

# Tag-derived simple sequence repeat markers of olive



**ASSIGN  
BUSTER**

## Identification and functional annotation of expressed sequence Tag-Derived Simple Sequence Repeat Markers of olive ( *Olea europaea* )

Olive tree ( *Olea europaea* L.) is one of the most important oil producing crops in world, the genetic identification of several genotypes by using molecular markers is the first step in breeding programs. A large number of *Olea europaea* expressed sequence tags (ESTs) 11, 215 were done from the NCBI database and used to search for microsatellites. Our result Explained that 8295 SSRs were present and it's percentage of occurrence which about 77. 6%, 11. 84%, 8. 62%, 0. 84%, 0. 77% and 0. 29% for Mononucleotide, trinucleotide, dinucleotide, hexanucleotide, pentanucleotide and tetranucleotide respectively. The appearance of the AAG/CTT repeat was highly percentage in trinucleotide and percentage of AG/CT was highly in dinucleotide repeats. By using flanking region of SSRs repeat we designed 1, 801 EST-SSR primer pairs. The result obtained from Functional annotation of olive EST sequences containing SSRs indicated that 81% of these sequence having homology with known proteins, while 1. 55% was homologous to hypothetical or unknown proteins and the 17. 37% sequences did not possess homology with any known proteins. Gene Ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) annotation revealed that EST containing SSRs were implicated in diverse biological process include cellular and metabolic process, while in molecular function includes catalytic activity, binding and enzyme regulator activity. A total of 93 different pathways were significant matches in the KEGG database, which divided onto Carbohydrate metabolism such as glycolysis/gluconeogenesis pathway and the Energy metabolism such as Carbon fixation in photosynthetic organism pathway,

also this included 11 different pathways from Lipid metabolism such as Fatty acid biosynthesis pathway. We isolate a genomic DNA from 9 olive cultivars and tested with 25 random selected primer pairs for amplification and polymorphism detection. All tested primers, exhibited successfully amplified and detected polymorphism.

Olive tree (*Olea europaea* L.) is one of the most superannuated and important long lived fruit species in Mediterranean [1], its diploid species ( $2n = 2x = 46$ ), and the genome size range between 2.90 pg/2C and 3.07 pg/2C, with 1C = 1,400 - 1,500 Mbp [2]. *Olea europaea* is one of the first domesticated crops from *Oleaceae* family, and it cultivated for table olives and edible oil [3], a long history of olive cultivation in the Middle East was descriptions by archaeology and botanists [4]. The olive cultivars are high of number that more than 1200 cultivars [5], also the accessions are available in a large numbers in olive producing countries, that occurrence a problems for germplasm preservation and its management [6].

The genetic identification and characterizing of several genotypes by using molecular markers is the first step in breeding programs [7], and by increased rate of mutation in microsatellites repeats that show a highly level of length polymorphism [8]. With the improvement and increasing of DNA sequencing technology, sequencing of expressed gene are used to obtain a large collection of EST which are isolated from a specific tissue and stage on organism [3]. Recent EST-SSR studies have reported that the EST is uses a source of SSRs and that reveal highly polymorphism [9]. EST sequences Available in public database and by using bioinformatics tools can determine and development of SSR markers in that EST sequences [10]. In olive that <https://assignbuster.com/tag-derived-simple-sequence-repeat-markers-of-olive/>

can be allow to development of new functional marker and use it in molecular breeding [11]. Also it can use as useful tools for gene and marker discovery, gene mapping and functional comparative studies. EST-SSRs recently reported in several plant species, such as Musa [12], Finger Millet [13], Jatropha Curcas [14], Pineapple [15], Celery [16], Lettuce [17], Barley [18], Radish [19], Citrus [20], Watermelon [21], Sugarcane [22], grapes [23], Cereal species [24] and bread wheat [25]. A large number of EST sequences in olive are available on database it can be a useful resource to develop gene based markers. The aim of this study was to use bioinformatics tools to develop and identify a new genic marker EST-SSR in Olive, to compare the frequency and distribution of different repeat types in genic sequences. Also determine the localization of these primers in different pathways in plant, to use it as tools to differences between the olive cultivars.

