

**YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

**OLEA EUROPAEA cv GEMLİK (ZEYTİN) BİTKİSİNE AİT
cDNA KÜTÜPHANESİNİN HAZIRLANMASI VE
SONUÇLARIN BİOİNFORMATİK ANALİZ YÖNTEMLERİ
İLE DEĞERLENDİRİLMESİ**

Biyolog Fatma ORUÇ

**F B E Biyoloji Anabilim Dalında
Hazırlanan**

YÜKSEK LİSANS TEZİ

Tez Danışmanı: Yrd. Doç. Dr. Nehir ÖZDEMİR

İSTANBUL, 2009

İÇİNDEKİLER

SİMGE LİSTESİ.....	iv
KISALTMA LİSTESİ	v
ŞEKİL LİSTESİ.....	vii
ÇİZELGE LİSTESİ	viii
ÖNSÖZ	ix
ÖZET	x
ABSTRACT.....	xi
1. GİRİŞ	12
1.1 BİYOİNFORMATİK.....	14
1.2 cDNA KÜTÜPHANELERİ.....	16
1.3 EXPRESSED SEQUENCES TAGS.....	18
1.3.1 EST'lerin Biyoinformatik için iyi bir kaynak olması	20
1.3.2 Ham EST'lerin İşlenmesi ve Birleştirilmesi	23
1.4 GenBank	25
1.4.1 GenBank Kayıtları Ve Kısımları	25
1.4.2 Dizi belirleyicileri ve Erişim numaraları	26
1.4.3 EST'ler ve NCBI.....	26
2. MATERYAL VE METOD.....	28
2.1 MATERYAL	28
2.1.1 Bitki Materyali	28
2.1.2 Kullanılan Kitler	28
2.1.3 Programlar.....	28
2.1.3.1 Phred Ve Phrap	28
2.1.3.2 CAP3.....	29
2.1.3.3 Consed / Autofinish	29
2.1.3.4 BLAST2GO	29
2.1.3.5 VecScreen	29
2.1.3.6 Sequin	30
2.2 METOD	30
2.2.1 RNA İzolasyonu.....	30
2.2.2 mRNA İzolasyonu	30
2.2.3 cDNA Kütüphanesinin Oluşturulması	30
2.2.4 Kütüphanenin Karakterizasyonu.....	32
2.2.5 Vektör Uzaklaştırma	32
2.2.6 Contig Oluşturma.....	33
2.2.7 EST Dizilerinin Kontrolü.....	36
2.2.8 Blast Yöntemi İle Fonksiyonların Belirlenmesi.....	38
2.2.9 Sonuçların GenBank Veritabanına Gönderilmesi.....	40
3. SONUÇLAR	42
3.1 CDNA KÜTÜPHANELERİNİN KARAKTERİZASYON SONUÇLARI.....	42

3.2	BİYOİNFORMATİK ANALİZLERİ	
3.2.1	BLAST Analizi Sonuçları.....	45
3.2.2	Gen Ontoloji Çalışmaları Sonuçları.....	52
3.2.2.1	Yaprak Dokusuna Ait Contiglerin BLAST Çalışması Sonuçları.....	52
3.2.2.2	Zeytin (Meyve) Dokusuna Ait Contiglerin BLAST Çalışması Sonuçları	60
3.2.2.3	Kütüphanedeki Bütün EST' ler (Toplam) İçin Yapılan Biyoinformatik Analizlerden Elde Edilen Contiglerin BLAST Çalışması Sonuçları	67
3.3	SONUÇLARIN GENBANK VERİTABANINA KAYDEDİLMESİ.....	76
4.	TARTIŞMA	77
4.1	EST ANALİZLERİ.....	77
4.1.1	Est Analiz Programları.....	79
4.2	EST ÖN İŞLEME	80
4.3	EST BİRLEŞTİRME	84
4.4	EST FONKSİYONLARININ ANALİZLERİ	88
4.4.1	Fonksiyon Analizlerinin Sonuçlarının İncelenmesi.....	90
4.4.1.1	Blast Sonuçlarının İncelenmesi.....	90
4.4.1.2	Gen Ontoloji Sonuçlarının İncelenmesi.....	95
4.4.1.3	Housekeeping İncelemesi	98
	KAYNAKLAR	101
	EKLER.....	112
	ÖZGEÇMİŞ	155

SİMGE LİSTESİ

oC	Santigrat derece
g	gram
L	litre
μ	mikro
μ l	mikrolitre
μ mol	mikromol
μ g	mikrogram
mg	miligram
ml	mililitre
%	Yüzde oranı
kDA	kilo Dalton
Mb	Mega baz
Kb	Kilo baz

KISALTIMA LİSTESİ

ATP	Adenosine triphosphate
BAC	Bakteri yapay kromozom vektörü (Bacterial Artificial Chromosome)
BLAST	Dizi homolojisi arama motoru (Basic Local Alignment Search Tool)
Bsp1407I	Restriksiyon Enzimi
CAP3	EST birleştirme programı
CDD	Conserved Domain Veritabanı
cDNA	Tamamlayıcı DNA dizisi (complementary Deoxyribonucleic acid)
COG	Orthologous Gruplarının Kümelemesi
cv	Kültivar (cultivar)
dbEST	EST veritabanı (Database EST)
dC	Sitozin nükleotidleri dizisi
DDJB	Japon DNA veritabanı (DNA Data Bank of Japan)
dG	Guanin nükleotidleri dizisi
DNA	Deoksiribonükleik Asit (Deoxyribonucleic acid)
dNTP	Deoksiribonükleotid Trifosfat (Deoxyribonucleotide triphosphate)
dT	Timin nükleotidleri dizisi
EMBL	Avrupa Moleküler Biyoloji Laboratuvarı (European Molecular Biology Laboratory)
EST	Anlatımı yapılmış dizi işaretleri (Expressed sequence tag)
e-value	BLAST için bir parametre (Expect value)
FMN	Flavin mononucleotid
FTP	Dosya aktarım iletişim kuralı (File Transfer Protocol)
GSS	Genom haritalama dizileri (Genome Survey Sequence)
GO	Gen Ontoloji
GTP	Guanosine triphosphate
HTG	Yüksek verimli genomik veriler (High Throughput Genomic)
KEGG	Kyoto Encyclopedia of Genes and Genomes
LINE	Long interspersed nuclear elements
LLR skor	Phrap programının kullandığı bir skor (Log Likelihood Ratio)
LTP	Lipid Transfer Protein (yağ transferi proteinleri)
LTR	Long Terminal Repeat
M13	PCR primeri
mRNA	Mesajcı Ribonükleik Asit (Messenger Ribonucleic acid)
NAD	Nicotinamide adenine dinucleotide
NADP	Nicotinamide adenine dinucleotide phosphate
NCBI	Ulusal Biyoteknoloji Merkezi (National Center for Biotechnology Information)
NID	GenBank'da nükleotid dizilerini tanımak için her bir diziye verilen numara (Nucleotide ID)
NIH	Ulusal Sağlık Enstitüsü (National Institutes of Health)
NLM	Ulusal Tıp Kütüphanesi (National Library of Medicine)
ORF	Açık Okuma Bölgesi (Open Reading Frame)
PCR	Polimeraz Zincir Reaksiyonu (Polymerase Chain Reaction)
pDONR TM 222	Klonlama vektörü
PEP	Phosphoenolpyruvic acid
RNA	Ribonükleik Asit (Ribonucleic acid)
RNaz	RNA'yı parçalayan enzim
S1	Bir tür nükleaz

