, NC, MR, and FB performed the experiments. FC, RM, MR, NC, FA, and FB analyzed the data. LB, GP, and EA contributed the reagents/materials/analysis tools. LB, RM, SM, MR, SM, FC, NC, and FB wrote the paper. LB and RM performed the maintenance of germplasm. GP and FC performed the assembly and mining of sequencing data. LB, RM, SM, MR, NC, FB, and EA reviewed the manuscript.

Compliance with ethical standards

Competing interests The authors declare that they have no competing interests.

Data archiving statement The sequence data of EST-SSRs and predicted genes developed in the present work are available in GenBank EST and predicted gene databases (<http://www.ncbi.nlm.nih.gov/genbank/>) under the accession numbers reported in Table 4.

References

- Acuna CV, Fernandez P, Villalba PV, García MN, Esteban Hopp H, Marcucci Poltri SN (2012) Discovery, validation, and in silico functional characterization of EST-SSR markers in *Eucalyptus globulus*. *Tree Genet Genomes* 8:289–301
- Aggarwal RK, Hendre PS, Varshney RK, Bhat PR, Krishnakumar V, Singh L (2007) Identification, characterization and utilization of EST-derived genic microsatellite markers for genome analyses of coffee and related species. *Theor Appl Genet* 114:359–372
- Alagna F, D’Agostino N, Torchia L, Servili M, Rao R, Pietrella M et al (2009) Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development. *BMC Genomics* 10:399
- Alagna F, Cirilli M, Galla G, Carbone F, Daddiego L, Facella P et al (2016) Transcript analysis and regulative events during flower development in olive (*Olea europaea* L.). *PLoS One* 11(4):e0152943
- Amos W, Hoffman JI, Frodsham A, Zhang L, Best S, Hill AVS (2007) Automated binning of microsatellite alleles: problems and solutions. *Mol Ecol Notes* 7:10–14
- Asadi AA, Monfared SR (2014) Characterization of EST-SSR markers in durum wheat EST library and functional analysis of SSR-containing EST fragments. *Mol Gen Genomics* 289(4):625–640
- 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* 54:2045–2053
- Baldoni L, Cultrera NGM, Mariotti R, Ricciolini C, Arcioni S, Vendramin GG et al (2009) A consensus list of microsatellite markers for olive genotyping. *Mol Breed* 24:213–231
- Baldoni L, Belaj A (2010) Olive. In: *Oil crop breeding*. ed. J. Vollmann and I. Rajcan, Springer
- Belaj A, del Carmen Dominguez-García M, Atienza SG, Urdiroz NM, De la Rosa R, Satovic Z et al (2012) Developing a core collection of olive (*Olea europaea* L.) based on molecular markers (DArTs, SSRs, SNPs) and agronomic traits. *Tree Genet Genomes* 8:365–378
- Besnard G, El Bakkali A (2014) Sequence analysis of single-copy genes in two wild olive subspecies (*Olea europaea* L.): nucleotide diversity and potential use for testing admixture. *Genome* 57:145–153
- Besnard G, Hernández P, Khadari B, Dorado G, Savolainen V (2011) Genomic profiling of plastid DNA variation in the Mediterranean olive tree. *BMC Plant Biol* 11:80
- Besnard G, Garcia-Verdugo C, De Casas RR, Treier UA, Galland N, Vargas P (2008) Polyploidy in the olive complex (*Olea europaea*): evidence from flow cytometry and nuclear microsatellite analyses. *Ann Bot* 101:25–30
- Biswas MK, Chai L, Mayer C, Xu Q, Guo W, Deng X (2012) Exploiting BAC-end sequences for the mining, characterization and utility of new short sequences repeat (SSR) markers in *Citrus*. *Mol Biol Rep* 39:5373–5386
- 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* 35:1–14
- Bocacci P, Beltramo C, Prando MS, Lembo A, Sartor C, Mehlenbacher SA et al (2015) In silico mining, characterization and cross-species transferability of EST-SSR markers for European hazelnut (*Corylus avellana* L.). *Mol Breed* 35:1–14
