Generation and Analysis of Expressed Sequence Tags (ESTs) from Olea europaea L.

Nehir Ozdemir Ozgenturk[1] Fatma Oruç [1]Ugur Sezerman[2] Alper Kuçukural[2] Senay Vural Korkut[1], Feriha Toksoz[3], Cemal Un[3]

[1]Yildiz Technical University, Faculty of Science and Arts, Department of Biology. Davutpasa Str. 124, 34210 Merter/Istanbul, TURKEY

[2]Sabanci University, Faculty of Engineering and Natural Sciences, 34956 Tuzla/Istanbul, TURKEY

[3]Ege University, Faculty of Science, Department of Biology. 35100 Bornova/Izmir, Turkey

Corresponding Author:

Assistant Prof. Dr. Nehir Ozdemir Ozgenturk

Yıldız Technical University

Faculty of Science and Arts

Department of Biology

34210 Merter/Istanbul

nehirozdemir@yahoo.com

Abstract

Olive (Olea europaea L.) is an important source of edible oil which was originated in Near-East region. In this study Gemlik, an important olive cultivar for black table olive in Turkey, was used as a plant material to construct cDNA libraries. Two cDNA libraries were constructed from young olive leaves and immature olive fruits for generation of ESTs to discover the novel genes and search the function of unknown genes of olive. The randomly selected 3840 colonies were sequenced for EST collection from both libraries. Readable 2228 sequences for olive leaf and 1506 sequences for olive fruit were assembled into 205 and 69 contigs respectively whereas 2478 were singletons. Putative functions of all 2752 differentially expressed unique sequences were designated by gene homology based on BLAST and annotated using BLAST2GO. While 1339 EST`s show no homology to the database, 2024 ESTs have homology (under 80%) with hypothetical proteins, putative proteins, expressed proteins and unknown proteins in NCBI-GenBank. 635 EST`s unique genes sequence have been identified by over 80% homology to known function in other species which were not previously described in Olea family. Only 3.1 % of total EST's was shown similarity with olive database existing in NCBI. This generated EST's data and consensus sequences were submitted to NCBI as valuable source for functional genome studies of olive.

Introduction

Oleacea family comprises 600 species in 24 genus and disseminates all around the world. The olive *Olea europaea* L, which is one of the first domesticated agricultural tree crops in the family *Oleaceae*, is cultivated mainly for both edible oil and table olives. The domestication of *Olea* europaea is supposed to be realized some 5700–5500 years ago in the Near-East (Zohary and Hopf 1994). Therefore Anatolia is one of the most important areas of the olive origin which over 86 varieties of Europea species present in Turkey (Anatolia). It is known that Olive is native to coastal areas of the Mediterranean region such as Spain, Italy, Greece, France, Turkey, Algeria and Morocco. Olive is the most extensively cultivated fruit crop with its orchards cover about 9,8 mil. ha. in the world. According to the statistics published by FAO, Turkey is the fourth largest producer of olive oil in the world, after Spain, Italy, Greece. Turkey is the first producer of black table olive in the world and Gemlik cuv. represents 80%

of black table olives production in Turkey. Because of economical importance of Gemlik, a lot of research centers in Turkey continue their molecular and classical breeding program for this cultivar.

Most of the genetic studies in cultivated plants are focused on the understanding of genetic mechanisms and improvement of product quality and quantity. With the improvement of DNA sequencing technology, large-scale single-pass cDNA sequencing is commonly used to obtain large expressed sequence tag (EST) collection which is generated with expressed gene at a particular stage and/or tissue of organism. The sequenced cDNA show direct information on the mature transcripts for coding part of the genome, so EST databases are very useful tools for gene and marker discovery, gene mapping and functional studies.

After the completion of the genome projects in different species, the number of ESTs has increased rapidly and become available in databases for further applications. Over 40 plant species EST libraries are currently available providing valuable resource for functional genomics studies (Martienssen, 2000; Yammanoto and Sasaki, 1997; Yu et al., 2002; Van der Hoeven et al., 2002; Moyle et al., 2005; Moser et al., 2005; Grimplet et al., 2005; Newcomb et al., 2006).

By using information from these EST databases the possible functions of many genes can be deduced by homologies to known genes.

Although many molecular markers have been developed in olives (Wiesman et al., 1998; Mekuria et al., 1999; Angiolillo et al., 1999; Besnard et al., 2000; Sefc et al., 2000; Rallo et al., 2000; Belaj et al., 2001; Besnard and Berville 2002; De Caraffa et al., 2002; Cipriani et al., 2002), EST studies for olives are not sufficent. By the end of 2008 around one thousand ESTs were generated for searching development of olive fruits and deposited in NCBI database (Galla et. al. 2009) Before we submit the olive EST collection to database, there were just around 1126 sequences available in GenBank databases (February 2009). In this paper, we report a rich EST collection from two separate cDNA libraries constructed from the fresh germinated leaves and immature olive fruits for Turkish olive cultivar Gemlik. 2304 clones were sequenced from the leaf cDNA library and 1536 clones were sequenced from the fruit cDNA library. After removal of low quality ESTs, generated 3734 high-quality olive ESTs were analyzed by using Phred-Phrap and Contig Assembly Program 3 (CAP3) software

and were submitted to GenBank (dbEST). Annotation is performed by using BLAST and BLAST2GO.