SAGE	Gen anlatımının seri analizi yöntemi (Serial Analysis of Gene Expression)
SINE	Short interspersed nuclear elements
SSR	Simple Sequence Repeats
STS	Dizi işaret bölgeleri (Sequence tagged site)
TIGR	EST birleştirme programı
UCSC	Universty of California Santa Cruz
USPTO	Amerika Birleşik Devletleri Patent ve Marka Ofisi (United States Patent and Trademark Office)
WUSTL	Genom Dizileme Merkezi

ŞEKİL LİSTESİ

Şekil 1.1: EST analizindeki genel aşamalar.....	21
Şekil 2.1.: cDNA kütüphanesinin oluşturulma aşamaları.....	31
Şekil 2.2.: pDONR vektör.....	32
Şekil 2.3.: 5' ve 3' kırpma bölgelerinin f dizisinde hesaplanması.....	35
Şekil 3.1.: Plazmid DNA'ların jelde görünümü	42
Şekil 3.2.: EST analizindeki genel aşamaları	43
Şekil 3.3.: Consed sonuçları	45
Şekil 3.4.: Meyve dokusuna ait 17. contigin NCBI web sitesinde BLASTN analizi sonucu...47	
Şekil 3.5.: BLAST2GO default değerleri	51
Şekil 3.6.: Yaprak için Species distribution.....	54
Şekil 3.7.: CAP3 programından elde edilen yaprak dokusuna ait contig dizilerinin BLASTX sonuçlarına göre GenBanktaki diğer nükleotid dizileri ile olan benzerlik oranı55	
Şekil 3.8.: Yaprak dokusuna ait contiglerin moleküler fonksiyonlarına göre sınıflandırılması.57	
Şekil 3.9.: Yaprak dokusuna ait contiglerin hücre bileşenlerine göre sınıflandırılması	58
Şekil 3.10.: Yaprak dokusuna ait contiglerin biyolojik proseslerine göre sınıflandırılması....	60
Şekil 3.11.: Meyve için Species distribution	62
Şekil 3.12.: CAP3 programından elde edilen meyve dokusuna ait contig dizilerinin BLAST sonuçlarına göre GenBanktaki diğer nükleotid dizileri ile olan benzerlik oranı62	
Şekil 3.14.: Meyve dokusuna ait contiglerin moleküler fonksiyonlarına göre sınıflandırılması64	
Şekil 3.15.: Meyve dokusuna ait contiglerin hücre bileşenlerine göre sınıflandırılması.....	65
Şekil 3.16.: Meyve dokusuna ait contiglerin biyolojik prosese göre sınıflandırılması.....	67
Şekil 3.17.: Species distribution: CAP3 programından kütüphaneye ait contig dizilerinin BLASTX sonuçlarına göre GenBanktaki diğer organizmalarla olan benzerliği69	
Şekil 3.18.: CAP3 programından elde edilen kütüphaneye ait contig dizilerinin BLAST sonuçlarına göre GenBanktaki diğer nükleotid dizileri ile olan benzerlik oranı70	
Şekil 3.19.: Kütüphaneye ait contiglerin moleküler fonksiyonlarına göre sınıflandırılması....	72
Şekil 3.20.: Kütüphaneye ait contiglerin hücre bileşenlerine göre sınıflandırılması.....	73
Şekil 3.21.: Kütüphaneye ait contiglerin biyolojik proseslerine göre sınıflandırılması	76
Şekil 4.1.: Bir EST dizisi genellikle vektör kontaminasyonu, bölünemeyen tekrar olasılıkları veya düşük kaliteli karmaşık bölgeler ile başlar ve biter (Nagaraj ve ark., 2006).....	77
Şekil 4.2.: Phred quality skorları.....	81

ÇİZELGE LİSTESİ

Çizelge 3.1.: PHRAP programı ile biyoinformatik analizleri sonucu kütüphaneden elde edilen değerler.....	44
Çizelge 3.2.: CAP3 programı ile biyoinformatik analizleri sonucu kütüphaneden elde edilen değerler.....	44
Çizelge 3.3.: Meyve dokusuna ait EST'lerden CAP3 programı ile üretilen 69 contigin http://www.ncbi.nlm.nih.gov/blast/Blast.cgi?PAGE=Nucleotides web sitesinde yapılan BLASTN analizlerinin sonuç çizelgesi.	48
Çizelge 3.4.: CAP3 programından elde edilen yaprak dokusuna ait contig dizilerinin BLAST sonuçlarına göre GenBanktaki zeytin dizileri ile benzerlik gösteren contigler	53
Çizelge 3.5.: BLAST sonuçlarına göre GenBanktaki zeytin dizileri ile benzerlik gösteren zeytin (meyve) dokusuna ait contigler	61
Çizelge 3.6.: BLAST sonuçlarına göre GenBanktaki zeytin dizileri ile benzerlik gösteren kütüphaneye ait contigler	68
Çizelge 4.1.: EST analiz aşamalarının karakteristiği.....	80
Çizelge 4.2.: EST ön işleme için kaynaklar	83
Çizelge 4.3.: EST dizilerini birleştirmek ve contig oluşturmak için kullanılan programlar	85
Çizelge 4.4.: EST / cDNA dizilerinin genom DNA dizileri ile eşleştirilmesi için programlar	89
Çizelge 4.5.: Phrap ve CAP3 programlarıyla elde edilen contiglerin karşılaştırılmasından örnekler. İki program için ortak olan fonksiyonlar.	90
Çizelge 4.6.: Phrap ve CAP3 programlarıyla elde edilen contiglerin karşılaştırılmasından örnekler. İki program arasındaki farklı fonksiyonlar.	91
Çizelge 4.7.: Phrap ve CAP3 programlarıyla elde edilen contiglerin farklı dokular bakımından karşılaştırılmasından örnekler.	93
Çizelge 4.8.: Gen Ontoloji sonuçları	97
Çizelge 4.9.: Populus housekeeping genleri ile ortak fonksiyona sahip contig dizileri	100
Çizelge Ek 1: İşlenmiş " <i>Olea europaea L. cv. Gemlik</i> " EST dizilerinin GenBank'a kaydedilmesinden sonra GenBank veritabanından alınan "accession numaraları"	112
Çizelge Ek 2: Yaprak dokusuna ait contig dizilerinden her birinin içerdiği yaprak EST'lerinin adı ve accession numaraları.....	113
Çizelge Ek 3: Meyve dokusuna ait contig dizilerinden her birinin içerdiği meyve EST'lerinin adı ve accession numaraları.....	135