- 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* 35:449–462
- Bradbury D, Smithson A, Krauss SL (2013) Signatures of diversifying selection at EST-SSR loci and association with climate in natural *Eucalyptus* populations. *Mol Ecol* 22:5112–5129
- Butler JM (2007) Short tandem repeat typing technologies used in human identity testing. *BioTechniques* 43:2–5
- Cabezas JA, Ibanez J, Lijavetzky D, Velez D, Bravo G, Rodriguez V et al (2011) A 48 SNP set for grapevine cultivar identification. *BMC Plant Biol* 11:153
- 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
- Chankaew S, Isemura T, Isobe S, Kaga A, Tomooka N, Somta P et al (2014) Detection of genome donor species of neglected tetraploid crop *Vigna reflexo-pilosa* (creole bean), and genetic structure of diploid species based on newly developed EST-SSR markers from azuki bean (*Vigna angularis*). *PLoS One* 9:e104990
- Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. *Mol Biol Evol* 24:621–631
- Chen M, Wu J, Wang L, Zhang X, Blair MW, Jia J, Wang S (2014) Development of mapped simple sequence repeat markers from common bean (*Phaseolus vulgaris* L.) based on genome sequences of a Chinese landrace and diversity evaluation. *Mol Breed* 33:489–496

- Chiappetta A, Muto A, Bruno L, Woloszynska M, VanLijsebettens M, Bitonti MB (2015) A dehydrin gene isolated from feral olive enhances drought tolerance. *Front Plant Sci* 6:392
- 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. *Theor Appl Genet* 104:223–228
- 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* 8:127
- Cipriani G, Spadotto A, Jurman I, Di Gaspero G, Crespan M, Meneghetti S et al (2010) The SSR-based molecular profile of 1005 grapevine (*Vitis vinifera* L.) accessions uncovers new synonymy and parentages, and reveals a large admixture amongst varieties of different geographic origin. *Theor Appl Genet* 121:1569–1585
- Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674–3676
- Corrado G, Alagna F, Rocco M, Renzone G, Varricchio P, Coppola V et al (2012) Molecular interactions between the olive and the fruit fly *Bactrocera oleae*. *BMC Plant Biol* 12:86
- Cubry P, Pujade-Renaud V, Garcia D, Espeout S, Le Guen V, Granet F, Seguin M (2014) Development and characterization of a new set of 164 polymorphic EST-SSR markers for diversity and breeding studies in rubber tree (*Hevea brasiliensis* Müll. Arg.). *Plant Breed* 133: 419–426
- Dai F, Tang C, Wang Z, Luo G, He L, Yao L (2015) De novo assembly, gene annotation, and marker development of mulberry (*Morus atropurpurea*) transcriptome. *Tree Genet Genomes* 11:1–11
- De Keyser E, De Riek J, Van Bockstaele E (2009) Discovery of species-wide EST-derived markers in *Rhododendron* by intron-flanking primer design. *Mol Breed* 23:171–178
- 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 Notes* 2:265–267
- De la Rosa R, Belaj A, Muñoz-Mérida A, Trelles O, Ortíz-Martín I, González-Plaza JJ et al (2013) Development of EST-derived SSR markers with long-core repeat in olive and their use for paternity testing. *J Amer Soc Hort Sci* 138:290–296
- Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc Ser B Stat Methodol*:1–38
- Díaz A, de la Rosa R, Martín A, Rallo P (2006) 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
- Díez CM, Imperato A, Rallo L, Barranco D, Trujillo I (2012) Worldwide core collection of olive cultivars based on simple sequence repeat and morphological markers. *Crop Sci* 52:211–221
- Díez CM, Trujillo I, Martínez-Urdiroz N, Barranco D, Rallo L, Marfil P, Gaut BS (2014) Olive domestication and diversification in the Mediterranean Basin. *New Phytol* 206:436–447
