International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2013): 4.438

Development of Annotated EST-SSR Database in Olive (*Olea europaea*)

Sami S. Adawy[1], Morad M. Mokhtar[2], Alsamman M. Alsamman[3], Mahmoud M. Sakr[4]

[1, 2, 3]Molecular Markers and Genome Mapping Department; Agricultural Genetic engineering Research Institute; Agricultural Research Center, Giza, 12619- Egypt.

[4]Academy of Scientific Research and Technology, Egypt

Abstract: ***Olive tree (Olea europaea L.) is one of the most important oil producing crops in the world and the genetic identification of several genotypes by using molecular markers is the first step in its breeding programs. A set of 1,801 well-informative EST-SSR primers targeting specific Olive genes included in different biological processes and pathways were generated using 11,215 Olive EST sequences acquired from the NCBI database. Our bioinformatics analytical procedure showed that 8295 SSR motifs were detected which belonged to different motif types with occurrences of 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 represented in trinucleotide and the representation of AG/CT was high in dinucleotide repeats. Results obtained from functional annotation of olives EST sequences targeted with our primers set indicated that 78.5% of these sequences having homology with known proteins, while 4.2% was homologous to hypothetical, predicted, unnamed or uncharacterized proteins and the 17.3% sequences did not possess homology with any known proteins. Our EST-SSR primer set cover a total of 92 biological pathways such as carbohydrate metabolism pathway, energy metabolism & carbon fixation in photosynthetic organism pathway including 11 pathways associated with lipid metabolism. A twenty five randomly selected primers were applied to 9 Egyptian cultivated olive accessions to test its amplification and polymorphism detection efficacy. All tested primers were successfully amplified and only 10 exhibited detectable polymorphism.***

Keywords: EST-SSR, bioinformatics, functional annotation, biological pathways.

1. Introduction

Olive tree (*Olea europaea* L.) is one of the most superannuated and important long lived fruit species in Mediterranean [1], it is a diploid species (2n = 2x = 46) with a genome size ranging 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 for oil production and the second most important oil fruit cultivated crop worldwide [3].Olive is a dependable source of edible oil and food for several thousands of years [4]–[6] . The large number of accessions cultivated in olive producing countries make the olive germplasm preservation and management a major problem as far as olive breeders are concerned [7].

The development for early selection strategies in olive breeding programs is a main goal at present [8] and in this view ,using molecular markers techniques for the identification and characterizing of several genotypes is the first step in modern olive breeding programs [9] and choosing a co-dominant , reliable and well amplified marker type is very crucial to start this process in order to significantly minimize the quantity of breeding starting materials and promotes the selection of desirable genotypes, which posses desired genes in its homozygous state [10].

Reflecting its increasing rate of mutation, micro-satellites repeats shows a highly level of length polymorphism [11] with a high evolution rates and a possible impact on the modification genes they are associated with. Not to mention that the typical role of mutation is to add or subtract repeat units which are both reversible and frequent, making SSR influence on genes regulation depending on the repeats number and provide a source of qualitative and quantitative variations [12].

These features granted SSR derived techniques its high heterozygosity [13]–[15] and the ability to differentiate between different accessions with distinct agronomical advantages , despite synonymous problems in many plant species [16]–[18].This arise the need of developing new derived SSR markers with a PCR primers rich resources more linked to desired genic regions in different plant species ,mean while the improvement and increasing of DNA sequencing technologies aid the increasing and sequencing of expressed genes was used to construct a large collection of EST libraries isolated from different tissues of various organism under distinct environmental conditions and through different development stages (Ozgenturk et al., 2010).

Recent studies reported the using EST libraries as a reliable resource for SSR derived markers taking in advance the availability of EST sequences in public databases and bioinformatics tools which detected SSR repeats and developed a PCR-based EST-SSR markers could reveal a high polymorphism in genic regions related to important agronomic traits [20], [21]. EST-SSRs markers reported in several plant species, such as *Musa* [22], Finger Millet [23], *Jatropha Curcas* [24], Pineapple [25] , Citrus [26], Watermelon [27], Sugarcane [28], and bread wheat [29].

In olive this technique could develop new functional markers with a flexibility to be used in marker-assisted selection in breeding programs and a useful tool for genes discovery,

gene mapping, and gene-gene interaction, functional and comparative studies.

Sequence public databases contain a large number of EST sequences derived from different olive cultivars under a variety of environmental conditions, stand as useful resources for developing gene based markers. The aim of this study was to use bioinformatics analytical procedures to detect SSRs in Olive's ESTs, compare the frequency and distribution of different repeat types in genic sequences, develop new genic EST-SSR markers suited for Olive genome, determine the localization of these primers targeted ESTs in different pathways and offer these primers in an informative illustration style to simplify the searching for trait - related markers in Olive breeding programs.

2. Materials and Methods

Source of Data: A total of 11,215 Olea europaea ESTs sequences were acquired from NCBI-EST database, these ESTs were isolated under distinct environmental conditions and through different developmental stages (http://www.ncbi.nlm.nih.gov).

Bioinformatics tools: SSRs identification was performed using the PERL script MISA (MIcroSAtellite identification tool; http://pgrc.ipk-gatersleben.de/misa/) and the criteria to determine SSR repeats were: mononucleotide (mono-) ≥ 10, dinucleotide (di-) ≥ 6, trinucleotide (tri-), tetranucleotide (tetra-), pentanucleotide (penta-), and hexanucleotide (hexa-) ≥ 5, and the number of maximum bases interrupting two SSRs to produce a compound microsatellite is 100 bp.

The flanking regions of SSR motifs were used to design SSR PCR-based primers using primer3_core [30]. The parameters used: optimum length of primer was 20 nucleotides, optimum annealing temperature (Tm) of 58°C, expected to amplify products size of 100-500 bp and optimum G/C content of 50 %.

Validation of designed primers: Twenty five PCR EST-SSR primers were randomly selected to validate its amplification efficacy, these EST-SSR primers were synthesized and applied on nine Olive cultivars adapted to the Egyptian environment (Maraki, Tofahi, Koratina, Pekoal, Manzanillo, Dolici, OjaziShami, Kronaki and Calamata).

Total genomic DNA was extracted from olive leaves using the Plant Genomic DNA Kit (Qiagen). PCR reaction content and PCR program cycles were summarized in (Supporting Information File S1).

Olive ESTs GO enrichment analysis: Only Olive EST sequences contain detectable SSR motifs and has generated valid primers through previous mentioned criteria were used in GO enrichment analysis by using Blast2GO pipeline tool [31] to assign gene ontology terms to EST products. BlastX search against the non-redundant (nr) NCBI database was used to analyze selected EST sequences with an Expect value (E-value) ≥1.0E-3 and the maximum hits for every gene was 20 hits. In the mapping and annotation steps of GO analysis, the default evidence codes weights (default=5) and Cutt-Off value score (default=55), respectively were used. The annotation step with GO-weight of 5 was given to map children terms of all EST sequences have hits.

3. Results and Discussion

Distribution of various repeat types of olive: Our result referred to 4,088 of Olea europaea EST sequences 36.45% out of 11,215 contains detectable SSR motifs matching our criteria, these ESTs contain 8,295 various SSR motifs. The gap between sequences contains simple repeats and repeat occurrence was due to the possibility that one SSR could contain more than one motif (Table 1).