ÖNSÖZ

Tez çalışmamın her aşamasında beni yönlendiren, yardım eden, her konuda desteğini esirgemeyen, bu çalışmayla kazandıklarımı borçlu olduğum değerli hocam, tez danışmanım Sayın Yrd. Doç. Dr. Nehir ÖZDEMİR'e;

Tezimin hazırlanmasında bilmediğim pek çok şeyi öğrenme fırsatını, üstesinden gelmekte zorlandığım problemlerin çözümünü ve ihtiyacım olan desteği sağlayan, hoşgörüsünü, sabrını ve emeğini esirgemeyen Sabancı Üniversitesi Mühendislik ve Doğa Bilimleri Fakültesi Biyomühendislik Bölümü öğretim üyesi Sayın Doç. Dr. Uğur SEZERMAN'a;

Çalışmalarım sırasında bilgilerini ve zamanlarını büyük bir yardımseverlikle bağışlayan, tezim için çok önemli katkıları olan Sabancı Üniversitesi Mühendislik ve Doğa Bilimleri Fakültesi Biyomühendislik Bölümün'den Alper KÜÇÜKURAL ve Cem MEYDAN'a;

Çalışmalarımı yakından izleyen, beni motive eden, yüksek lisans eğitimimde önemli bir yeri olan değerli hocam, Yrd. Doç. Dr. Şenay VURAL KORKUT'a;

Bana olan inancını sık sık belirten değerli hocam Sayın Yrd. Doç. Dr. Cemal ÜN'e;

Yıldız Teknik Üniversitesi, Fen-Edebiyat Fakültesi Biyoloji Bölümü başkanı Prof. Dr. Nezhun GÖREN'e bana sağladığı bölüm olanaklarından dolayı;

Hayatımın her döneminde olduğu gibi tezimin hazırlanma aşamasındaki bu zorlu süreçte de ilgi ve desteklerini yanımda hissettiğim sevgili dostlarıma;

Tüm hayatım boyunca hiçbir fedakârlıktan kaçınmayan, maddi ve manevi desteklerini esirgemeyerek beni bugünlere getiren biricik annem Nimet ORUÇ ve babam Ömer Faruk ORUÇ'a;

Ve buraya isimlerini yazamasam da hayatımın her anında bana inanan ve destek veren tüm herkese;

Sonsuz saygı ve teşekkürlerimi sunarım.

Mayıs 2009

Fatma ORUÇ

ÖZET

Anavatanı Akdeniz havzası olan Zeytin (*Olea europaea L. cv. Gemlik*) bitkisi ekonomik açıdan ülkemiz için önemli olan tarım ürünleri arasında yer almaktadır. Bu çalışmada zeytin bitkisinden alınan yaprak ve meyve örnekleri kullanılarak iki ayrı cDNA kütüphanesi yeni gen keşfi ve gen fonksiyon analizleri için önemli olan EST'leri oluşturmak için hazırlanmıştır. Kurulan cDNA kütüphanelerinden tesadüfi seçilen 3840 adet klon örneklerinin dizi analizleri yapılarak zeytin için EST (Expressed Sequence Tags) koleksiyonu oluşturulmuştur.

Zeytin bitkisine ait ham EST verilerinin analiz edilmesi için Phred/ Phrap/ Consed bilgisayar programları ile düşük kaliteli diziler ve vektör fragmentleri uzaklaştırılmıştır. Fragment eşleştirme işlemi için hem Phrap hem CAP3 programı ayrı ayrı kullanılarak yapılmış ve sonuçlar karşılaştırılmıştır. Karşılaştırma için Consed programı her iki programın sonucu kontrol edilerek değerlendirilmiştir. CAP3 programına ait sonuçlar daha uygun bulunduğu için çalışmaya bu verilerle devam edilmiştir. CAP3 programından elde edilen veriler NCBI'a (National Center for Biotechnology Information) ait GenBank veritabanıyla BLASTN uygulaması kullanılarak karşılaştırılmıştır ve GenBank'a kaydedilmiştir.

Anahtar Kelimeler: Zeytin (*Olea europaea L. cv. Gemlik*), EST, cDNA kütüphanesi, Biyoinformatik, Phred/ Phrap/ Consed, CAP3

ABSTRACT

Olive (*Olea europaea L.*) is a very important agriculture product for our country which was originated in Mediterenian region. In this study we prepared a cDNA library from young olive leaves and fruits for generation of ESTs (Expressed Sequence Tags) which are useful for the discovery of novel genes and searching for the function of unknown genes. 3840 colonies which are randomly selected from the libraries were sequenced for obtained olive ESTs collection.

For analysis of the raw EST data in olive genom, low-quality sequences and the vector fragment was removed with Phred / Phrap / Consed computer programmes. Fragment assembly was done with the CAP3 and Phrap softwares and results were compared with each other and evaluated with the Consed program. According comparasion two program, CAP3 was accepted as a reliable source and their result was acceptapted to continue further analysis. The data obtained from the CAP3 program were compared to the GenBank database of NCBI using BLASTN tool and submitted to NCBI data base. The results showed that new information about olive genom and also putative function of the sequences.

Keywords: Olive (*Olea europaea L. cv. Gemlik*), EST, cDNA library, Bioinformatik, Phred/ Phrap/ Consed, CAP3

1. GİRİŞ

Biyoinformatik; matematik, bilgisayar ve biyolojik bilimlerin bileşiminden oluşmuş, 1990'lı yıllarda bilgisayar teknolojilerinin ve genom çalışmalarının gelişmesi ve değerlendirilecek bilginin artması sonucunda önemi artmış bir bilim dalı olmuştur.

Genom çalışmaları sonucunda elde edilen gen dizilerinin, bilgisayar ortamına aktarılmasıyla birlikte oluşturulan veritabanları sadece dizilerle ilgili değil aynı zamanda dizilerin izole edildiği organizmalar, oluşturduğu protein yapıları ve fonksiyonları hakkında da bilgi içerdiği için önemli bir kaynak olmaktadır. Bu veritabanlarından en büyüğü olan NCBI (National Center for Biotechnology Information) ait veritabanları sadece kendi içindeki kayıtlı bilgilerle değil, farklı veritabanlarından da gelen günlük bilgilerle de her gün güncellenmektedir. Laboratuvar çalışmalarından elde edilen sonuçların anlamlı hale gelmesini sağlayacak bilgisayar programları kullanılarak oluşturulan veriler, bu veritabanlarının bir fonksiyonu olan BLAST (Dizi homolojisi arama motoru) gibi basit ve hızlı yöntemler ile daha önce çalışılmış organizmalara ait verilerle karşılaştırılabilir. Aynı çalışmaları bilgisayar programları ve veritabanları olmadan yapmaya çalışmak nerdeyse imkansız olduğu için veritabanlarının varlığı genom araştırmacıları için vazgeçilmez bir kaynak olmuştur.