### **The Source of Sequence, screening and primer designing of microsatellites.**

EST database used a source of olive EST sequences from NCBI (<http://www.ncbi.nlm.nih.gov>). A total of 11, 215 ESTs sequences of *Olea europaea* are available and used in this study. Identification of SSRs by using the perl script MISA (MicroSATellite identification tool; <http://pgrc.ipk-gatersleben.de/misa/>). The criteria used to determine SSR repeat were: mononucleotide  $\geq 10$ , dinucleotide  $\geq 6$ , trinucleotide, tetranucleotide, pentanucleotide, and hexanucleotide  $\geq 5$ , and the maximal number of bases interrupting two SSRs in a compound microsatellite are 100 bp. The flanking region of SSR motifs used to design SSR primers by using primer3\_core [26]. The parameters used were: optimum length of primer 20 nucleotides, optimum annealing

temperature ( $T_m$ ) of 58°C, expected amplified products size of 100-500 bp and optimum G/C content 50 %.

## **Validation of designed primer**

For primer validation, we designed 25 EST-SSR primers and test these primers on 9 olive cultivars. Total genomic DNA was extracted from olive leaves using Plant Genomic DNA Kit (QiGen). PCR amplification was conducted in 25  $\mu$ l reactions containing 50 ng of template DNA, 2.5 mM  $MgCl_2$ , 5  $\mu$ l 5X PCR buffer, 0.5 mM each primer, 0.5 U Taq DNA polymerase, and 2.5 mM dNTPs. The PCR cycling profile was 94°C for 5 min, 35 cycles at 94°C for 45 s, the optimum annealing temperature for each primer pair shown on (Table S1) for 50 s, 72°C for 45 s, and a final extension at 72°C for 10 min. The quality of the PCR product was checked by mixing it with an equal volume of loading buffer and then visualizing the band on a 1.5% agarose gel in TBE buffer at 100 W for 120 min.

## **Identification of EST-SSRs Putative Function annotation**

The putative function annotation of EST sequences contains SSR performed by used Blast2go program [27] to BLAST against a reference database. Also blast2go program are containing many features such as Gene Ontology (GO), Enzyme Commission (EC), and KEGG annotation.

## **Distribution of various repeat type in olive**

Our result referred to 4,088 EST sequences about 36.45% from 11,215 of *Olea europaea* EST sequences as containing 8,295 various motif SSRs that Due to the EST sequences maybe contain more than one SSR motif (Table 1),

and this number based on the criteria we used it to identify SSR motifs in the EST sequence.

The investigation of different types of SSR repeats in our result showed that the highest percentage of appearance mononucleotide repeats were 77.64%, followed by trinucleotide 11.84%, dinucleotide 8.62%, hexanucleotide 0.84%, pentanucleotide 0.77% and tetranucleotide 0.29%(Fig. 1). The higher abundant of trinucleotide in coding regions were consistent with the previous studies in eukaryotic genomes [28, 31].

In mononucleotide A/T repeats 88.8% were higher than the G/C 11.2% motifs, and these results were proportionate with SSRs analysis of chloroplast SSRs on *Olea* species [29] and with SSRs analysis of major cereal organelle genome [28]. GA motifs were representing 55% from dinucleotide motifs in olive EST sequences. According to previously studies from foxtail millet [31], barley, maize, rice, sorghum and wheat [30], GA motifs were the most abundant motifs in these crops. AG/CT and GA/TC motifs were the most frequent respectively and CG motifs the lowest frequencies were found in olive, this case reported in the distribution of microsatellites on three different plant families that Brassicaceae, Solanaceae and Poaceae [32]. Dinucleotide motif can represent to multiple codons that depending on the reading frame and can translate into different amino acids such as, AG/CT motif can represent AGA, GAG, CUC and UCU codons in mRNA and translate into the amino acids Glu, Arg, Leu and Ala respectively, Ala and Leu were present in protein at higher frequencies, hence the higher incidence of GA, CT motifs in the EST sequences [33]. This could be one of the reasons why GA, CT motifs are present at such highly appearance in EST collections [34], <https://assignbuster.com/tag-derived-simple-sequence-repeat-markers-of-olive/>

dinucleotide repeats that located on coding regions are more sensitive to any change such as any addition or deletion because that causes a frame shifts and will give different amino acids [35]. As for trinucleotide TCT, TTC were the most common repeat motif in olive EST (Table 2), While AAG/CTT motifs were the most common in chloroplast of *Olea* species SSRs [29], however, in other crops such as barley, maize, rice, sorghum and wheat CCG or AAC were the most common trinucleotide repeat [30].

Our results revealed that tetranucleotide motifs AATC, CTTT are the most common; however the most common in *Olea* species SSRs chloroplast were AAAG, CTTT [29]. Pentanucleotide and hexanucleotide AAAAT and GAAAAA respectively are the most common repeat motif in our results while [29] found AATCC was the most common on pentanucleotide in *Olea* species SSRs chloroplast and hexanucleotide was not found.