- Dominguez-Garcia MDC, Belaj A, De la Rosa R, Satovic Z, Heller-Uszynska K, Kilian A (2012) Development of DArT markers in olive (*Olea europaea* L.) and usefulness in variability studies and genome mapping. *Sci Hortic* 136:50–60
- Du FK, Xu F, Qu H, Feng S, Tang J, Wu R (2013) Exploiting the transcriptome of Euphrates poplar, *Populus euphratica* (Salicaceae) to develop and characterize new EST-SSR markers and construct an EST-SSR database. *PLoS One* 8:e61337
- Dutta S, Kumawat G, Singh BP, Gupta DK, Singh S, Dogra V et al (2011) Development of genic-SSR markers by deep transcriptome sequencing in pigeonpea [*Cajanus cajan* (L.) Millspaugh]. *BMC Plant Biol* 11:17
- Duran C, Singhania R, Raman H, Batley J, Edwards D (2013) Predicting polymorphic EST-SSRs in silico. *Mol Ecol Resour* 13:538–545
- 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
- El-Rodeny W, Kimura M, Hirakawa H, Sabah A, Shirasawa K, Sato S et al (2014) Development of EST-SSR markers and construction of a linkage map in faba bean (*Vicia faba*). *Breed Sci* 64:252
- Emebiri LC (2010) An EST-SSR marker tightly linked to the barley male sterility gene (*msg6*) located on chromosome 6H. *J Hered* 101:769–774
- Essalouh L, El Aabidine AZ, Contreras S, Ben Sadok I, Santoni S, Khadari B, Jakse J, Bandelj D (2014) Genomic and EST microsatellite loci development and use in olive: molecular tools for genetic mapping and association studies. *Acta Hort* 1057:543–549
- Feng SP, Li WG, Huang HS, Wang JY, Wu YT (2009) Development, characterization and cross-species/genera transferability of EST-SSR markers for rubber tree (*Hevea brasiliensis*). *Mol Breed* 23: 85–97
- Ferrao LFV, Caixeta ET, Pena G, Zambolim EM, Cruz CD, Zambolim L (2015) New EST-SSR markers of *Coffea arabica*: transferability and application to studies of molecular characterization and genetic mapping. *Mol Breed* 35:1–5
- Fu N, Wang Q, Shen HL (2013) De novo assembly, gene annotation and marker development using Illumina paired-end transcriptome sequences in celery (*Apium graveolens* L.). *PLoS One* 8:e57686
- Fungtammaman A, Ananda G, Hile SE, Su MSW, Sun C, Harris R et al (2015) Accurate typing of short tandem repeats from genome-wide sequencing data and its applications. *Genome Res* 25(5):736–749
- Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A (2003) ExpASY: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 31:3784–3788
- Guo DL, Guo MX, Hou XG, Zhang GH (2014) Molecular diversity analysis of grape varieties based on iPBS markers. *Biochem Sys Ecol* 52:27–32
- Haouane H, El Bakkali A, Moukhli A, Tollon C, Santoni S, Oukabli A, Modafar CE, 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* 139:1083–1094
- Heesacker A, Kishore VK, Gao W, Tang S, Kolkman JM, Gingle A et al (2008) SSRs and INDELs mined from the sunflower EST database: abundance, polymorphisms, and cross-taxa utility. *Theor Appl Genet* 117:1021–1029
- Hinchliffe DJ, Turley RB, Naoumkina M, Kim HJ, Tang Y, Yeater KM, Li P, Fang DD (2011) A combined functional and structural genomics approach identified an EST-SSR marker with complete linkage to the Ligon lintless-2 genetic locus in cotton (*Gossypium hirsutum* L.). *BMC Genomics* 12:445
- Hosseini-Mazinani M, Mariotti R, Torkzaban B, Sheikh-Hassani M, Ataei S, Cultrera NGM, Pandolfi S, Baldoni L (2014) High genetic diversity detected in olives beyond the boundaries of the Mediterranean Sea. *PLoS One* 9:e93146
- Hu J, Wang L, Li J (2011) Comparison of genomic SSR and EST-SSR markers for estimating genetic diversity in cucumber. *Biol Plant* 55: 577–580