Ülkemizde yürütülen genom çalışmalarının artmaya başlamasıyla birlikte biyoinformatik kaynaklarının araştırmacılar için doğru ve kolay bir şekilde kullanımının artması beklenmektedir. Ancak biyoinformatik bilim dalının yeni gelişiyor olması ve bu alanda ülkemizde çok az sayıda bilim insanının çalışıyor olması, bu konuda çalışma yapmak isteyen araştırmacıları zorlamaktadır. Bu çalışmalar için gerekli teknik bilgilerin eksikliği karşılaşılan ilk sorunlardan biri olması yanı sıra multidisipliner alanda organizasyon ve iletişim zorlukları bulunmaktadır. Uygun program tespiti ve bu programa ulaşmak, kullanımına hakim olmak karşılaşılan diğer sorunlardır. Bu çalışmanın bu alanda çalışmak isteyen araştırmacılara kaynak oluşturması beklenmektedir.

Bu çalışmada anavatanı Akdeniz ülkeleri olan 658.000 hektar alanda ve 95 milyon ağaç varlığı ile ülkemizdeki toplam tarım alanında % 2.5, bağ-bahçe alanında %20'lik bir paya sahip zeytin bitkisi (*Olea europaea L. cv. Gemlik*) ile çalışılmıştır. Zeytin bitkisi 37 ilimizde yetiştiriciliği yapılmakta ve doğrudan geçim kaynağı olmaktadır. Türkiye dünya

zeytin üretiminin % 10'unu karşılarken siyah zeytin üretiminde birinci, sofralık zeytin üretiminde ikinci ve zeytinyağı üretiminde dördüncü sırada yer alan önemli bir ülke konumunda bulunmaktadır [7]. Türkiye'de üretilen sofralık zeytinin %85'i siyah, % 15'i yeşil ve rengi dönük olarak işlenmektedir. Bu veriler Türkiye'de en fazla siyah sofralık zeytin üretilen çeşit olarak da Gemlik çeşidinin ilk sırayı aldığını ifade etmektedir. Türkiye ihracatının %80'nini karşılayan Marmara Bölgesi'nin önemli çeşitlerinden olan Gemlik zeytini, dünyada kabul gören kalite özelliklerine sahip bir çeşittir (Tunalıoğlu , 2003). Taşıdığı ticari önemle beraber zeytin bitkisinin genomu hakkında bilinenlerin az olması bu çalışmanın çıkış noktasını oluşturmuştur.

Zeytin bitkisinin mensubu olduğu *Oleaceae* familyası, 25 cins ve 500-600 kadar türe sahiptir. Zeytinin dahil olduğu *Olea* cinsine ait 30 kadar tür, Akdeniz havzası ve çevresinde geniş bir yayılım göstermektedir (Besnard ve ark , 2002) . Bunlardan *Olea europaea*'nın kromozom sayısı $2n = 46$ olup, iki varyetesi bulunmaktadır (Baldoni ve ark., 2006). Bu varyeteler; "*Olea europaea L.subp. oleaster*" ve "*Olea europaea L.subp. europaea*"'dır. Yapılan çekirdek , plastid ve mitokondrial DNA analizleri ve parmakizi profillemeye çalışmaları *Olea europaea L.subp. europaea*'nın Anadolu da delice olarak bilinen *Oleaster*'den kültüre edildiğini göstermektedir (Besnard ve ark., 2001;2002 ; Koçkar ve İlkçı, 2003 ; Elbaum ve ark., 2006).

Bu çalışmada biyoinformatik analizlerden yararlanılarak zeytin genleri ile ilgili yeni bilgilere ulaşmak amacıyla laboratuarda zeytin yaprak ve meyve dokularına ait iki farklı cDNA kütüphanesi hazırlandı. cDNA, ters transkriptaz enzimi ile kalıp olarak mRNA kullanılarak sentezlenen komplementer DNA'dır. Genomik DNA'nın tam tersine, cDNA sadece ekson adı verilen eksprese olmuş bölgeleri içermesinden dolayı cDNA'lar sadece anlatım yapan dizi bölgelerinden oluşur, bu avantaj bize gen bölgelerine kolayca ulaşma imkanı sağlar.

Bir organizmaya, bir dokuya veya bir organa ait normal şartlarda veya belirli şartlar altında elde edilmiş çok sayıda cDNA klonları ile cDNA kütüphaneleri oluşturulur ve genom çalışmalarında gen fonksiyonlarına ulaşmak için kullanılırlar. Bu kütüphanelerden elde edilen klonların dizi analizleri sonucu ortaya çıkan anlatımı yapılmış dizi işaretleri EST' ler (Expressed Sequence Tags), bilgisayar ortamına fonksiyonel gen dizileri ve parçaları olarak aktarılarak çeşitli biyoinformatik analizler için veri olarak kullanılmaktadır. Böylece cDNA kütüphanelerinden elde edilen EST' ler bilinmeyen

genlerin tanımlanması ve yerlerinin tespit edilmesinde, gen ekspresyonu ve regülasyonu hakkında veri toplanması ve genom haritalarının oluşturulması konusunda önemli bir veri sağlarken, biyoinformatik alanına ait kaynaklar ve analiz teknikleri kullanılarak hızlı, güvenilir bir şekilde istenilen sonuçlara ulaşılmaktadır.

Günümüzde çok sayıda genom projelerinin tamamlanması ile birlikte EST'lerin sayısı hızla artmıştır ve daha fazla uygulama için veritabanında kullanılabilir hale gelmiştir. Bilinen önemli EST koleksiyonlarından bazıları; Arabidopsis thaliana (L.) (Martienssen, 2000), pirinç (Oryza sativa L.; Yamamoto ve Sasaki, 1997; Yu ve ark., 2002), domates (Lycopersicon esculentum; Van der Hoeven ve ark., 2002), ananas (Ananas comosus; Moyle ve ark., 2005), üzüm (Vitis vinifera; Moser ve ark., 2005), kayısı (Prunus armeniaca; Grimlet ve ark., 2005) ve elma (Malus domestica; Newcomb ve ark., 2006) gibi farklı bitki türleri için üretilmiştir. Bu EST veritabanlarındaki bilgilerin kullanımı sonucu bilinen genlerin homolojisi ile pek çok genin olası fonksiyonlarının tahmini mümkün olmaktadır.

Zeytinde EST koleksiyonu oluşturulması çalışmalarının başında zeytin için pek çok moleküler markır geliştirilmiştir. (Wiesman ve ark., 1998; Mekuria ve ark., 1999; Angiolillo ve ark., 1999; Besnard ve ark., 2000; Sefc ve ark., 2000; Rallo ve ark., 2000; Belaj ve ark., 2001; Besnard ve Berville 2002; De Caraffa ve ark., 2002; Cipriani ve ark., 2002). Bu çalışmada veritabanındaki zeytin genlerinin miktarını arttırmada etkili olabilecek zeytin yaprakları ve meyve taneleri için üretilen iki adet cDNA kütüphanesinden elde edilmiş zengin bir EST koleksiyonu oluşturulmuştur.