Material and Method

The olive breeding line of *O.europea,* Gemlik cuv. (G 20/1) is used as a plant material research in this study. Plant materials were supplied by The Ataturk Central Horticultural Research Institute (ACHRI).

Library Construction

Total RNA was isolated from 10 g fresh germinated leaves and immature olive fruits with the RNeasy Plant Miniprep kit (Qiagen) and pooled. mRNA was purified from total RNA using the Oligotex Spin-Column Protocol (Oligotex mRNA Mini Kit, Qiagen, Valencia, CA). The mRNAs were pooled and final concentration of mRNA was adjusted to 1-3 µg. Two separate cDNA libraries were established with 1,5 µg and 3 µg mRNA leaf and immature olive fruit, respectively. cDNA libraries were constructed with the CloneMiner cDNA Library Construction Kit according to the manufacturer's instructions (Invitrogen, Carlsbad, CA, USA). Double stranded cDNA was cloned into pDONR222 vector and transformed into E.coli strain DH5 (Invitrogen, Carlsbad, CA, USA). Each cDNA library was plated onto LB-kanamycin agar medium and individual grown clonies were picked into 384-well plates with SOB medium and inoculated overnight. After the addition of glycerol (10% v/v) the library stored at -80℃.

Plasmid DNA Purification and DNA Sequencing

Plasmid DNA was isolated from randomly selected sixty clones with alkaline lysis method (Sambrook et al., 1989, Feliciello and Chinalli 1993). Isolated DNA was digested with Bgl1701 and analyzed by a 1% agarose gel electrophoresis to identify insert size.

Randomly selected 3840 clones were used as template for PCR amplification of the cloned cDNA by M13 universal primers. Automated sequencing was performed on an automated high-throughput pipeline using the ABI 3730 capillary sequencer (PE Applied Biosystems, Foster City,CA) at the Genome Sequencing Center, Washington University in St. Louis (WUSTL).

EST Analysis

EST sequences were trimmed of vector, adapter, and low quality sequence by using Phred software (Ewing and Green 1998; Ewing et al. 1998) (CodonCode Crop., Dedham, MA.) 106 low quality EST sequences were removed with the program Phred (version 3/19/99, default 20. The remaining 3734 EST sequences are reprocessed with 'cross-match' application of Phrap for the vector sequence trimming (Ewing ve Green 1998; Ewing et al., 1998).

Total EST sequences, leaf and fruit EST sequences, were assembled separately into contigs by using Contig Assembly Program 3 (Cap3) (Huang et al., 1992, Huang and Madan, 1999). The default values were used for all the parameters. Also the assembly result was controlled with Consed/Autofinish software (Gordon et. al., 1998, 2001). Plausible functions for the established contigs were designated by gene homology based on BLAST. The biological meaning of the unique sequences was investigated according to gene ontology (GO) terms based on BLAST definitions using the program BLAST2GO which is a comprehensive bioinformatics tool for functional annotation and analysis of gene or protein sequences (Conesa et. al., 2005, Conesa and Gotz ,2007).

RESULT

Quality of cDNA libraries and clustering of ESTs

Two separate, cDNA libraries were constructed from a pool of RNA extracted from young leaves and fruits independently. The insert size distribution ranged from 200 to 2500 bps in the leaf cDNA library which consisted of 2.4×10^6 clones with an average insert length of 1.6 kb. In the immature olive fruit cDNA library the average insert size was 1.1 kb (min 70 bp to max 1500 bp) and the library consisted of 2.2×10^5 clones. After construction of cDNA libraries 2304 clones were sequenced from the leaf library; 1536 clones were sequenced from the fruit library. Consequently a total of 3840 EST sequences was generated. Raw EST sequence data was processed and base-called by using Phred. The Olive EST sequences were trimmed from the start and to the end of the sequences on the basis of trace quality to remove vector, adapter and low-quality bases with the default value of 0.05. After this process, 106

clones were removed and the average length of 3734 ESTs was determined as 874 bp.

For contig assembly, designated 2228 high-quality leaf EST sequences and 1506 high-quality fruit EST sequences were analyzed as individual and total by program CAP3. While assembling the 2228 leaf EST sequences into 205 contigs, length ranged from 514 bases to 1924 bases and the number of EST ranged from 2-33, 1506 fruit EST‟s were assembled in to 69 contig, length ranged from 461 bases to 1909 bases and the number of EST ranged from 2-385 (Table 1). When we assembled two libraries together since there are some common genes expressed in the leaf as well as in the fruit,some of the ESTs obtained from the leaf and fruit established new contigs increasing the total contig number of the assmebled libraries to 299. Some of the singlets of the leaf and fruit libraries established new contigs when the libraries assembled together decreasing the total singlet number of the joint library by 100 to 2368. All 3734 EST sequences and the 249 of high quality consensus sequences were submitted to GenBank (dbEST) and EST‟s can be accessed through the accession numbers GO242703-GO246436. Consensus sequences of olive can be reached on the accession numbers EZ421546-EZ421794.