- Huang H, Lu J, Ren Z, Hunter W, Dowd SE, Dang P (2011) Mining and validating grape (*Vitis* L.) ESTs to develop EST-SSR markers for genotyping and mapping. *Mol Breed* 28:241–254
- Jiao Y, Ma RJ, Shen ZJ, Yu ML (2014) Development of Ty1-*copia* retrotransposon-based SSAP molecular markers for the study of genetic diversity in peach. *Biochem Sys Ecol* 57:270–277

- Kalia RK, Rai MK, Kalia S, Singh R, Dhawan AK (2011) Microsatellite markers: an overview of the recent progress in plants. *Euphytica* 177:309–334
- Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol Ecol* 16:1099–1106
- Kaur S, Panesar PS, Bera MB, Kaur V (2015) Simple sequence repeat markers in genetic divergence and marker-assisted selection of rice cultivars: a review. *Crc Cr Rev Anal Chem* 55:41–49
- Kaya HB, Cetin O, Kaya H, Sahin M, Sefer F, Kahraman A, Tanyolac B (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
- Kaya HB, Cetin O, Kaya HS, Sahin M, Sefer F, Tanyolac B (2016) Association mapping in Turkish olive cultivars revealed significant markers related to some important agronomic traits. *Biochem Genet* 54:506–533
- Khadari B, El Bakkali A, Zine El Aabidine A, Essalouh L, Contreras S, Sadok B, Costes E, Moukhli A (2014) How can we efficiently characterize genes of agronomic interest in olive: towards the genetic association studies? *Acta Hort* 1057:551–558
- King DG (2012) Evolution of simple sequence repeats as mutable sites. In: Tandem repeat polymorphisms: genetic plasticity, neural diversity and disease. Ed. A.J. Hannan, Landes Bioscience and Springer Science+Business Media
- Leonarduzzi C, Spanu I, Labriola M, González-Martínez SC, Piotti A, Vendramin GG (2016) Development and characterization of three highly informative est-ssr multiplexes for *Pinus halepensis* Mill. and their transferability to other Mediterranean pines. *Plant Mol Biol Rep* 1–10
- Li D, Deng Z, Qin B, Liu X, Men Z (2012) De novo assembly and characterization of bark transcriptome using Illumina sequencing and development of EST-SSR markers in rubber tree (*Hevea brasiliensis* Muell. Arg.). *BMC Genomics* 13:192
- Li S, Yin T, Wang M, Tuskan GA (2011) Characterization of microsatellites in the coding regions of the *Populus* genome. *Mol Breed* 27:59–66
- Lima LS, Gramacho KP, Gesteira AS, Lopes UV, Gaiotto FA, Zaidan HA, Pires JL, Cascardo JCM, Micheli F (2008) Characterization of microsatellites from cacao—*Moniliophthora perniciosa* interaction expressed sequence tags. *Mol Breed* 22:315–318
- Liu C, Yuan D, Lin Z (2014) Construction of an EST-SSR-based interspecific transcriptome linkage map of fibre development in cotton. *J Genet* 93:689–697
- Long Y, Wang Y, Wu S, Wang J, Tian X, Pei X (2015) De novo assembly of transcriptome sequencing in *Caragana korshinskii* Kom. and characterization of EST-SSR markers. *PLoS One* 10(1):e0115805
- Loureiro J, Rodriguez E, Costa A, Santos C (2007) Nuclear DNA content estimations in wild olive (*Olea europaea* L. ssp. *europaea* var. *sylvestris* Brot.) and Portuguese cultivars of *O. europaea* using flow cytometry. *Genet Res Crop Evol* 54:21–25
- Luro FL, Costantino G, Terol J, Argout X, Allario T, Wincker P, Talon M, Ollitrault P, Morillon R (2008) Transferability of the EST-SSRs developed on Nules clementine (*Citrus clementina* Hort ex Tan) to other *Citrus* species and their effectiveness for genetic mapping. *BMC Genomics* 9:287
- Ma JQ, Zhou YH, Ma CL, Yao MZ, Jin JQ, Wang XC, Chen L (2010) Identification and characterization of 74 novel polymorphic EST-SSR markers in the tea plant, *Camellia sinensis* (Theaceae). *Am J Bot* 97:e153–e156