1.1 BİYOİNFORMATİK

Biyoinformatik; biyoloji, bilgisayar bilimleri ve bilgi teknolojilerinin birleşiminden oluşmuş tek bir disiplindir. Bu disiplin yeni biyolojik çalışma anlayışlarının oluşmasını ve temelini biyoloji biliminden alan yeni bakış açılarının yaratılmasını hedeflemektedir. Primrose ve Twyman (2006) biyoinformatiği gen anlatımı bilgilerinin toplanması, analizi ve yorumlanması ile ilgili bir biyoloji branşı olarak tanımlamıştır.

Biyoinformatik 1990'ların ortalarında ortaya çıkmış olan genomik devrimle göze çarpan bilimsel bir disiplin olmuştur. Pek çok dizinin, yapının ve biyokimyasal verinin birikmesi ve bunların doğru ve karşılaştırmalı olarak incelenebilmesi ihtiyacı bu bilimin doğmasına sebep olmuştur. Biyoinformatik nükleotid ve aminoasit dizileri, protein domainleri ve protein yapıları gibi çeşitli tipteki verilerin analizini ve yorumlanmasını da içeren bir şekilde genişlemiştir [3].

Bununla birlikte biyoinformatiğin kökleri 1960'lı yıllarda Margaret Dayhoff tarafından oluşturulan protein dizilerine ait ilk veritabanının yapıldığı zamana da dayanmaktadır. Bu ilk veritabanı protein dizilerinin ve yapılarının atlası olarak adlandırılmıştır. (Dayhoff ve ark. , 1965)

Fakat bazı kaynaklar 1951 ile 1955 yılları arasında Frederick Sanger ve arkadaşlarının ilk kez tam olarak bir protein dizisini belirlemesini biyoinformatiğin başlangıcı olarak kabul edebilmektedir. Tam olarak belirlenen ilk protein dizisi sığır insülinidir. (Ryle ve ark. , 1955).

Bugün bir DNA dizisini belirlemek bir protein dizisini belirlemekten daha kolay fakat 1977 yılından önce DNA dizileme için güvenilir bir yöntemin elde edilmediği zamanlarda durum bunun tam tersiydi.

İlk nükleotid dizisinin belirlenmesinin ardından (Madison ve ark. , 1966) 1970'lerin sonu ve 1980'lerin başında, DNA dizilerinin sayısı literatürlerde artmaya başlamıştı. Dizilenmiş genlerin translasyonu ile tahmini protein dizileri de ortaya çıkmaktaydı. Bu yüzden karakterize olmamış protein sayısı giderek artmaktaydı. 1982' de ilk nükleotid dizi veritabanı Genbank' ı oluşturmaya yetecek kadar DNA dizisi vardı. 1982 yılının sonunda Genbank 606 diziyi kapsamaktaydı. 1994 yılına kadar Genbank düzenli bir şekilde büyüdü. Genomik çalışmaların odak noktası haline gelmesi ile genomik klonların ve EST' lerin sayısının hızla artması Genbank' ında hızla büyümesine neden oldu.

1994'de Genbank' daki dizi sayısı yalnızca 200.000 iken 10 yıl sonra bu sayı 30 milyona ulaşmıştır (Benson ve ark. , 2004). Günümüzde Genbank' ta yaklaşık 82 milyon dizi bulunmaktadır.

Biyoinformatik biliminin genom hakkında daha fazla bilgi sahibi olunması için amacı; dizi bilgisinin elde edilmesi ve veri grupları arasındaki ilişkiyi tanımlamaktır. Veri grupları genellikle nükleotid ve protein dizilerini, protein yapılarını, gen anlatım profillerini moleküler ağırlıkları, biyokimyasal veya metabolik sentez yollarını kapsar.

Dizilere ait veritabanları; organize olmuş bir şekilde verileri sunan, genellikle veri girişi ve sistemin bileşenlerine ulaşmak için birleştirilmiş bilgisayar programları ile tasarlanmış geniş yapılardır. Basit bir veritabanı kayıtların her biri için aynı tür bilgileri taşıyan tek bir dosya gibi olmalıdır. Herhangi bir kayıt; nükleotid dizisi, “BAĞLANTI ADI (Contact_Info)” altında bağlantı bilgileri, giriş dizisi ile molekülün ne tip olduğuna dair tanım, kaynak olarak kullanılan organizmanın bilimsel adı ve literatür bilgilerini içermesi gerekir.

Günümüzde en yaygın kullanılan veri tabanı NCBI’ da tek bir sorgulamada birden çok veritabanına ulaşılmasını sağlayan Enterez sistemi bulunmaktadır. Enterez kullanıcıya spesifik bir bilgi için tek bir veritabanından değil, bünyesindeki tüm veritabanlarından cevap verir. Örneğin; Enterez protein veritabanı Enterez taksonomi veritabanı ile çapraz bağlıdır. Bu durum araştırmacının protein dizisinden türe ait taksonomik bilgilere ulaşmasını sağlar [3]. Böylece araştırmacıların herhangi bir bilgiye ait detaylara hızla ulaşması bu yolla mümkün olur.

1.2 cDNA KÜTÜPHANELERİ

Karşılaştırmalı genom çalışmalarında gen dizilerini çözmek genlerin fonksiyonlarını bulmak için önemli bir adımdır. Büyük bir genomda sadece genlerin kodlandığı bölgelere ulaşmak için cDNA sentezinin yapılması araştırmacıların hedefe ulaşmasını kolaylaştıran bir yöntemdir. cDNA, hücredeki mRNA’dan revers transkriptaz enzimi ile oluşturulur. Çekirdekten ayrılıp sitoplazmaya ulaşan intronları uzaklaşmış olgun mRNA’nın karşılığı olarak oluşturulan DNA zincirine cDNA (complementary DNA) denir. Bu süreç mRNA’nın kalıp olarak kullanılmasından çift zincirli DNA’nın sentezlenmesine kadar uzanır (Primrose ve Twyman, 2006). Böylece bütün organizma genomunda anlatımı yapılan genleri tespit etmenin en temel yolu olan mRNA izolasyonunda, mRNA’nın hücre dışında kararsız bir yapıda olması sorunu mRNA’ların daha kararlı bir bileşik olan cDNA

olarak çevrilmesi ile aşılmış olur.Oluşturulan cDNA lardan kütüphaneler kurulur ve buradaki cDNAların nükleotid dizi bilgisinin çözülmesiyle EST ler oluşturulur. EST ler bütün bir organizma için yada organizmaların istenilen hücre, doku veya organları için oluşturularak eksprese olmuş genlere ulaşılmasını sağlar [4].

cDNA kütüphanesi ile ilgili ilk çalışma 1970 ortalarında yapıldı (Homopolimer kuyruk tekniği ile cDNA). Pekçok cDNA kütüphanesi kurma yöntemleri içinden en popüler olanı 1976'da Maniatis ve arkadaşları tarafından kullanılan yöntemdir. Bu yöntemde oligo dT kuyruğu kullanılır. Oligo dT ler poliadenilat kuyruğu ile komplementerdir. Bu şekilde oluşan cDNA ipliği geçici olarak ikinci cDNA ipliğini oluşturmak için saç tokası yapısını oluşturma eğilimlidir (Efstratiadis ve ark., 1976). İkinci cDNA ipliğinin sentezinden sonra spesifik bir nukleaz olan S1 nukleazı gibi bir nukleaz cDNA'nın klonlama vektörüne girmesini sağlar.