Table 1: The assembly analysis of EST for two cDNA libraries independently and together by CAP3

	Leaf	**Fruit**	**Total**
Number of EST's	2228	1506	3734
Number of Contig	205	69	299
Number of Singlet	1.591	887	2.368
Average Length of Contigs	2194 bp	1912 bp	2134 bp
Number of EST Range in the Contig	2-33	2-385	2-379

Identification of ESTs' Putative Function

The annotation of the 3734 ESTs were designated by database search algorithms BLASTN for

nucleic acids and the BLASTX for proteins at The National Center for Biotechnology Information (NCBI) web server.

Among the 3734 ESTs, 682 of them (18.2 %) showed significant sequence similarities to putative genes registrated in NCBI with score of $\geq$ 80 bits or e value $\leq 10^{-10}$ according to BLASTN similarity search against the nucleotid collection database ((last verified on July 2010). The 1647 ESTs (44.1%) resulted in some hits but with weak similarity scores ($\leq$ 80-40 bits) out of these 896 ESTs (23.9 %) had a score between 60- 79 bits and 751 ESTs (20.2 %) had a score between 40-59 bits. The 1405 ESTs (37.7 %), which gave very low similarity scores but stil gave some hits (0-39 hits) or gave no hits since they have no similarity to exisiting sequences in the databases, thats why they were classified in the "No hit" category. Some of the low scoring hits may also be considered as no hits as well. But since the algorithms provided some hits we put them into weak similarity match category. BLASTN analysis against the nucleotid collection database between our EST and olea sequences in NCBl database has shown that there are only 116 ESTs have similarities, and 38 % of these (45 ESTs) have 80% or higher homology (with the score of $\geq$ 80 bits). 96.9 % of the ESTs generated by us in these studies are different than the ones in olive sequences database already presented by NCBI. On the other hand, with BLASTN analysis against EST database only 81 EST have similarities to olea ESTs in NCBI, and 29 % of these have 80% or higher homology (with the score of $\geq$ 80 bits).

According to the BLASTN result, 13 different total contigs sequences have similarities with *Olea Europaea* EST sequences in GenBank Table 2. These are; specifically those acting on the CH-OH group of donor with NAD+ or NADP+ as acceptor from oxidoreductases family “mannitol dehydrogenase1”, polypeptide that was employed the phases involved in photosystem II “photosystem II 10 kDa polypeptide mRNA”, “glycolate oxidase-like FMN-binding domain protein mRNA”, responsible for the shuttling of phospholipids and other fatty acid groups between cell membranes also able to bind acyl groups “plant lipid transfer protein mRNA”, most commonly known by the shorter name RuBisCO, is an enzyme that is used in the Calvin cycle to catalyze the first major step of carbon fixation, a process by which the atoms of atmospheric carbon dioxide are made available to organisms in the form of energy-rich molecules such as sucrose “ribulose-1,5-bisphosphate carboxylase/oxygenase activase mRNA”, enzyme that acts upon β1->4 bonds linking two glucose or glucose-substituted

molecules “beta-glucosidase (bglc) mRNA”, vacuolar membrane protein in plants “tonoplast intrinsic protein (tip) mRNA”, to transmit signals between cells and binding large family of proteins “polyubiquitin OUB2 mRNA”, some sequences previously identified in olive and a protein that is involved in gluconeogenesis, the synthesis of glucose from smaller molecules “glyoxisomal malate dehydrogenase mRNA”.

Table 2: Homolog genes with *Olea Europaea* consensus EST sequences in GenBank

Contig name	**Homology of *Olea europea* in NCBI data base**	**Query coverage (%)**	**Max. Ident. (%)**	**Length of Contig (bp)**	**The number of EST in the contig**
Contig 7	Olea europaea putative mannitol dehydrogenase 1 (MTD1) mRNA	56	86	895	2
Contig 14	Olea europaea photosystem II 10 kDa polypeptide mRNA, partial cds	30	99	724	4
Contig 24	Olea europaea putative glycolate oxidase-like FMN-binding domain protein mRNA	22	99	2819	9
Contig 85	Olea europaea putative plant lipid transfer protein mRNA	25	100	914	5
Contig 93	Olea europaea Cu/Zn super-oxide dismutase (ole e 5 allergen)	79	93	891	2
Contig 98	Olea europaea putative cytochrome P450 mRNA, partial cds	28	99	1756	8
Contig 111	Olea europaea putative ribulose-1,5-bisphosphate carboxylase/oxygenase activase mRNA, partial cds	43	98	1797	22
Contig 137	Olea europaea subsp. europaea beta-glucosidase (bglc) mRNA, complete cds	93	98	2018	13
Contig 155	Olea europaea tonoplast intrinsic protein (tip) mRNA, complete cds	87	86	911	2
Contig 157	Olea europaea polyubiquitin OUB2 mRNA, complete cds	89	91	1393	7