- Morgante M, Hanafey M, Powell W (2002) Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nat Genet* 30:194–200
- Mousavi S, Mazinani MH, Arzani K, Ydollahi A, Pandolfi S, Baldoni L, Mariotti R (2014) Molecular and morphological characterization of Golestan (Iran) olive ecotypes provides evidence for the presence of promising genotypes. *Genet Resour Crop Evol* 61:775–785
- Peakall R, Smouse P (2012) GenALEX 6.5: genetic analysis in excel. Population genetic software for teaching and research—an update. *Bioinformatics* 28:2537–2539
- Perrier X, Jacquemoud-Collet JP (2006) DARwin software. <http://darwin.cirad.fr/darwin>
- Qi X, Cui F, Li Y, Ding A, Li J, Chen G, Wang H (2010) Molecular tagging wheat powdery mildew resistance gene pm21 by EST-SSR and STS markers. *Mol Plant Breed* 1
- Qiu L, Yang C, Tian B, Yang JB, Liu A (2010) Exploiting EST databases for the development and characterization of EST-SSR markers in castor bean (*Ricinus communis* L.). *BMC Plant Biol* 10:278
- Rallo P, Dorado G, Martín A (2000) Development of simple sequence repeats (SSRs) in olive tree (*Olea europaea* L.). *Theor Appl Genet* 101:984–989
- Rallo P, Tenzer I, Gessler C, Baldoni L, Dorado G, Martín A (2003) Transferability of olive microsatellite loci across the genus *Olea*. *Theor Appl Genet* 107:940–946
- Ramchiary N, Nguyen VD, Li X, Hong CP, Dhandapani V, Choi SR, Yu G, Piao ZY, Lim YP (2011) Genic microsatellite markers in *Brassica rapa*: development, characterization, mapping, and their utility in other cultivated and wild *Brassica* relatives. *DNA Res* 18: 305–320
- Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J. Hered* 86: 248–249
- Rousset F (2008) Genepop'007: a complete reimplemention of the Genepop software for Windows and Linux. *Mol Ecol Resour* 8: 103–106
- Sabino Gil F, Busconi M, da Câmara MA, Fogher C (2006) Development and characterization of microsatellite loci from *Olea europaea*. *Mol Ecol Notes* 6:1275–1277
- Sarri V, Baldoni L, Porceddu A, Cultrera NGM, Contento A, Frediani M, Belaj A, Trujillo I, Cionini PG (2006) Microsatellite markers are powerful tools for discriminating among olive cultivars and assigning them to geographically defined populations. *Genome* 49: 1606–1615
- Scott KD, Egger P, Seaton G, Rossetto M, Ablett EM, Lee LS, Henry RJ (2000) Analysis of SSRs derived from grape ESTs. *Theor Appl Genet* 100:723–726
- Sefc KM, Lopes MS, Mendonça D, Rodrigues dos Santos M, da Câmara Machado ML, da Câmara MA (2000) Identification of microsatellite loci in olive (*Olea europaea*) and their characterization in Italian and Iberian olive trees. *Mol Ecol* 9:1171–1173
- Shen D, Bo W, Xu F, Wu R (2014) Genetic diversity and population structure of the Tibetan poplar (*Populus szechuanica* var. *tibetica*) along an altitude gradient. *BMC Genet* 15(Suppl 1):S11
- Shiferaw E, Pè ME, Porceddu E, Ponnaiah M (2012) Exploring the genetic diversity of Ethiopian grass pea (*Lathyrus sativus* L.) using EST-SSR markers. *Mol Breed* 30:789–797
- Shirasawa K, Oyama M, Hirakawa H, Sato S, Tabata S, Fujioka T et al (2011) An EST-SSR linkage map of *Raphanus sativus* and comparative genomics of the Brassicaceae. *DNA Res* 18:221–232
- Sun J, Yin H, Li L, Song Y, Fan L, Zhang S, Wu J (2015) Evaluation of new IRAP markers of pear and their potential application in differentiating bud sports and other Rosaceae species. *Tree Genet Genomes* 11:1–13
- Targońska M., Bolibok-Bragoszewska H., Rakoczy-Trojanowska M (2015) Assessment of genetic diversity in *Secale cereale* based on SSR markers. *Plant Mol Biol Rep* 1–15