Saç tokası yönteminin en belirgin dezavantajı S1 nukleazının 5' ucunda yaptığı kesimlerin kayıplar oluşturabilmesidir. Saç tokası yöntemi ikinci ipliği parça parça sentezleyen yöntemlerin temelidir (Land ve ark., 1981). Bu yöntemlere göre; oligo dT bağlanmış, ilk cDNA ipliği sentezlendikten sonra oligo dT'nin ucuna dNTP'ler eklenerek yeni iplik sentezlenirken terminal transferaz enzimi kullanılır. İkinci ipliğin sonuna ise sentetik oligo dC kuyruğu ilave edilir. Oligo dC kuyruğu sentetik oligo dG ile komplementer olması için eklenmiştir. Böylece tekrar yeni bir cDNA ipliği sentezlenebilecektir. Bu stratejiyi ilk kez kullanan Land ve ark., 1981 de cDNA ipliğini tavuk lizozim geninden oluşturmuştur. Bununla birlikte, bu yöntem farklı cDNA'larda daha etkin sonuç verebilmektedir (ör:Cooke ve ark., 1980).

Oluşturulacak cDNA ekspresyon kütüphanesi için, cDNA klonlarının vektöre aktarılması sırasında problem yaşanmaması, eksiksiz bir şekilde her bir klonun bir vektöre aktarımın tamamlanması gerekmektedir. Farklı bir yöntemde ise, kullanılan ilk yöntemde olduğu gibi (hairpin-loop) saç tokası yapısı elde edildikten sonra, çift zincirli cDNA'nın kesiminden önce karışıma sentetik linker (bağlayıcı) eklenir. Daha sonra, linker diziler taşıyan oligo dT primer kullanılarak ikinci cDNA ipliğinin klonlanması sağlanabilir. Bu yöntemde de klasik saç tokası yönteminde olduğu gibi ikinci iplik parça parça klonlanır. (ör:Coleclough ve Erlitz).

cDNA sentezini plazmidle sağlamak için primerleri kullanmak da alternatif bir yöntemdir. Bu strateji Okayama ve Berg tarafından tasarlanmıştır ve iki aşamada gerçekleşir (1982). Birinci aşamada; cDNA ipliği daha kolay elde edilebilir. Önce RNA-DNA hibrid molekülü ilk cDNA ipliğinin sentezi için terminal transferaz reaksiyonunun substratıdır. Ancak bu birinci aşamada cDNA sentezi ancak, mRNA'nın korunan (metillenmiş) 3'-hidroksil grubuna kadar devam edebilir. İkinci aşamada ise; ikinci ipliğin sentezlenmesi RNA üzerinden RNaz enzimi ile oldukça duyarlı olan nick-translasyon tipi reaksiyonla oluşur. Bilinen cDNA klonlama yöntemlerinin yerine ikinci cDNA ipliğinin sentezlenmesini de içeren iki aşamalı yöntemler oldukça yaygındır.(ör:Gubbler-Hoffman 1983, Lapeyre ve Amalric, 1985). Bunlardan Gubbler-Hoffman yöntemi oldukça yaygın olarak bilinmektedir.

1.3 EXPRESSED SEQUENCES TAGS

Birçok organizmanın genom dizilerini belirleme çalışmaları günümüzde hızla devam etmektedir. Bu dizileme projelerinin amacı, genomik dizilerin belirlenmesi ve tüm gen setlerinin elde edilmesi olmasına rağmen; asıl önemli hedef ise, genlerin ne zaman nasıl ve ne şekilde aktif olduklarını belirlemektir. Normal koşullarda genlerin nasıl çalıştıkları anlaşılırsa, hastalık durumu gibi değişik şartlarda genlerin nasıl tepki verdikleri anlaşılabilir. Bu amaca ulaşmak için ise, araştırmacılar mutlaka protein veya proteinleri tanımlamalı, proteinleri kodlayan genleri bulup dizileri incelemelidir. Araştırmacılar son yıllarda cDNA klonlarına ait dizi bilgilerinin bilgisayar ortamına aktarılması ile elde edilmiş, eksprese olmuş dizi bilgilerini yani EST'leri (Expressed Sequence Tags, EST) bu amaç doğrultusunda kullanmaktadır.

EST'ler bilinmeyen genlerin tanımlanması ve yerlerinin tespit edilmesinde, gen ekspresyonu ve regülasyonu hakkında veri toplanması ve genom haritalarının oluşturulması konusunda kısa süreli ve ekonomik bir yol sağlamaktadır [4].

EST koleksiyonları için başlangıç materyalinin cDNA olması, gen tespit edilmesi işleminde genomik dizilere üstünlük sağlar. Genomun boyutları organizmadan organizmaya çeşitlilik gösterir ancak genellikle genomun büyük bir kısmı gen kodlanmayan bölgelerden oluşmaktadır bu yüzden genomik dizilerden genlerin tespit

edilmesi daha zordur.

Yalnızca bir defa eksprese olmuş cDNA klonu izole edilerek dizi analizi yapıldığında, araştırmacı elde edilen bu diziden birkaç yüz nükleotid uzunluğunda iki farklı EST dizisi oluşturur. cDNA geninin başlangıç kısmından elde edilen diziye 5'EST adı verilir. Bu dizi genelde cDNA'nın protein kodlayan kısmını oluşturmaktadır. Ayrıca bu dizi türler arasında korunmuştur ve gen ailesi içerisinde fazla bir değişiklik göstermez. cDNA geninin bitişine yakın kısımlarda elde edilen diziye ise 3'EST adı verilir. Bu diziler eksprese olmuş dizinin 3' ucuna yakın kısımlarından elde edildiklerinden, daha çok kodlama yapmayan ve translasyonu olmayan dizileri içermektedir ve kodlama yapan bölgelere göre türler arasında daha az korunmuşluk göstermektedir.

Bilinen en güçlü haritalama yöntemi, genom haritalaması için de kullanılabilen, dizi işaret bölgelerine (STS sequence-tagged sites) dayanan haritalama yöntemidir. STS genom veya kromozomda kolaylıkla tanımlanabilen ve sadece bir kere bulunan kısa DNA dizileridir. 3'EST dizileri, belirli türler arasında tek oluşları, benzerlikleri ve doğrudan eksprese olmuş geni işaret etmesi nedeniyle en temel STS kaynağıdır [4].