Table 1: Summary of SSR repeats identified on Olea europaea EST sequences.

Searching item	Numbers
Total number of sequences examined:	11215
Total size of examined sequences (bp):	6566149
Total number of identified SSRs:	8295
Number of SSR containing sequences:	4088
Number of sequences containing more than one SSR:	1910
Number of SSRs present in compound formation:	2447

Our investigation of different SSR repeats types showed that the highest appearance percentage of mono- repeats were 77.64% , followed by tri- 11.84%, di- 8.62%, hexa- 0.84%, penta- 0.77% and tetra- 0.29% (Figure 1).The higher abundant of tri- in coding regions were consistent with previous studies in eukaryotic genomes [32], [33].

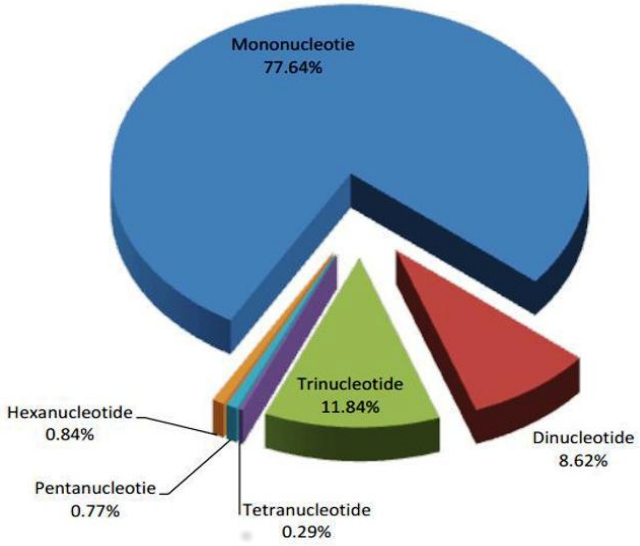

Figure 1: Frequencies (%) of different repeats type on Olea europaea EST sequences

The mono- motifs A/T were 88.8% higher than G/C motifs (11.2%), and these results were proportionate with SSR analysis of chloroplast on Olea species [34] and with SSR analysis of major cereal organelle genome [32]. In di- motifs, GA represented 55% of the di- motifs in olive EST sequences ,this agree with previous studies suggested that GA are the most abundant repeats type in foxtail millet [33], barley, maize, rice, sorghum and wheat [35]. AG/CT and GA/TC motifs were the most frequent respectively, while CG repeats were the lowest frequencies, this case was reported in

microsatellites distribution for Brassicaceae, Solanaceae and Poaceae [36]. The motifs Type of di- could represented in multiple codons depending on the open reading frame (ORF) regions which will be translated into different amino acids , for instance AG/CT motifs could represents AGA, GAG, CUC and UCU codons in mRNA , in this case it will be translated into the amino acids Glu, Arg, Leu and Ala respectively, therefore Ala and Leu will be presented in proteins at higher frequencies, hence the higher incidence of GA, CT motifs in the EST sequences [37]. This could be one of the reasons suggested to explain the highly representation GA, CT motifs appearance in EST collections [38]. di-repeats that located in coding regions are more sensitive to any change , such as substitutions , additions or deletions , as it causes a frame shifts which could give alternative amino acids [39]. Regarding tri- , the TCT and TTC motifs were the most common repeats in olive EST (Table 2), on the other hand AAG/CTT motifs were the most common in other studies focused on SSR types occurred in the chloroplast of Olea species [34], despite the fact that , CCG or AAC were the most common tri- repeats types in other crops such as barley, maize, rice, sorghum and wheat [35].

Table 2: Distribution of different repeat types on Olea europaea EST sequences.

Type of motif	Num. of motif	Distribution (%)*
A\T	5720	88.80%
G\C	720	11.20%
GA\TC	531	74%
AG\CT	91	12.70%
AC\AT\CA\CG\GT\TA\TG	93	13.30%
TCT\TTC	620	63.70%
AAG\AAT\ACC\AGA\AGC\AGG\ATA	62	6%
CCT\CGC\CTC\CTG\CTT\GAA\GAC	81	8%
ATG\ATT\CAC\CAG\CAT\CCA\CCG	67	7%
GAG\GAT\GCA\GCT\GGA\GGC\GGT	70	7%
GTG\TAA\TAT\TCA\TCC\TGA\TGC\CAA	48	5%
TTA\TTG\TGG\TGT\TCG\AAC	26	2.30%
GTT\GTA\GTA\GTA\GCG\GCC\CGG\ATC	8	1%

* Motif representation percentage against its repeat type.

Our results revealed that tetra- motifs AATC, CTTT are the most common; however the most common in Olea species SSRs chloroplast were AAAG, CTTT [34]. Penta- AAAAT and hexa-GAAAAA were the most common motifs in our result, while AATCC was the most common on penta- in Olea species chloroplast and hexa- was not found in this organelle [34].

EST-SSR PCR-based primer design : In this study, we used 4,088 EST sequences to design and select one of the most suitable PCR primer pairs. Only 1,801 EST sequences which contain detectable SSR motifs generated suitable primer pairs. The other ESTs 2287 sequences neither contain enough flanking regions to design a specific primer, or the generated primers didn't match our criteria which we managed by primer3_core tool [30]. The designed primers were referred as Oe-ESSR_xxxx, where Oe-ESSR is an abbreviating for Olea europaea EST-SSRs and xxxx are referring to the index of EST-SSR primers (start with 1 and end with 1801).

Gene ontology enrichment analysis for Olive EST-SSR sequences : All EST sequences which have generated an EST-SSR primer pairs by our mentioned criteria were annotated with Blast2go pipeline tool. In the BLAST step, out of the 1,801 EST sequences used, only 1413 have a homology with known proteins, while hypothetical, predicted, unnamed or uncharacterized proteins were 75 and only 313 sequences did not possess homology with any known proteins. Most of these hits have Expected values ≥ 1.E-27 (Figure 2-A) and the homology degrees ranging from 40.5% to 100% (Figure 2-B).