Contig 169	Olea europaea cultivar Bianchera tRNA-His (trnH) gene, partial sequence; trnH-psbA intergenic spacer, complete sequence; PSII 32 kDa protein (psbA) gene, complete cds; psbA-trnK intergenic spacer, complete sequence; and tRNA-Lys (trnK) gene, partial sequence; chloroplast	97	96	1225	2
Contig 201	Olea europaea putative glyoxisomal malate dehydrogenase mRNA, partial cds	46	97	1383	2
Contig 255	Olea europaea putative metallophosphatase/diphosphonucleotide phosphatase 1 mRNA, partial cds	28	96	969	2

In addition to BLAST results, Gene ontology (GO) annotations of the leaf, fruit and all contig sequences of *Olea Europea* L. cv. Gemlik were performed by using Blast2GO. The software performed BLASTX similarity search against the GenBank non-redundant protein database, retrieved GO terms for the top 20 BLAST results and annotated the sequences based on default criteria (Conesa et. al., 2005, Conesa and Gotz ,2007). GO terms were distributed among the Biological Process, Molecular Function and Cellular Component categories (Table 3).

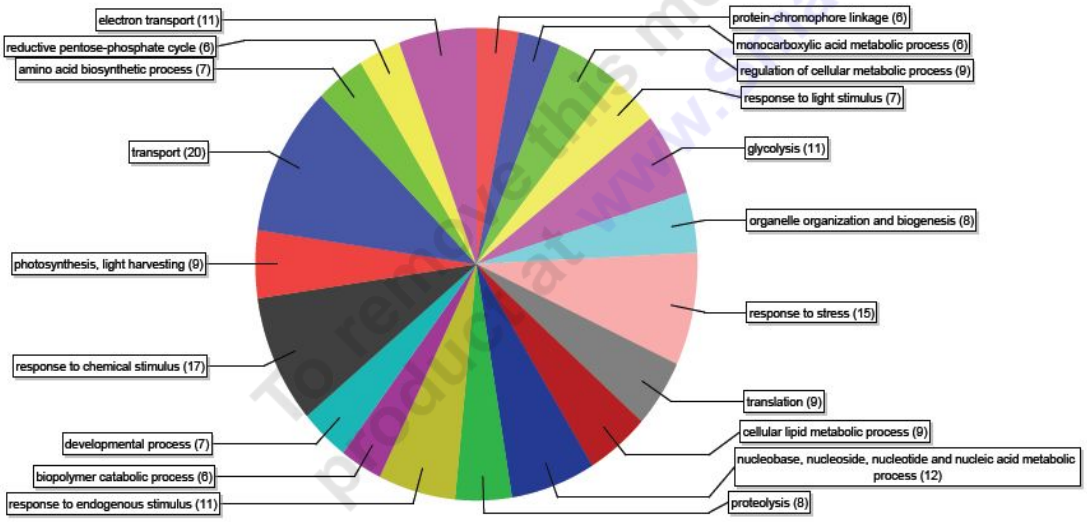
Table 3.: Gene Ontology results of leaf, fruit and total contigs with the program of BLAST2GO

	Molecular Function/ Number of Contig (existent percentage)	**Cellular Component / Number of Contig (existent percentage)**	**Biological Process / Number of Contig (existent percentage)**
Leaf **(Total 205 Contig)**	**1)** Protein binding /24 (11,7%) **2)** ATP binding /13 **3)** DNA binding /9 **4)** Structural molecule activity /9 **5)** Iron ion binding/ 9 **6)** Peptidase activity /9 **7)** Nucleoside-triphosphatase activity/ 8 **8)** Carbon carbon lyase activity /7 **9)** Hydrolase activity, acting on ester bonds/ 7 **10)** GTP binding /7	**1)** Integral to membrane /15 **2)** Photosystem II /15 **3)** Mitochondrion /14 **4)** Cytoplasmic membrane-bounded vesicle /8 **5)** Nucleus /8 **6)** Photosystem I /8 **7)** Chloroplast stroma /6	**1)** Transport /20 (9,7%) **2)** Response to chemical stimulus /17 **3)** Response to stres /15 **4)** Nucleobase, nucleoside, nucleotide and nucleic asit metabolic proses /12 **5)** Glycolysis /11 **6)** Response to endogenous stimulus /11 **7)** Electron transport /11 **8)** Cellular lipid metabolic process /9 **9)** Translation /9 **10)** Regulation of cellular metabolic process /9