## **Design and validate of EST-SSR**

In this study, we designed 1,801 PCR primer pairs from the 8295 SSR motifs of *Olea europaea* EST, The designed primers were referred as Oe-ESSR\_XXXX, where Oe-ESSR referred to *Olea europaea* EST SSR, XXXX was referred the number of EST-SSR from 1-1801 and that different of the previously SSR primers designed from chloroplast sequences of *Olea* species [29]. This primers were listed in the (Table S1), and provide with all information related it such as Primer name, “ GenInfo Identifier” gi number of EST sequence, Repeat type, Repeat Sequence, Length of Repeat, Repeat start on sequence, Repeat end on sequence, Forward and Reverses Primer, T<sub>m</sub> (°C), Length of Primer (bp), product Length (bp), sequence of EST, Sequence Description, gene ontology, Enzyme code and Enzyme Name.

<https://assignbuster.com/tag-derived-simple-sequence-repeat-markers-of-olive/>

We use a sample of 25 primers randomly from these 1, 801 EST-SSR primers to validate it by using a genomic DNA isolated from 9 olive cultivars. All tested primers, exhibited successfully amplified and detected polymorphism(Fig. 2).

### **Putative Function annotation of EST-SSRs**

The putative function annotation of the EST sequences containing SSR performed by used Blast2go program [27]. According to the Blast2go result, 81% from EST sequence as homology with known proteins, while hypothetical or unknown proteins were 1. 55%, and 17. 37% of this EST sequences did not homology with any known proteins. The gene ontology of olive EST sequences containing SSRs using Blast2GO revealed that in the biological processes, the highly appearance of SSR were involved in cellular processes, metabolic, response to stimulus, biological regulation and developmental process, while Signaling, rhythmic processes and growth had the lowest SSR contents among these EST. The molecular function category includes catalytic activity and binding, while cell membrane and organelle were assigned in the cellular component category(Fig. 3). The Similar results were found on functional annotation of the citrus and date palm EST sequences containing SSRs [20, 36]. Our results agreement with the similar results obtained in [20] which suggested that genes were involved in protein metabolism and biosynthesis were well conserved in plants.

### **Functional classification by KEGG pathway analyses**

The KEGG Pathway analysis is useful tool to understand the molecular interaction and biological functions [37]. Our study exhibited a total of 93 different pathways include 253 enzymes target by 381 EST-SSR primers were <https://assignbuster.com/tag-derived-simple-sequence-repeat-markers-of-olive/>



significant matches in the KEGG database(Table S2), this data can Visualization by using circos software[38](fig. 4).

The higher occurrence of SSR on pathways indicated a good potential for using these molecular markers to targeting the enzyme related to the trait subjected in our study. This EST sequences contain SSR were categorized into metabolism, as well as its subcategories, including lipid metabolism(Table 3), carbohydrate metabolism, energy metabolism, amino acid metabolism, nucleotide metabolism and ' metabolism of cofactors and vitamins'.

In details, the mapping result can further investigated against the glycolysis/gluconeogenesis(Fig. 5), Oxidative phosphorylation(Fig. 6)and Fatty acid degradation(Fig. 7)pathways as an example of Carbohydrate metabolism, Energy metabolism and Lipid metabolism respectively.

## **Conclusion**

SSR markers are very important because it is co-dominant, highly polymorphic and can generate from functional regions of the genome. EST-SSR technique have the potential to generate phenotypically linked functional markers and a useful tool can use on genetic diversity, marker assisted selection and genome mapping in olive. This study exhibited the functional categorization of olive EST sequences containing SSR revealed that these ESTs representing in genes with cellular component, biological process and molecular function. This EST-SSR primers also providing with useful information to understand the biological functions and genes

interactions according to the localization of this primers in different pathways related to possible phenotypic differences between the olive cultivars.