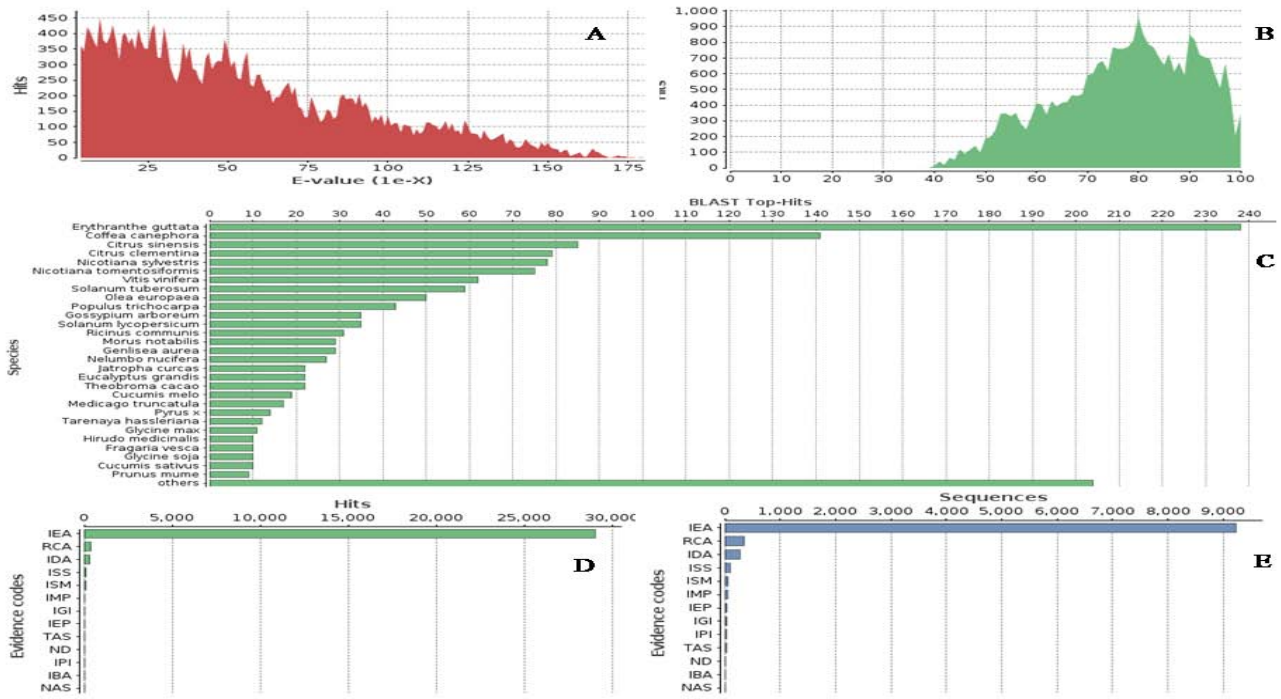

Figure 2: Gene ontology enrichment analysis statics for Olive EST-SSR sequences which have generated an EST-SSR primer pairs with Blast2go pipeline tool. A: E-value distribution for BLAST step against NCBI database, B: the homology degrees and similarity distribution for sequence queries with BLAST hits, C: Top hit distribution against different databases related to different genomes, D: evidence code distribution for blast hits, E: evidence code distribution for blast sequences.

Olea europea came in the ninth place in top-hit species distribution, maybe due to that, only sequences revealed SSR and produced PCR primers were used or lower number of olive sequences in the NCBI database compared to other species with finished and published genomes (Figure 2-C).

In the GO terms mapping step, only 1264 sequences were mapped with a total of GO terms reaches 6432.The number of GO terms assigned to every EST sequence differs from one to 49 terms and most EST sequences were mapped to terms inferred from electronic annotation (IEA), which is higher in evidence code distribution for both blast hits and sequences (Figure 2-D & 2-E).

In the annotation step, about 5090 GO terms were mapped to 1264 EST sequences, giving a GO mean level of 6.9 and revealing 256 sequences with known enzyme code (EC). The average length of sequences was 823 and sequences with length higher than 750 bp gain more annotation than other sequences. The other 537 EST sequences, which generated PCR primers and didn't reveal any annotation could be used as a tool to discover genomic regions with unknown function.

The three major GO functional groups: molecular function (GO: 0003674), biological process (GO: 0008150) and cellular components (GO: 0008370) revealed subgroups with related biological functions. Out of 5090 GO terms revealed in our result, about 1348 are linked to molecular function, 1244 GO are related to cellular components and 2498 GO terms associated with biological processes (Figure

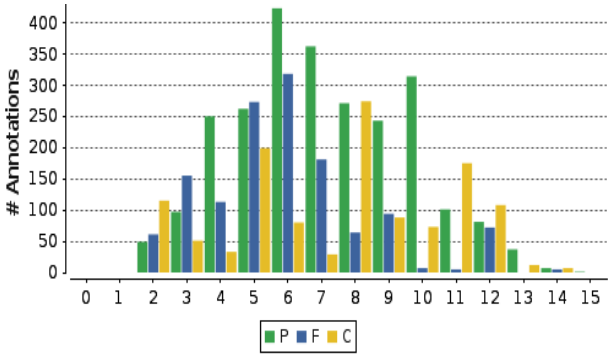

Figure 3: The distribution of GO terms linked to our EST-SSR primers along the three major functional groups: GO Biological process (P), Molecular function (F) and cellular components (C).Go level with a total annotation = 5090 and mean level = 6.939.

In the biological processes about 22% of the total EST-SSR sequences with PCR-based primers are associated with genes involving in cellular processes (GO:0009987) like cell communication , which its activation is reported under Olive environmental stresses and fruit development [40], [41]. Also metabolic processes (GO: 0008152) were covered with (21%) of EST-SSR primers ,this processes involves beta-glucosidase, a gene that shaping the phenolic profile of virgin olive oil [42].

Other processes like single-organism process (GO:0044699) which includes genes that enhance the salt tolerance in some plant like CIPKs family [43],localization (GO:0051179), response to stimulus (GO:0050896) has gain 16% , 12% , 8% of ESTs , respectively , while signaling (GO:0023052),

rhythmic processes (GO:0048511) and growth(GO:0040007) are covered with the lowest number of EST-SSR primers.

The molecular function category are covered with SSR primers targeting ESTs associated with catalytic activity (GO: 0003824) (37%), binding (GO: 0005488) (36%) including SEUSS-LIKE genes , which has been reported as transcriptional adaptors regulate the development of flower and embryo [44] and transporter activity (GO: 0005215) (16%) like aquaporin genes.

Cellular components category are assigned by cell (GO: 0005623) 42.7% primers targeting cell membrane genes and organelle (GO: 0043226) (13%) primers for organelle ESTs and macromolecule complex (GO: 0032991) (13%) (Figure 4).