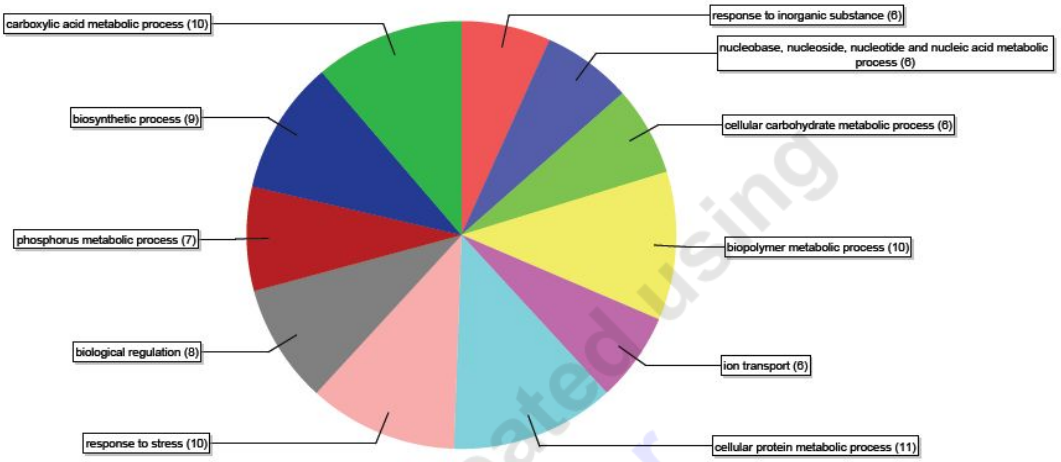
	11) Magnesium ion binding /7 **12)** Coenzyme binding /6 **13)** Transferase activity transfering acyl groups /6 **14)** Chlorophyl binding /6 **15)** Electron carrier activity /6 **16)** Zinc ion binding /6 **17)** Oxidoreductase activity acting on CH-OH 7group of donors /6 **18)** Transferase activity transfering phosphorus containing groups /6 **19)** Transmembrane transporter activity /6 **20)** Isomerase activity /5	**8)** Cytosol /6 **9)** Chloroplast thylakoid membrane /6 **10)** Ribosome /6 **11)** Peroxisome /6	**11)** Photosynthesis, light harvesting /9 **12)** Organelle organization and biogenesis /8 **13)** Proteolysis /8 **14)** Amino acid biosynthetic process /7 **15)** Developmental process /7 **16)** Response to light stimulus /7 **17)** Protein-chromophore linkage /6 **18)** Monocarboxylic acid metabolic process /6
Fruit **(Total 69 Contig)**	**1)** Hydrolase activity /9 (13 %) **2)** Transferase /8 (11,5%) **3)** Metal ion binding /8 (11,5%) **4)** Ion transmembrane transporter activity /6 **5)** Antiporter activity /6 **6)** Oxidoreductase activity /6 **7)** Cation binding /6 **8)** Nucleotide binding /6	**1)** Mitochondrion /6 **2)** Integral to membrane /6 **3)** Vacuolar membrane /5 **4)** Chloroplast /4 **5)** Plastid /4 **6)** Membrane /3 **7)** Nucleus /2 **8)** Cytoplasm /2 **9)** Golgi aparatus oxygen evolving complex /1 **10)** Microtubulle /1 **11)** Cytosolic small ribosomal subunit /1	**1)** Cellular protein metabolic process /11 (15,4%) **2)** Carboxylic acid metabolic process /10 (14,4%) **3)** Response to stres /10(14,4%) **4)** Biopolymer metabolic process /10 (14,4%) **5)** Biosynthetic process /9 (13%) **6)** Biological regulation /8 (11,5%) **7)** Phosphorus metabolic process /7 (10.1%) **8)** Nucleobase, nucleoside, nucleotide and nucleic asit metabolic proses /6 **9)** Ion transport /6 **10)** Cellular carbohydrate metabolic process /6 **11)** Rresponse to inorganic substance /6
3734 EST **(Total 299 Contig)**	**1)** ATP binding /19 **2)** DNA binding /11 **3)** Zinc ion binding /11 **4)** Iron ion binding /10 **5)** Structural constituent ribosome /9 **6)** Hydrolase activity, acting on ester bonds /9 **7)** Nucleoside-triphosphatase activity /9 **8)** Carbon carbon lyase activity	**1)** Mitochondrion /23 **2)** Integral to membrane /22 **3)** Photosystem II /16 **4)** Cytoplasmic membrane-bounded vesicle /14 **5)** Nucleus /12 **6)** Ribosome /10	**1)** Translation /14 **2)** Electron transport /13 **3)** Glycolysis /12 **4)** Organelle organization and biogenesis /12 **5)** Response to endogenous stimulus /11 **6)** Cellular lipid metabolic process /11 **7)** Photosynthesis, light harvesting /10 **8)** Proteolysis /10 **9)** Protein folding /9