## References

1. Gaby E, Mbanjo N, Tchoumboungang F, Mouelle AS, Oben JE, Nyine M, et al. Development of expressed sequence tags-simple sequence repeats ( EST-SSRs ) for Musa and their applicability in authentication of a Musa breeding population. *Afr J Biotechnol.* 2012; 11(71): 13546–59.
2. Naga BLRI, Mangamoori LN, Subramanyam S. Identification and characterization of EST-SSRs in finger millet (*Eleusine coracana* (L.) Gaertn.). *J Crop Sci Biotechnol.* 2012; 15(10): 9–16.
3. Wen M, Wang H, Xia Z, Zou M, Lu C, Wang W. Development of EST-SSR and genomic-SSR markers to assess genetic diversity in *Jatropha Curcas* L. *BMC Res Notes.* 2010; 3: 42.
4. Wo T. In silico mining for simple sequence repeat loci in a pineapple expressed sequence tag database and cross-species amplification of EST-SSR markers across Bromeliaceae. *Theor Appl Genet.* 2011; 123: 635–47.
5. Fu N, Wang PY, Liu XD, Shen HL. Use of EST-SSR markers for evaluating genetic diversity and fingerprinting celery (*apium graveolens* L cultivars). *Molecules.* 2014; 19: 1939–55.
6. Simko I. Development of EST-SSR markers for the study of population structure in lettuce (*Lactuca sativa* L.). *J Hered.* 2009; 100(2): 256–62.

7. Zhang M, Mao W, Zhang G, Wu F. Development and characterization of polymorphic ESTSSR and genomic SSR markers for tibetan annual wild barley. *PLoS One*. 2014; 9(4): 1-10.
8. Nakatsuji R, Hashida T, Matsumoto N, Tsuru M, Kubo N. Development of genomic and EST-SSR markers in radish ( *Raphanus sativus* L .). *Breed Sci*. 2011; 61: 413-9.
9. Liu S, Li W, Long D, Hu C, Zhang J. Development and Characterization of Genomic and Expressed SSRs in Citrus by Genome-Wide Analysis. *PLoS One*. 2013; 8(10): 1-10.
10. Campus P. Development of EST-SSRs in watermelon (*Citrullus lanatus* var. *lanatus*) and their transferability to *Cucumis* spp. *J Hortic Sci Biotechnol*. 2008; 83(6): 732-6.
11. Pinto LR, Oliveira KM, Ulian EC, Garcia AAF, de Souza AP. Survey in the sugarcane expressed sequence tag database (SUCEST) for simple sequence repeats. *Genome*. 2004; 47: 795-804.
12. Scott KD, Eggler P, Seaton G, Rossetto M, Ablett EM, Lee LS, et al. Analysis of SSRs derived from grape ESTs. *TAG Theor Appl Genet*. 2000; 100: 723-6.
13. Varshney RK, Thiel T, Stein N, Langridge P, Graner A. In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species. *Cell Mol Biol Lett*. 2002; 7: 537-46.
14. Gupta PK, Rustgi S, Sharma S, Singh R, Kumar N, Balyan HS. Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Mol Genet Genomics*. 2003; 270: 315-23.

15. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3-new capabilities and interfaces. *Nucleic Acids Res.* 2012; 40(15): 1-12.
16. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics.* 2005; 21(18): 3674-6.
17. Rajendrakumar P, Biswal AK, Balachandran SM, Sundaram RM. In silico analysis of microsatellites in organellar genomes of major cereals for understanding their phylogenetic relationships. *In Silico Biol.* 2008; 8: 87-104.
18. Filiz E, Koc I. In Silico chloroplast SSRs mining of *Olea* species. *BIODIVERSITAS.* 2012; 13(3): 114-7.
19. Kantety R V., La Rota M, Matthews DE, Sorrells ME. Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol Biol.* 2002; 48: 501-10.
20. Jia XP, Shi YS, Song YC, Wang GY, Wang TY, Li Y. Development of EST-SSR in foxtail millet (*Setaria italica*). *Genet Resour Crop Evol.* 2007; 54: 233-6.
21. Da Maia LC, De Souza VQ, Kopp MM, De Carvalho FIF, De Oliveira AC. Tandem repeat distribution of gene transcripts in three plant families. *Genet Mol Biol.* 2009; 32: 822-33.
22. Lewin B, Dover G. *Genes v.* Oxford University Press Oxford; 1994.
23. Cho YG, Ishii T, Temnykh S, Chen X, Lipovich L, McCouch SR, et al. Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *TAG Theor Appl Genet.* 2000; 100: 713-22.

24. Metzgar D, Bytof J, Wills C. Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome Res.* 2000; 10: 72–80.
25. Zhao Y, Williams R, Prakash CS, He G. Identification and characterization of gene-based SSR markers in date palm ( *Phoenix dactylifera* L .). *BMC Plant Biology*; 2012; 12: 237
26. Li D, Deng Z, Qin B, Liu X, Men Z. 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.* 2012; 13: 192.
27. Krzywinski M, Schein J, Birol I, Connors J, Krzywinski M, Schein J, et al. Circos•: An information aesthetic for comparative genomics  
Circos•: An information aesthetic for comparative genomics. *Genome Res.* 2009; 19: 1639–45.