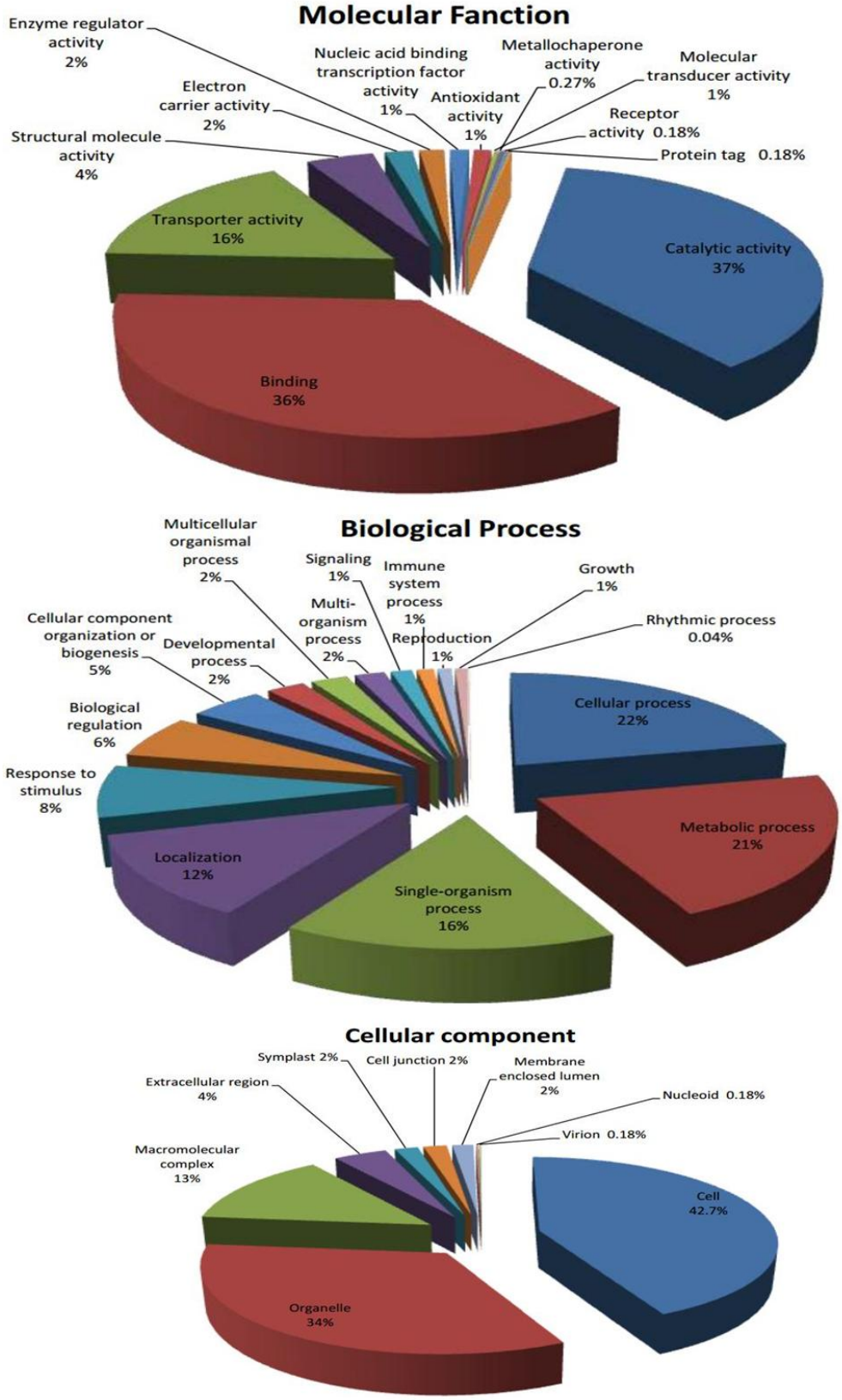

Figure 4: The Categories identified by gene ontology from Olea europaea EST-SSRs sequences

Functional classification by KEGG pathway analyses: The KEGG pathway database is a useful tool for understanding genes biological functions and its molecular interactions [45].To stand on the metabolic pathways that were covered by our EST-SSR primers set , we mapped all EST sequences that contains detectable SSR motifs and has generated valid primers to the KEGG reference pathways. The KEGG pathway analysis revealed that, our EST-SSR primers set covered a total of 92 different pathways and about 256 EST-SSR primers are associated with genes linked to 132 enzymes.

The major pathways were covered with EST-SSR primers using over than nine genes each. These pathways includes starch and sucrose metabolism which is related to depletion of stored carbohydrates (CHO) during the on-year (high yield) and suggested as a cause for alternate bearing in olives [46].Another targeted metabolic pathway is gluconeogenesis which controls the manipulation of non-carbohydrate carbon substrates to glucose [47] , also methionine metabolism which synthesized S-Adenosylmethionine as a donor of the methyl group in DNA methylation for gene expression regulation [48]. These ESTs has a significant match in the KEGG database (Supporting Information Table S2) and these results are visualized by using Circos software [49] (Figure 5).

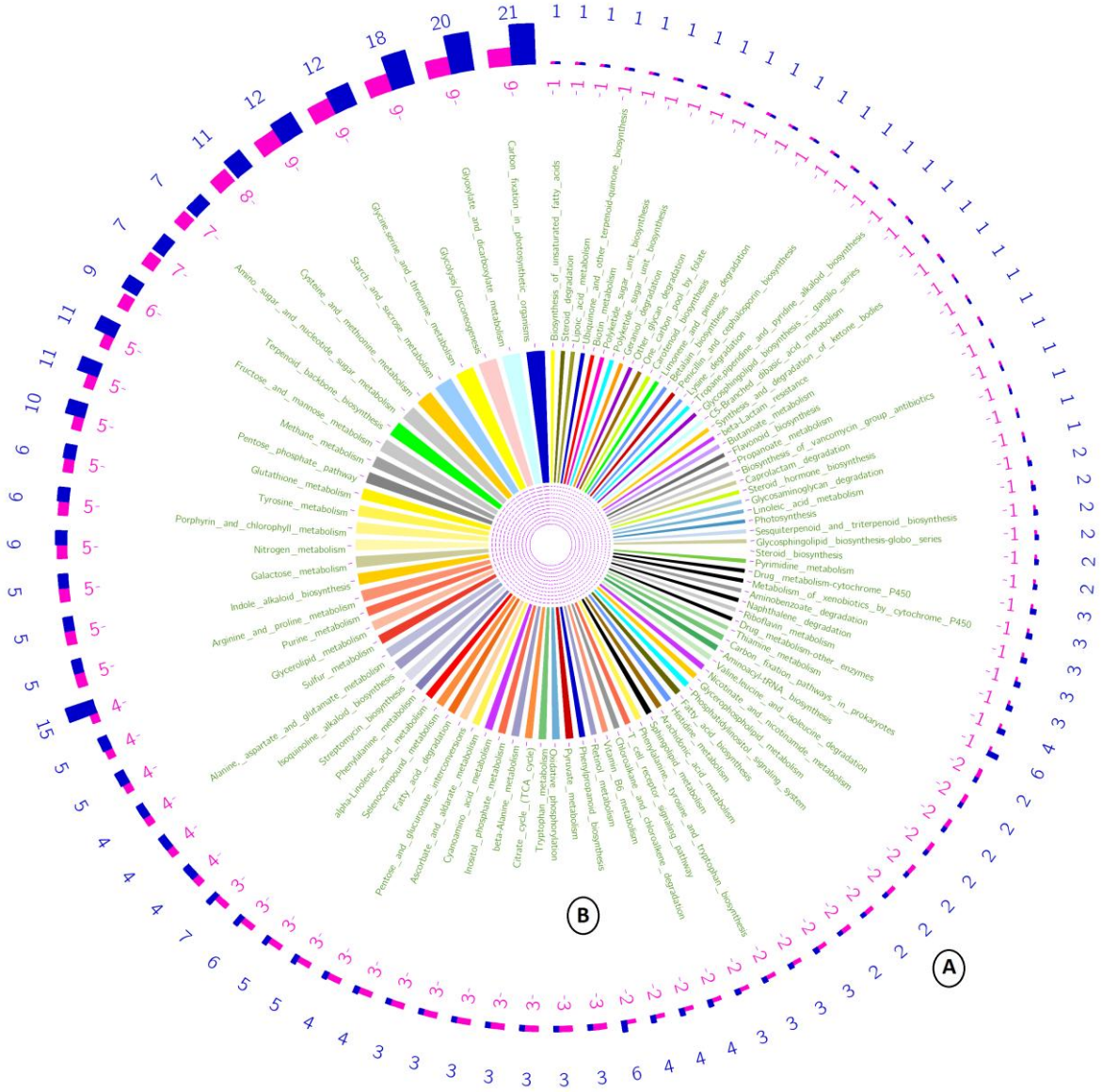

Figure 5. The KEGG analytical pathways for EST sequences contain detectable SSR motifs and generated valid primers. (A) The number of EST-SSR primers in every pathway (blue) , the number of enzymes linked to these primers (red) and the ratio between these two (red/blue chart). (B) The biological pathway title which contains these EST-SSR primers products.