	/9 **9)** GTP binding /8 **10)** Carbon transmembrane transporter activity /8 **11)** Ligase activity /8 **12)** Calcium ion binding /8 **13)** Magnesium ion binding /8 **14)** Coenzyme binding /8 **15)** Isomerase activity /8 **16)** Kinase activity /7 **17)** Electron carrier activity /7 **18)** Chlorophyl binding /7 **19)** Antiporter activity /7 **20)** Endopeptidase activity /6 **21)** Oxidoreductase activity, acting on the aldehyde or oxo group of donors /6 **22)** Phosphotransferase activity, alchole groups as acceptor /6 **23)** Transferase activity transfering acyl groups /6 **24)** Unfolded protein binding /5 **25)** Oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor /5	**7)** Photosystem I /9 **8)** Chloroplast stroma /7 **9)** Chloroplast thylakoid membrane /7 **10)** Cytosolic part /7 **11)** Endomembrane system /6 **12)** Cytoskeleton /6 **13)** Vacuolar membrane /6 **14)** Peroxisome /6	**10)** Response to salt stress /8 **11)** Coenzyme metabolic process /8 **12)** Lipid biosynthetic process /7 **13)** Phosphorylation /7 **14)** Response to cold stress /7 **15)** Response to light stimulus /7 **16)** Developmental process /7 **17)** Protein-chromophore linkage /7 **18)** Amino acid biosynthetic process /7 **19)** Reductive pentose-phosphate cycle /7 **20)** Monocarboxylic acid metabolic process /6 **21)** Biopolymer biosynthetic process /6 **22)** Response to oxidative stres /6 **23)** Protein catabolic process /6 **24)** Pesponse to metal ion /6 **25)** Cellular di-,tri-valent inorganic cation homeostasis /6 **26)** Metal ion transport /6 **27)** RNA metabolic process /6 **28)** Secondary metabolic process /6 **29)** Regulation of transcription /5 **30)** Establishment of cellular localization /5

A



B



C

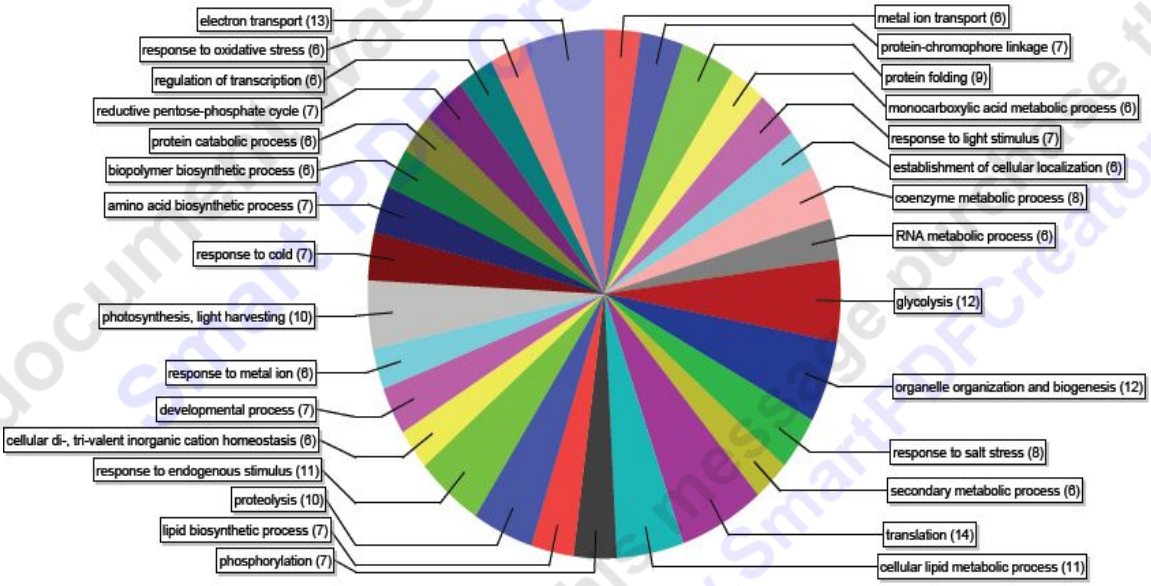


Figure 1: GO terms distribution in the biological process show with circle graphs for leaf (A), fruit (B) and total contigs (C)

20 different types of molecular functions were found for 162 leaf contigs by Blast2GO program. Also Blast2GO results showed that 47 fruit contigs have 8 different molecular function as GO terms, and the contigs that were prepared from all ESTs have 25 different types of molecular functions in 205 contigs. The common molecular function GO terms for all three results are “hydrolase activity”, “transferase activity”, “transmembrane transporter activity”, “oxidoreductase activity” and “ion binding”. Most of the assigned functional class

(11,7%) is binding proteins for the sequences obtained from the leaves. Fruit contigs also have binding proteins as functional class but not as common as leaf contigs. All molecular function results from revealed BLAST2GO program are shown in Table 3.

The biological process category refers to a biological objective to which a gene contributes, but does not identify pathways. Biological process results are identified by BLAST2GO program like molecular function results. Results are similar for all three contig groups. Especially "carboxylic acid metabolic process", "biosynthetic process", "response to stress", "transport", "biopolymer metabolic process", "nucleobase, nucleoside, nucleotide and nucleic asit metabolic process" are common for all three results. But there were a lot of different GO terms for biological process results. For instance, in fruit contigs "phosphorus metabolic process", "biological regulation", "cellular carbohydrate metabolic process", "cellular protein metabolic process" and "response to inorganic substance" GO terms weren't seen in leaf contigs. Some of GO terms like "response to chemical stimulus", "response to endogenous stimulus", "cellular lipid metabolic process", "glycolysis", "proteolysis" and "protein-chromophore linkage" weren't seen in fruit contigs. All the observed differences and similarities between contig groups are summarized in Table 3. When in Figure 1 the biological process which is most observed for leaf in GO terms are transport, response to chemical stimulus, response to stress, in total contigs, GO terms of translation, electron transport, glycolysis, and in fruit, cellular protein metabolic process, carboxylic acid metabolic process, response to stres are the most observed ones. Facing different GO terms in total contigs depend on the fact that the different sequences among the leaf and fruit contigs do form new consensus sequences.