- This P, Jung A, Boccacci P, Borrego J, Botta R, Costantini L et al (2004) Development of a standard set of microsatellite reference alleles for identification of grape cultivars. *Theor Appl Genet* 109:1448–1458

- Torkzaban B, Kayvanjoo AH, Ardalan A, Mousavi S, Mariotti R, Baldoni L et al (2015) Machine learning based classification of microsatellite variation: an effective approach for phylogeographic characterization of olive populations. *PLoS One* 10(11):e0143465
- Toth G, Gaspari Z, Jurka J (2000) Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res* 10:967–981
- Trujillo I, Ojeda MA, Urdiroz NM, Potter D, Barranco D, Rallo L, Diez CM (2014) Identification of the worldwide olive germplasm Bank of Cordoba (Spain) using SSR and morphological markers. *Tree Genet Genomes* 10:141–155
- Varshney RK, Graner A, Sorrells ME (2005) Genic microsatellite markers in plants: features and applications. *Trends Biotechnol* 23(1):48–55
- Wang Z, Yan H, Fu X, Li X, Gao H (2013) Development of simple sequence repeat markers and diversity analysis in alfalfa (*Medicago sativa* L.). *Mol Biol Rep* 40:3291–3298
- Weeks DE, Conley YP, Ferrell RE, Mah TS, Gorin MB (2002) A tale of two genotypes: consistency between two high-throughput genotyping centers. *Genome Res* 12:430–435
- Xie WG, Zhang XQ, Cai HW, Liu W, Peng Y (2010) Genetic diversity analysis and transferability of cereal EST-SSR markers to orchardgrass (*Dactylis glomerata* L.). *Biochem Syst Ecol* 38:740–749
- Xu SC, Gong YM, Mao WH, Hu QZ, Zhang GW, Fu W, Xian QQ (2012) Development and characterization of 41 novel EST-SSR markers for *Pisum sativum* (Leguminosae). *Am J Bot*:e149–e153
- Yang AH, Zhang JJ, Tian H, Yao XH (2012) Characterization of 39 novel EST-SSR markers for *Liriodendron tulipifera* and cross-species amplification in *L. chinense* (Magnoliaceae). *Am J Bot* e460–e464
- Yao MZ, Ma CL, Qiao TT, Jin JQ, Chen L (2012) Diversity distribution and population structure of tea germplasm in China revealed by EST-SSR markers. *Tree Genet Genomes* 8:205–220
- You Y, Liu D, Liu H, Zheng X, Diao Y, Huang X, Hu Z (2015) Development and characterisation of EST-SSR markers by transcriptome sequencing in taro (*Colocasia esculenta* (L.) Schoot). *Mol Breed* 35:1–11
- Zeng S, Xiao G, Guo J, Fei Z, Xu Y, Roe BA, Wang Y (2010) Development of a EST dataset and characterization of EST-SSRs in a traditional Chinese medicinal plant, *Epimedium sagittatum* (Sieb. Et Zucc.) maxim. *BMC Genomics* 11:94
- Zhang H, Wei L, Miao H, Zhang T, Wang C (2012) Development and validation of genic-SSR markers in sesame by RNA-seq. *BMC Genomics* 13:316
- Zhang LY, Bernard M, Leroy P, Feuillet C, Sourdille P (2005) High transferability of bread wheat EST-derived SSRs to other cereals. *Theor Appl Genet* 111:677–687
- Zhang M, Mao W, Zhang G, Wu F (2014) Development and characterization of polymorphic EST-SSR and genomic SSR markers for Tibetan annual wild barley. *PLoS One* 9:e94881
- Zhang L, Yuan M, Tao A, Xu J, Lin L, Fang P, Qi J (2015) Genetic structure and relationship analysis of an association population in jute (*Corchorus* spp.) evaluated by SSR markers. *PLoS One* 10:e0128195
- Zhao Y, Williams R, Prakash CS, He G (2012) Identification and characterization of gene-based SSR markers in date palm (*Phoenix dactylifera* L.). *BMC Plant Biol* 12:237
- Zhou C, He X, Li F, Weng Q, Yu X, Wang Y, Li M, Shi J, Gan S (2014) Development of 240 novel EST-SSRs in *Eucalyptus* L'Hérit. *Mol Breed* 33:221–225
- Zhu H, Senalik D, McCown BH, Zeldin EL, Speers J, Hyman J et al (2012) Mining and validation of pyrosequenced simple sequence repeats (SSRs) from American cranberry (*Vaccinium macrocarpon* Ait.). *Theor Appl Genet* 124:87–96