Breeding Olive trees for achieving a higher Olive oil quantity and quality is one of the most important goals for Olive breeding programs worldwide [50], [51].There is a high occurrence of EST-SSR primers in metabolic pathways for enzymes related oil contents indicates a good potential opportunity for using a marker type related to oil traits in olives. The primer-targeted ESTs were categorized by the metabolism it involves in, including lipid metabolism (Table 3), carbohydrate metabolism, energy metabolism, amino acid metabolism, nucleotide metabolism and metabolism of cofactors and vitamins.

Table 3: List of lipid metabolism pathways which have been assigned to EST-SSR sequences targeted with PCR primers.

Pathway	Enzyme name	Enzyme Commission	EST-SSR primer
Fatty acid biosynthesis	Desaturase	ec:1.14.19.2	Oe-ESSR_1411
	synthase I	ec:2.3.1.41	Oe-ESSR_487
Fatty acid degradation	1-monooxygenase	ec:1.14.15.3	Oe-ESSR_1370
	dehydrogenase (NAD+)	ec:1.2.1.3	Oe-ESSR_1089
	Dehydrogenase	ec:1.1.1.1	Oe-ESSR_714,Oe-ESSR_1464,Oe-
Synthesis and degradation of ketone bodies	Synthase	ec:2.3.3.10	Oe-ESSR_1159
Steroid biosynthesis	Synthase	ec:2.5.1.21	Oe-ESSR_1791,Oe-ESSR_568
Steroid hormone biosynthesis	Dehydrogenase	ec:1.1.1.145	Oe-ESSR_1048
Glycerolipid metabolism	dehydrogenase (NAD+)	ec:1.2.1.3	Oe-ESSR_1089
	Lipase	ec:3.1.1.3	Oe-ESSR_1375
	Melibiase	ec:3.2.1.22	Oe-ESSR_197,Oe-ESSR_1318
	dehydrogenase (NADP+)	ec:1.1.1.2	Oe-ESSR_350
Glycerophospholipid metabolism	A1	ec:3.1.1.32	Oe-ESSR_1375
	N-methyltransferase	ec:2.1.1.103	Oe-ESSR_624
Sphingolipid metabolism	lactase (ambiguous)	ec:3.2.1.23	Oe-ESSR_713
	Melibiase	ec:3.2.1.22	Oe-ESSR_197,Oe-ESSR_1318
Arachidonic acid metabolism	Peroxidase	ec:1.11.1.9	Oe-ESSR_990,Oe-ESSR_980
	1-monooxygenase	ec:1.14.15.3	Oe-ESSR_1370
alpha-Linolenic acid metabolism	13S-lipoxygenase	ec:1.13.11.12	Oe-ESSR_726,Oe-ESSR_451
	A1	ec:3.1.1.32	Oe-ESSR_1375
	Dehydrogenase	ec:1.1.1.1	Oe-ESSR_714,Oe-ESSR_1464,Oe-ESSR_1461
Biosynthesis of unsaturated fatty acids	Desaturase	ec:1.14.19.2	Oe-ESSR_1411

In details, the mapping results can further investigated against the glycolysis/gluconeogenesis (Figure 6) and Oxidative Phosphorylation pathways (Figure 7) as an example of carbohydrate metabolism and lipid metabolism respectively.

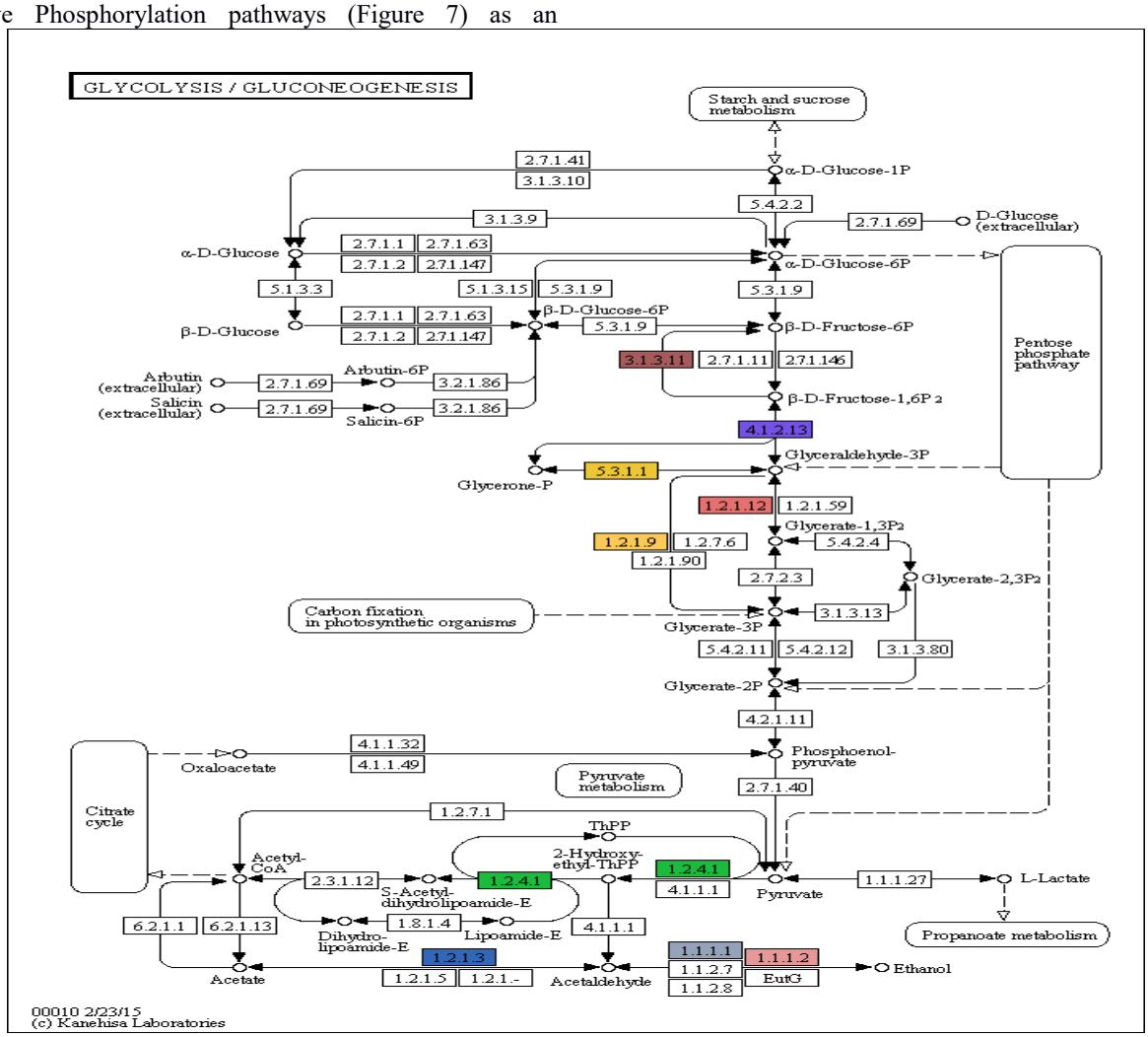

Figure 6: The Distribution of The 18 EST-SSR designed primers targeting 9 enzymes (Enzyme Commission colors) in the glycolysis/gluconeogenesis pathway.