The final GO term category identifies the locations in the cell where the gene products are found. The Olea europaea gene products were found generally associated with the cellular components, in the intracellular space or in organelles such as the mitochondrion, cytoskeleton, vacuolar membrane, peroxisome, ribosome. Despite the fact that the most represented GO terms for cellular components of all contigs are integral to membrane and mitochondrion, in the meantime, as expected photosystem II has also been most observed GO term for the leaf.

DISCUSSION

The EST's give very remarkable information about gene expression patterns at a certain stage of the organism. ESTs have been used for gene discovery (Schmitt et al 1999; Lee et al., 2005) tissue- or stage- specific gene expression (Audic and Claverie 1997) and alternative splicing (Gupta et al., 2004). In this project, we aimed to obtain more information about olive genome, and we have planned to produce a large EST collection for *Olea Europea* L. which has limited number of ESTs in databases. In order to achieve this goal of creating a larger and richer collection, we have constructed two different cDNA libraries from leaves, and fruits for increasing our chance to capture different genes.

According to BLASTN result we have observed some common putative genes between leaves and fruit contigs assembled by CAP3 such as reductase, cytochrome P450, GDP-mannose-3',5'-epimerase (GME), tubulin, ascorbate peroxidase, beta-glucosidase, polyubiquitin, aldolase-like protein, ubiquitin, chlorophyll a/b binding protein. Among the assembled leaves contigs some specific putative genes were observed such as asparagine synthetase (AS), germacrene D synthase, desacetoxyvindoline 4-hydroxylase-like (D4H), plastid transketolase 1, ABC transporter family protein, glutamate synthase 1, chloroplast ferredoxin I, glyceraldehyde-3-phosphate dehydrogenase, chlorophyll a/b-binding protein, malate dehydrogenase, alcohol dehydrogenase, mannitol dehydrogenase 1. Equally among the assembled fruit contigs have some different putative genes than leaves such as SDH2-1, UDP-glucuronate decarboxylase 3, cytoplasmic ribosomal protein, aspartic protease, S-RNase-binding protein, chloroplast oxygen-evolving protein, elongation factor 1 alpha subunit, myb-related transcription factor, Tic20-like protein, Ca2+ antiporter/cation exchanger. Since less than 10% of olive genes were tagged in each tissue in this study some of the GO terms occuring on one tissue and not on the other tissue could be due to the less representative ESTs obtained or sampling variation and may not infer to tissue specific genes.

On the other hand, the Blast2GO analysis of assembled EST‟s enabled the identification of GO terms on three different categories, such as molecular function, biological process and cellular location. While the leaf contigs gave hits on 20 different functional classes and fruit contigs gave hits on 8 functional classes, but contigs obtained from the combined library yielded in hits on 25 functional classes, some of them were not observed in functional classes obtained from the leaf and fruit libraries alone. This may be the result of new contigs generated by the combination of the libraries which are giving hits to genes belonging to new functional classes which maybe expressed both in the leaf and the fruit tissues.

It has been the widest olive genome EST collection of *Olea Europea* L. cv. Gemlik which was constructed to the date. The number of ESTs of *Olea europea* is 4860 in NCBI (last verified on May 2010), and 3734 out of this figure were generated within this study. This project has dramatically increased the number of Olive ESTs in NCBI GenBank database which is a very useful source for the scientists working on olive genome or on comparative genome researches. For further researches, more ESTs should be generated, and be annotated in order to increase the identified number of expressed olive genes for functional analysis.

ACKNOWLEDGMENTS

This study was supported by the Scientific and Technological Research Council of Turkey (TUBİTAK), Project No. 104T146.

REFERENCES

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

Audic, S. ve Claverie J. The significance of digital gene expression profiles. Genome Res.1997; 7,986-995.

Belaj A, Trujillo I, De la Rosa R, Rallo L. Polymorphism and discrimination capacity of Randomly Amplified Polymorphic markers in an Olive Germplasm Bank. J Am Soc Hortic Sci. 2001;126:64–71

Besnard G, Green PS, Bervillé A. The genus *Olea*: molecular approaches of its structure and relationships to other *Oleaceae*. Acta Bot Gallica, 2002a; 149 (1):49-66

Besnard, G, and Berville A. On chloroplast DNA variations in the olive (*Olea europaea* L.) complex: comparison of RFLP and PCR polymorphisms. Theor. Appl. Genet. 2002;104: 1157–1163.