Transkript örnekleri için, EST'ler ve cDNA dizileri doğrudan kanıttır ve bunlar için günümüzde en önemli kaynaklar transcriptome araştırmalarıdır. High-throughput EST'ler herhangi bir organizmada transkripsiyonu yapılan dizilerde bir cDNA klonunun 5' veya 3' ucundan itibaren dizi hakkında bilgi almak için etkin olabilir. 1991 yılında EST'ler insanda gen keşfinde birincil kaynak olarak kullanılmışlardır (Adams ve ark., 1991). Daha sonra, sayısız organizmalar için EST veri birikiminde ve üretiminde hızlı bir büyüme olmuştur. Şu anda, EST'ler gen keşfini etkinleştirmede, genom açıklamada, gen yapısının tanımlanmasına destek olmada, alternatif transkriptleri belirlemede, tek nükleotit polimorfizminin (SNP) karakterizasyonu için rehber olmakta ve proteom analizini kolaylaştırmada etkin olmaktadır (Jongeneel CV. 2000; Dong ve ark., 2005; Rudd, 2003).

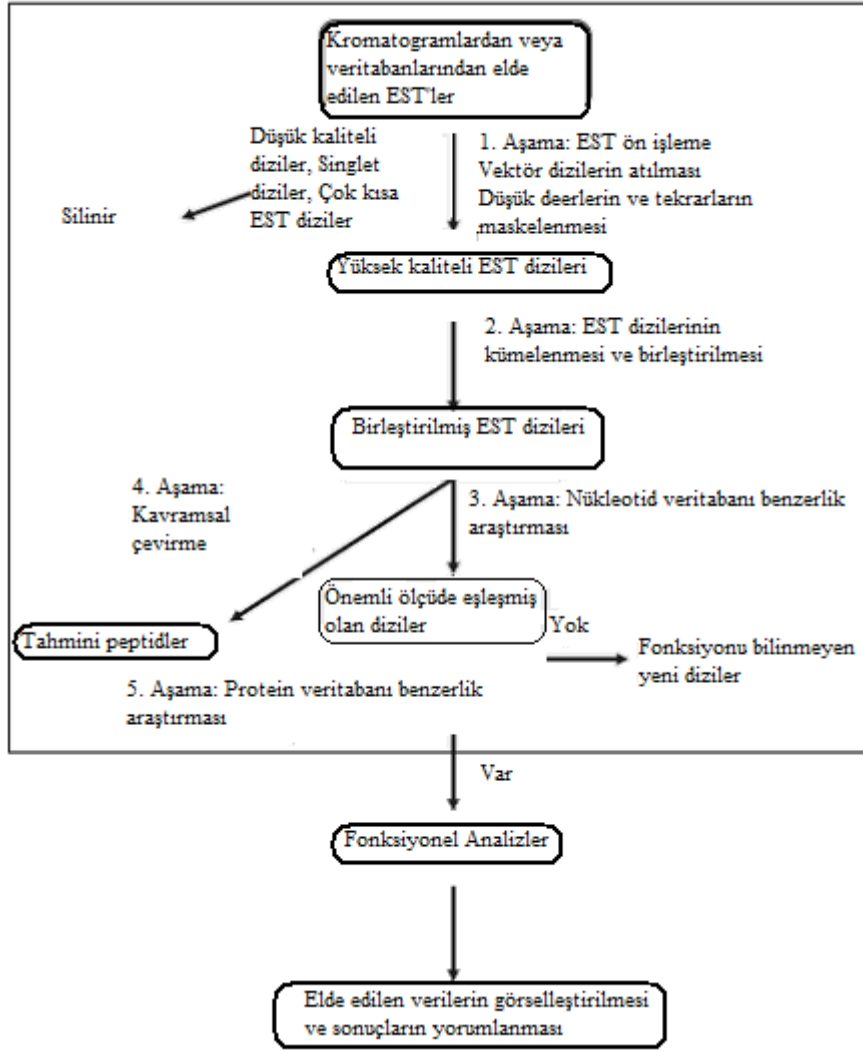
Tüm genom dizilemesi büyük genomlu organizmalar için zor ve pahalıdır. Ayrıca, mısır gibi bitkilerdeki retrotransposon tekrarlarının bir sonucu olarak meydana gelen genom genişlemesi, tüm genom dizilemesi uygulamasının pek tercih edilmemesine neden olur (Bennetzen, 2002). Bu çalışmada Gemlik zeytin bitkisi için transkripsiyonu yapılan dizilerden EST'ler üretilmiştir. Böylece uygulanabilecek en kolay yöntem ile genomu hakkında fazla bilgi edinilememiş zeytin bitkisi için en kapsamlı genom çalışmalarından

biri yapılmıştır.

1.3.1 EST'lerin Biyoinformatik için iyi bir kaynak olması

High throughput (mikroaraylar, otomatik dizileme ve kütle spektrometrisi gibi) dizileme teknolojisinin gelişmesi, cDNA kütüphanelerinden rastgele seçilmiş binlerce klondan single-pass dizileme yani 5 've 3' uçlarından tek yönlü dizileme ile 300 baz çiftlik EST (Expressed Sequence Tags) adı verilen anlatımı yapılan dizi parçalarının oluşmasına olanak sağlamıştır (Wilcox ve ark. , 1991; Okubo ve ark. , 1992).

EST veri kümelerinin analiz çalışmalarında genellikle benzer aşamalar takip edilir (Şekil 1). Öncelikle ilk aşama olarak veritabanından elde edilen veya laboratuvar çalışmalarında üretilen EST dizilerinin ön-işlemesi yapılır, dizi tekrarları taranarak ve düşük kaliteli veya kontaminant diziler elimine edilerek yüksek kaliteli EST'ler belirlenir. Daha sonraki aşamada, dizi benzerliklerine dayanarak yüksek kaliteli EST'ler kümeler halinde gruplandırılırlar. Bu kümelerden 'birleşme (assembling)' ile maksimum bilgilendirici olduğu düşünülen her biri tahmini (putative) bir gen olabilecek diziler türetilir. Bu aşama, birkaç kısa EST dizisinden aynı çakışan bilgilerin toplanması ile dizi uzunluğunun artırılması sağlanır. Elde edilen ve contig olarak adlandırılan dizi veritabanlarıyla karşılaştırılarak benzerlik sorgulaması yapılır ve eğer diziler için önemli veritabanı eşleşmeleri bulunursa bu diziler için olası fonksiyonlar belirlenir. Ayrıca, birleşmiş bir dizi kavramsal olarak tahmini bir peptide çevirilebilir ve ardından protein dizi veritabanları ile karşılaştırılabilir. Tüm transkriptomlar protein ürünlerine çevrilebilir olmasa da domain ve motif analizini de içeren protein merkezli fonksiyonel açıklama, protein analiz aracı olarak uygulanabilir. Birleşme aşamasında herhangi bir contig dizisi içinde yer almayan EST'ler ise "sinleton" veya "singlet" olarak adlandırılırlar. Singletler çoğunlukla bir gen dizisini ifade edecek yeterlilikte olmadıklarından fonksiyonel analizleri yapılmaz.