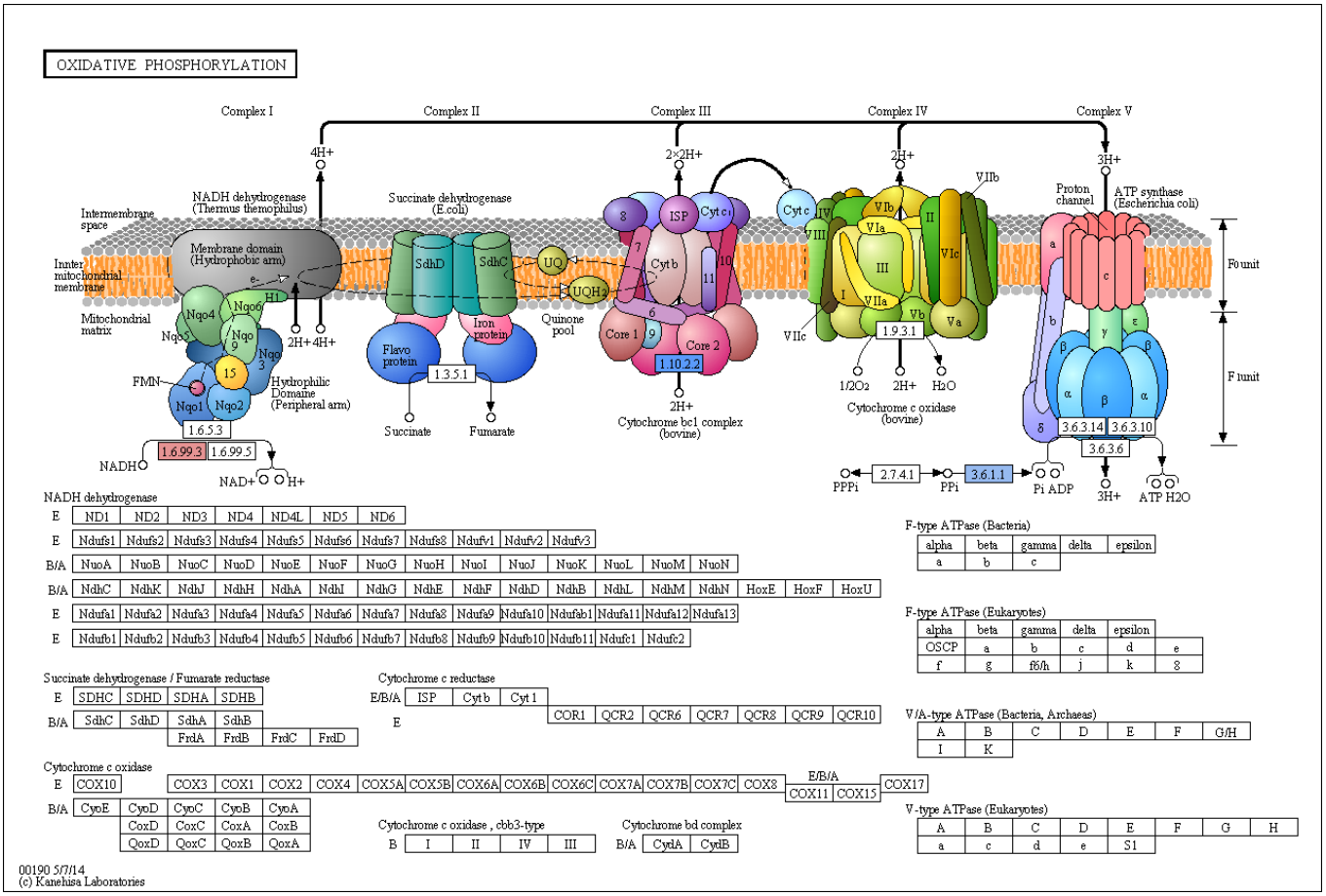

Figure 7: The distribution of the 3 EST-SSR designed primers targeting 3 enzymes (Enzyme Commission colors) in Oxidative Phosphorylation pathway

Olive EST-SSR primers database: All primers were listed in the (Supporting Information Table S3) and was provided with all related information such as primer name, NCBI GI number for the EST sequence which is targeted by this primer , repeat type, repeat sequence, repeat length , repeat start index in the sequence, repeat end index in the sequence, forward and reveres primer pairs, annealing temperature (Tm) (°C), primer length (bp), primer product length (bp), the sequence of the EST, sequence description, gene ontology, enzyme code and enzyme name.

Validation of designed primers: Twenty five primers were randomly selected to validate its efficacy to be used in polymerase chain reaction (PCR) procedures as a reliable molecular marker for marker-assisted selection programs by using a genomic DNA isolated from nine olive cultivars. All tested primers, exhibited successfully amplified and detectable PCR bands and only 10 exhibited detectable polymorphism (Figure 8).

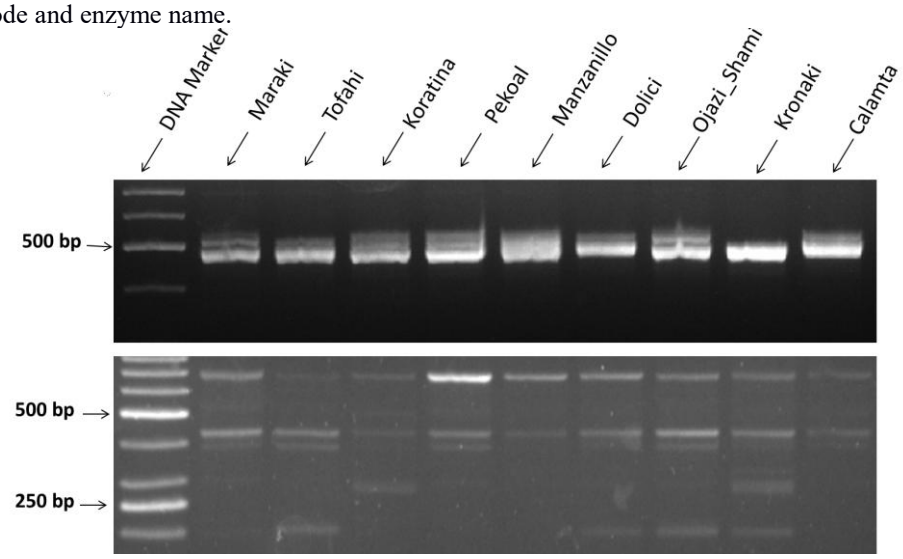

Figure 8: Sample of olive EST-SSR primers tested among nine olive cultivars and that exhibited successfully amplified and detected polymorphism.

4. Conclusion

SSR markers are very important co-dominant, highly polymorphic technique, which can be generated from functional regions in different plant genomes. The EST - SSR technique has the potential to generate prototypically linked functional markers and it is a useful tool could be used in genetic diversity, marker assisted selection and genome mapping in olives. This study exhibits the functional categorization of olive EST sequences containing SSR motifs which can be targeted by a valid set of PCR primers. These ESTs representing genes associates with cellular component, biological process and molecular functions in olives. Also EST-SSR primers could provide useful information to understand the biological functions and gene-gene interactions by taking in advance the localization of these primers in different pathways which has possible relationships with highly important pathways in olive cultivation.