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

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 2005; 21(18):3674-3676.

Conesa A and Gotz S. Blast2GO: A Comprehensive Suite for Functional Analysis in Plant Genomics. International Journal of Plant Genomics, 2007; Volume 2008, Article ID 619832,

12 pages doi:10.1155/2008/619832., 2007.

De Caraffa, VB, Maury J, Gambotti C, Breton C, Berville A, and Giannettini J. Mitochondrial DNA variation and RAPD mark oleasters, olive and feral olive from Western and Eastern Mediterranean. Theor. Appl. Genet. 2002;104: 1209–1216.

Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 1998; 8(3):175-185.

Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.* 1998; 8(3):186-194.

Feliciello I and Chinali G. A modified alkaline lysis method for the preparation of highly purified plasmid DNA from Escherichia coli. Anal. Biochem. 1993; 212: 394–401.

Galla G, Barcaccia G, Ramina A, Collani S, Alagna F, Baldoni L, Cultrera N GM, Martinelli F, Sebastiani L, Tonutti P. Computational annotation og genes differentially expressed along olive fruit development. BMC Plant Biology. 2009; 9:128

Gordon D, Abajian C, Green P. Consed: a graphical tool for sequence finishing. Genome Research. 1998; 8:195-202.

Gordon D, Desmarais C, Green P. Automated finishing with Autofinish. Genome Res. 2001; 11:614-625.

Grimplet J, Romieu C, Audergon J-M, Marty I, Albagnac G, Lambert P, Bouchet J-P, Terrier N. Transcriptomic study of apricot fruit (Prunus armeniaca) ripening among 13,006 expressed sequence tags. Physiol Plant. 2005;125: 281–292

Gupta S, Zink D, Korn B, Vingron M and Haas SA. Genome wide identification and classification of alternative splicing based on EST data. Bioinformatics. 2004;**20:** 2579–2585

Huang X. A contig assembly program based on sensitive detection of fragment overlaps. *Genomics.* 1992; **14**:18–25

Huang X, Madan A. CAP3: a DNA sequence assembly program. GenomeRes. 1999;9:868–77.

Lee Y, Tsai J, Sunkara S et. al. The TIGR Gene Indices: clustering and assembling EST and known genes and integration with eukaryotic genomes. Nucleic Acids Res. 2005;33:D71–4.

Martienssen RA. Weeding out the genes: The *Arabidopsis* genome project. Funct. Integr. Genomics 2000; 1:2–11

Mekuria GT, Collins GG and Sedgley M. Genetic variability between different accessions of some common commercial olive cultivars. J. Hortic. Sci. Biotechnol. 1999; 74: 309–314.

Moser C, Segala C, Fontana P, Salakhudtinov I, Gatto P, Pindo M, Zyprian E, Toepfer R, Grando MS, Velasco R. Comparative analysis of expressed sequence tags from different organs of Vitis vinifera L. Funct Integr Genomics 2005; 5: 208–217

376

377 Moyle R, Fairbairn DJ, Ripi J, Crowe M, Botella JR. Developing pineapple fruit has a small
378 transcriptome dominated by metallothionine. J Exp Bot. 2005; 56: 101–112

379 Newcom RD, Crowhurs RN, Gleave AP, Rikkerink EHA, Allan AC, Beuning LL, Bowen JH,
380 Gera E, Jamieson KR, Janssen BJ, Laing WA, McCartney S, Nain B, Ross GS, Snowden KC,
381 Soulyere EJF, Walton EF, Yauk Y-K. Analyses of expressed sequence tags from apple. Plant
382 Physiol 2006; 141:147–166

383 Rallo P, Dorado G, Martin A. Development of simple sequence repeats (SSRs) in olive tree
384 (Olea europaea L.). Theor. Apopl. Genet. 2000; 101 984-989.
385
386 Sambrook J, Fritsch EF,Maniatis T. Molecular Cloning: A Laboratory Manual. Cold Spring
387 Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

388 Schmitt AO, Specht T, Beckmann G, et. al. Exhaustive mining of EST libraries for genes
389 differentially expressed in normal and tumour tissues. Nucleic Acids Res. 1999; 27:4251–60.

390 Van der Hoeven R, Ronning C, Giovannoni J, Tanksley MG. Deductions about the number,
391 organization, and evolution of genes in the tomato genome based on analysis of a larger
392 expressed sequence tag collection and selective genomic sequencing. Plant Cell. 2002; 14:
393 1441–1456

394 Wiesman Z, Avidan N, Lavee S, Quebedeaux B. Molecular characterization of common olive
395 varieties in Israel and the West Bank using randomly amplified polymorphic DNA (RAPD)
396 markers. J Am Soc Hortic Sci. 1998; 123:837–841

397 Yammanoto K, and Sasaki T. Large-scale EST sequencing in rice. Plant Mol. Biol. 1997;
398 35:135–144.

399 Yu,J. et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science. 2002;
400 296:79–92.

401 Zohary M and Hopf M. *Domestication of Plants in the Old World,* 2d ed. (Oxford: Clarendon,
402 1994).