5. Supporting Information

https://drive.google.com/folderview?id=0B1nAji4taBMVOXFrLXVnb3BITEk&usp=sharing

References

[1] D. Zohary, M. Hopf, and E. Weiss, *Domestication of Plants in the Old World: The origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin*. Oxford University Press, 2012.

[2] J. Loureiro, E. Rodriguez, A. Costa, and C. Santos, "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. Resour. Crop Evol.*, vol. 54, pp. 21–25, 2007.

[3] L. Baldoni, N. G. Cultrera, R. Mariotti, C. Ricciolini, S. Arcioni, G. G. Vendramin, A. Buonamici, A. Porceddu, V. Sarri, M. A. Ojeda, and others, "A consensus list of microsatellite markers for olive genotyping," *Mol. Breed.*, vol. 24, no. 3, pp. 213–231, 2009.

[4] R. Ben-Ayed, C. Sans-Grout, F. Moreau, N. Grati-Kamoun, and A. Rebai, "Genetic Similarity Among Tunisian Olive Cultivars and Two Unknown Feral Olive Trees Estimated Through SSR Markers," *Biochem. Genet.*, vol. 52, no. 5–6, pp. 258–268, 2014.

[5] D. Calzada, M. Aguerri, S. Baos, D. Montaner, M. Mata, J. Dopazo, J. Quiralte, F. Florido, C. Lahoz, and B. Cárdaba, "Therapeutic targets for olive pollen allergy defined by gene markers modulated by Ole e 1-derived peptides," *Mol. Immunol.*, vol. 64, no. 2, pp. 252–261, 2015.

[6] C. Newton, J. Terral, and S. Ivorra, "The Egyptian olive (Olea europaea subsp. europaea) in the later first millennium BC: origins and history using the morphometric analysis of olive stones," *ANTIQUITY-OXFORD-*, vol. 80, no. 308, p. 405, 2006.

[7] A. A. Awan, M. Zubair, A. Iqbal, S. Abbas, and N. Ali, "Molecular analysis of genetic diversity in olive cultivars," *African J. Agric. Res.*, vol. 6, no. 21, pp. 4937–4940, 2011.

[8] S. G. Atienza, R. de la Rosa, L. León, A. Martín, and A. Belaj, "Identification of QTL for agronomic traits of importance for olive breeding," *Mol. Breed.*, vol. 34, no. 2, pp. 725–737, 2014.

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

[10] Y. M. Sivolap, "Molecular markers and plant breeding," *Cytol. Genet.*, vol. 47, no. 3, pp. 188–195, 2013.

[11] J. Sahu, P. Sen, M. D. Choudhury, M. Barooah, M. K. Modi, and A. Das Talukdar, "Towards an efficient computational mining approach to identify EST-SSR markers," *Bioinformation*, vol. 8, no. 4, p. 201, 2012.

[12] Y. Kashi and D. G. King, "Simple sequence repeats as advantageous mutators in evolution," *TRENDS Genet.*, vol. 22, no. 5, pp. 253–259, 2006.

[13] W. Powell, M. Morgante, C. Andre, M. Hanafey, J. Vogel, S. Tingey, and A. Rafalski, "The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis," *Mol. Breed.*, vol. 2, no. 3, pp. 225–238, 1996.

[14] A.-F. Adam-Blondon, C. Roux, D. Claux, G. Butterlin, D. Merdinoglu, and P. This, "Mapping 245 SSR markers on the Vitis vinifera genome: a tool for grape genetics," *Theor. Appl. Genet.*, vol. 109, no. 5, pp. 1017–1027, 2004.

[15] F. L. Luro, G. Costantino, J. Terol, X. Argout, T. Allario, P. Wincker, M. Talon, P. Ollitrault, and R. Morillon, "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*, vol. 9, no. 1, p. 287, 2008.

[16] F. Vantini, G. Tacconi, M. Gastaldelli, C. Govoni, E. Tosi, P. Malacrino, R. Bassi, and L. Cattivelli, "Biodiversity of grapevines (Vitis vinifera L.) grown in the Province of Verona," *VITIS-Journal Grapevine Res.*, vol. 42, no. 1, p. 35, 2015.

[17] E. Díaz-Losada, A. T. Salgado, I. Orriols-Fernández, A. M. Ramos-Cabrer, and S. Pereira-Lorenzo, "New synonyms and homonyms for cultivars from Northwestern Spain," *Am. J. Enol. Vitic.*, p. ajev–2012, 2012.

[18] I. Trujillo, M. A. Ojeda, N. M. Urdiroz, and D. Potter, "Identification of the Worldwide Olive Germplasm Bank of Córdoba (Spain) using SSR and morphological markers," 2013.

[19] N. Ozdemir Ozgenturk, F. Oruç, U. Sezerman, A. Kuçukural, S. Vural Korkut, F. Toksoz, and C. Un, "Generation and analysis of expressed sequence tags from Olea europaea L.," *Comp. Funct. Genomics*, vol. 2010, 2010.

[20] R. KAUR, N. SHARMA, and R. RAINA, "Identification and functional annotation of expressed sequence tags based SSR markers of Stevia rebaudiana," *Turkish J. Agric. For.*, p. 39, 2015.

[21] P. K. Gupta and R. K. Varshney, "The development and use of microsatellite markers for genetic analysis and

plant breeding with emphasis on bread wheat," *Euphytica*, vol. 113, no. 3, pp. 163–185, 2000.
[22] E. G. N. Mbanjo, F. Tchoumbougnang, A. S. Mouelle, J. E. Oben, M. Nyine, C. Dochez, M. E. Ferguson, and J. Lorenzen, "Development of expressed sequence tags-simple sequence repeats (EST-SSRs) for Musa and their applicability in authentication of a Musa breeding population," *Afr J Biotech*, vol. 11, pp. 13546–13559, 2012.
[23] B. L. R. I. Naga, L. N. Mangamoori, and S. Subramanyam, "Identification and characterization of EST-SSRs in finger millet (Eleusine coracana (L.) Gaertn.)," *J. Crop Sci. Biotechnol.*, vol. 15, no. 1, pp. 9–16, 2012.
[24] M. Wen, H. Wang, Z. Xia, M. Zou, C. Lu, and W. Wang, "Developmenrt of EST-SSR and genomic-SSR markers to assess genetic diversity in Jatropha Curcas L.," *BMC Res. Notes*, vol. 3, no. 1, p. 42, 2010.
[25] T. Wöhrmann and K. Weising, "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.*, vol. 123, no. 4, pp. 635–647, 2011.
[26] S.-R. Liu, W.-Y. Li, D. Long, C.-G. Hu, and J.-Z. Zhang, "Development and characterization of genomic and expressed SSRs in citrus by genome-wide analysis," 2013.
[27] M. Verma and L. Arya, "Development of EST-SSRs in watermelon (Citrullus lanatus var. lanatus) and their transferability to Cucumis spp.," *J. Hortic. Sci. Biotechnol.*, vol. 83, no. 6, p. 732, 2008.
[28] L. R. Pinto, K. M. Oliveira, E. C. Ulian, A. A. F. Garcia, and A. P. De Souza, "Survey in the sugarcane expressed sequence tag database (SUCEST) for simple sequence repeats," *Genome*, vol. 47, no. 5, pp. 795–804, 2004.
[29] R. K. Varshney, T. Thiel, N. Stein, P. Langridge, and A. Graner, "In silico analysis on frequency and distribution of microsatellites in ESTs of some cereal species," *Cell. Mol. Biol. Lett.*, vol. 7, no. 2A, pp. 537–546, 2002.
[30] A. Untergasser, I. Cutcutache, T. Koressaar, J. Ye, B. C. Faircloth, M. Remm, and S. G. Rozen, "Primer3—new capabilities and interfaces," *Nucleic Acids Res.*, vol. 40, no. 15, pp. e115–e115, 2012.
[31] A. Conesa, S. Götz, J. M. García-Gómez, J. Terol, M. Talón, and M. Robles, "Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research," *Bioinformatics*, vol. 21, no. 18, pp. 3674–3676, 2005.
[32] P. Rajendrakumar, A. K. Biswal, S. M. Balachandran, and R. M. Sundaram, "In silicoAnalysis of Microsatellites in Organellar Genomes of Major Cereals for Understanding Their Phylogenetic Relationships," *In Silico Biol.*, vol. 8, no. 2, pp. 87–104, 2008.
[33] X.-P. Jia, Y.-S. Shi, Y.-C. Song, G.-Y. Wang, T.-Y. Wang, and Y. Li, "Development of EST-SSR in foxtail millet (Setaria italica)," *Genet. Resour. Crop Evol.*, vol. 54, no. 2, pp. 233–236, 2007.
[34] E. Filiz and I. Koc, "In Silico chloroplast SSRs mining of Olea species," *Biodiversitas*, vol. 13, pp. 114–117, 2012.
[35] R. V. Kantety, M. La Rota, D. E. Matthews, and M. E. Sorrells, "Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat," *Plant Mol. Biol.*, vol. 48, pp. 501–510, 2002.
[36] L. C. da Maia, V. Q. de Souza, M. M. Kopp, F. I. F. de Carvalho, and A. C. de Oliveira, "Tandem repeat distribution of gene transcripts in three plant families," *Genet. Mol. Biol.*, vol. 32, no. 4, pp. 822–833, 2009.
[37] B. Lewin and G. Dover, *Genes v*, vol. 299. Oxford University Press Oxford, 1994.
[38] Y. G. Cho, T. Ishii, S. Temnykh, X. Chen, L. Lipovich, S. R. McCOUCH, W. D. Park, N. Ayres, and S. Cartinhour, "Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (Oryza sativa L.)," *Theor. Appl. Genet.*, vol. 100, no. 5, pp. 713–722, 2000.
[39] D. Metzgar, J. Bytof, and C. Wills, "Selection against frameshift mutations limits microsatellite expansion in coding DNA," *Genome Res.*, vol. 10, pp. 72–80, 2000.
[40] S. B. M. Hammami, T. Manrique, and H. F. Rapoport, "Cultivar-based fruit size in olive depends on different tissue and cellular processes throughout growth," *Sci. Hortic. (Amsterdam).*, vol. 130, no. 2, pp. 445–451, 2011.
[41] R. Gucci, E. M. Lodolini, and H. F. Rapoport, "Water deficit-induced changes in mesocarp cellular processes and the relationship between mesocarp and endocarp during olive fruit development," *Tree Physiol.*, p. tpp086, 2009.
[42] C. Romero-Segura, R. García-Rodríguez, A. Sánchez-Ortiz, C. Sanz, and A. G. Pérez, "The role of olive β-glucosidase in shaping the phenolic profile of virgin olive oil," *Food Res. Int.*, vol. 45, no. 1, pp. 191–196, 2012.
[43] D.-G. Hu, Q.-J. Ma, C.-H. Sun, M.-H. Sun, C.-X. You, and Y.-J. Hao, "Overexpression of MdSOS2L1, a CIPK protein kinase, increases the antioxidant metabolites to enhance salt tolerance in apple and tomato," *Physiol. Plant.*, 2015.
[44] F. Bao, S. Azhakanandam, and R. G. Franks, "SEUSS and SEUSS-LIKE transcriptional adaptors regulate floral and embryonic development in Arabidopsis," *Plant Physiol.*, vol. 152, no. 2, pp. 821–836, 2010.
[45] D. Li, Z. Deng, B. Qin, X. Liu, and Z. Men, "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*, vol. 13, no. 1, p. 192, 2012.
[46] A. Bustan, A. Avni, S. Lavee, I. Zipori, Y. Yeselson, A. A. Schaffer, J. Riov, and A. Dag, "Role of carbohydrate reserves in yield production of intensively cultivated oil olive (Olea europaea L.) trees," *Tree Physiol.*, p. tpr036, 2011.
[47] S.-J. S. Sung, D.-P. Xu, C. M. Galloway, and C. C. Black, "A reassessment of glycolysis and gluconeogenesis in higher plants," *Physiol. Plant.*, vol. 72, no. 3, pp. 650–654, 1988.
[48] S. C. Lu, "Sadenosylmethionine," *Int. J. Biochem. Cell Biol.*, vol. 32, no. 4, pp. 391–395, 2000.
[49] M. Krzywinski, J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S. J. Jones, and M. A. Marra, "Circos: An information aesthetic for comparative genomics," *Genome Res.*, vol. 19, pp. 1639–1645, 2009.

[50] M. El Riachy, F. Priego-Capote, L. Rallo, M. D. Luque-de Castro, and L. León, ―Phenolic profile of virgin olive oil from advanced breeding selections,‖ *Spanish J. Agric. Res.*, vol. 10, no. 2, pp. 443–453, 2012.
[51] Y. Ozdemir, N. Aktepe Tangu, and M. E. Akcay, ―Generating omega-3 rich olive oil by cross breeding,‖ *Eur. J. Lipid Sci. Technol.*, vol. 115, no. 9, pp. 977–981, 2013.

doi: http://dx.doi.org/10.1101/026294

Cite this article as: Adawy, Sami S., Morad M. Mokhtar, Alsamman M. Alsamman, and Mahmoud M. Sakr. 2015. ―Development of Annotated EST-SSR Database in Olive (Olea Europaea).‖ International Journal of Science and Research 4(9).

Author Profile

Prof. Sami Said Adawy is the Deputy Director of Agricultural Genetic engineering Research Institute; Agricultural Research Center, Giza, 12619- Egypt. Email address: samiadawy@ageri.sci.eg. Address: 9 Gamaa Street, Giza 12619- Egypt. Fax: (202)35731574

Prof. Mahmoud M. Sakr, is the President of Academy of Scientific Research and Technology (ASRT), Egypt

Mr. Morad Mokhtar Mokhtar, Molecular Markers and Genome Mapping Department; Agricultural Genetic engineering Research Institute; Agricultural Research Center, Giza, 12619- Egypt. Address: 9 Gamaa Street, Giza 12619- Egypt.

Mr. Alsamman Mahmoud Alsamman, Molecular Markers and Genome Mapping Department; Agricultural Genetic engineering Research Institute; Agricultural Research Center, Giza, 12619- Egypt. Address: 9 Gamaa Street, Giza 12619- Egypt.