Şekil 1.1: EST analizindeki genel aşamalar. 1. Ham EST dizilerinin kesilen veya maskelenen vektör kontaminasyonu, düşük değerlikli diziler ve tekrar bölgeleri bakımından kontrol edilmesi. Ayrıca, düşük kaliteli, singleton olan ve kısa diziler de atılır. 2. Kümelenen ve birleştirilen EST'lerin contig (birleşmiş dizi) oluşturması. (contig = ' tahmini transkript). 3. Homolog olan ve benzer fonksiyona sahip olan dizileri belirlemek için DNA veritabanı benzerlik araması uygulanır. 4. Contiglerin kavramsal translasyonu ile tahmini peptitler belirlenir. 5. Tahmini fonksiyonların belirlenmesi için protein veritabanı benzerlik araması uygulaması yapılır. Sonuçların görselleştirilmesi ve yorumu ile fonksiyonel anlamlandırmanın yapılması için analiz genişletilir (Nagaraj ve ark., 2006).

EST' ler genlerin keşfi için genom haritalarında fiziksel marker olarak ve genomik

klonlardaki genlerin tanımlanması için kullanılmaktadırlar. (Adams ve ark. , 1991, 1992, Banfi ve ark. , 1996). Bir genin analizi için başlangıç ve bitiş kodonlarının her ikisinin de bulunduğu tam boy olarak nitelenebilen (full length) olduğundan emin olunan bir cDNA'ya sahip olmak araştırmacı için çok önemlidir. Tam boy cDNA klonu daha az analiz ile çok fazla bilgi verir araştırmacıya. (Primrose ve Twyman, 2006). Full length olsun yada olmasın farklı organizmalardan gelen yaklaşık olarak 82 milyon EST araştırmacıların kullanımına sunulmak üzere EST veritabanlarında bulunmaktadır. EST verilerinin gelişimi de GenBank'ın kapsamını daha iyi gözlemlemek için çeşitli araştırmacılar tarafından birkaç kez incelenmiştir böylece geçen yıllar içinde GenBank'a kaydedilen bilgi miktarının artışı, GenBank kullanımının yaygınlaşma hızı ve GenBank'ın kullanıcılarına sunduğu olanakların gelişimi gözlemlenebildi (Boguski 1995, Gerhold ve Caskey 1996, Okubo ve Matsubara 1997).

EST' ler haritalamada kullanılabilmelerinin yanı sıra gen anlatımının analizi için de oldukça kullanışlıdır. cDNA microarrayleri (Schena ve ark., 1995) için çok sayıda hedef dizinin üretilmesi EST' lerden üretilen PCR primerleri ile sağlanmaktadır. Ayrıca genin anlatım dizisinin tamamını kapsamayan (full length olmayan) cDNA parçalarının kullanıldığı farklı doku veya farklı koşullarda farklı anlatım yapan genlerin belirlenmesine dayalı differential display (Liang ve Pardee,1992) tekniğinde de EST' ler önemli bir yer tutmaktadır. Gen anlatımının analizi için en son geliştirilen ve EST kullanılan bir diğer yöntem ise gen anlatımının seri analizi yöntemi olan SAGE (serial analysis of gene expression) yöntemidir (Velculescu ve ark., 1995). SAGE tekniğinde iki farklı dokudan veya iki farklı koşuldaki gelen her bir cDNA kütüphanesi için, EST uzunlukları yalnızca 9 – 10 baz çiftinden oluşan cDNA klonlarının birleşiminden elde edilen iki farklı koşula ya da iki farklı dokuya ait uzun iki dizinin karşılaştırılması yapılır. (sadece spesifik transkripti belirlemeye uygun minimum dizi).

Genom çalışmalarına büyük yarar sağlayan EST'ler ve diğer nükleotid dizileri ile protein dizileri üzerinde çalışılırken elde edilen sonuçların birbirleri ile karşılaştırılması yeni bulguların sağlanmasını kolaylaştırmaktadır. Araştırmacıların bu kaynakların hepsine aynı yerden ulaşması çalışmaların hızı ve güvenilirliği için gereklidir. Bu yüzden tüm verilerin toplandığı tek bir kaynak olan GenBank adında önemli bir veritabanı bulunmaktadır.

1.3.2 Ham EST'lerin İşlenmesi ve Birleştirilmesi

Eksprese olmuş dizi işaretleri (EST'ler) (Adams ve ark. , 1991) bölgesel olarak dizilenmiş tekil bir cDNA kopyasıdır. EST dizileri; gen tahmini, alternatif kesim (splicing), özelleşmiş dokulardaki gen anlatım farklılaşması gibi birçok farklı alanda kullanılmışlardır.

Dizi verilerinin işlenmesi ham EST dizisinin içerdiği vektör, adaptör ve düşük değerli dizilerin uzaklaştırılması verilerin daha iyi incelenmesi için temizlenmesi demektir. EST ön işleme olarak da bilinir. EST ön işleme sırasında vektör kontaminasyonunun giderilmesi, EST dizileme sürecinin önemli bir işlemidir. İnternette de ulaşılabilen Phred gibi, dizi verilerinin temizlenmesini sağlayan birçok program mevcuttur. Bunlara ek olarak, Phil Green'in Cross-Match, TIGR'den Lucy ve SeqClean, NCBI tarafından sağlanan VecScreen gibi çevirim içi olmayan programlar da bulunmaktadır (Chen ve ark. , 2007).

Cross-Match Smith-Waterman algoritması üzerine kurulmuştur (Smith ve Waterman, 1981). Program çok hassas ve verimli bir karşılaştırma algoritması çalıştırır ve iki dizi arasındaki benzerlikleri tanımlamaya çalışır. Ayrıca hedef diziden, vektör kontaminasyonunu çıkartmak için ek bilgisayar analiz süreçlerine ihtiyaç vardır. Cross Match; Partigene gibi EST işlenmesi ve birleştirmesi, dizileme ve fonksiyonel bildirim gibi farklı işleri yapan biyoinformatik programları arasında ilişkiyi sağlayan bir program ile diğer EST işleme paketlerine entegre edilmiştir (Chen ve ark. , 2007). Bu çalışmada Cross-Match programı kullanılmış ve ham dizilerin ön hazırlığı sorunsuz yapılmıştır.

Elde edilen farklı gen fragment dizilerinin bir araya getirilmesi EST analizlerinin önemli bir adımıdır. Sanger ve arkadaşları (1977) ile Maxam ve Gilbert (1977) adlı araştırmacıların başlattığı dizileme teknolojilerinin gelişmesi sonucu araştırmacılar 1000-1500 nükleotide kadar dizileme işlemleri yapmayı başarabilmişlerdir. Ancak dizi parçalarını manuel olarak birleştirmek oldukça zordur. İlk çalışmalar basit bilgisayar programları ile üst üste binen fragmentleri (parçaları) birleştirmekle başladı. Fakat günümüzde çok daha karmaşık analizlerde (İnsan genom projesi gibi çalışmalarda) yeni bilgisayar programları kullanılarak daha etkili sonuçlar alınmaktadır.

Fragmentleri birleřtirmek oldukça önemli problemlerden biri iken arařtırmacılar birleřen fragmentlerin oluřturduėu contig olarak bilinen dizilerin kalitesinin yüksek olmasının da çözümlenmesi gereken bir problem olduėunu fark ettiler. Bu yüzden contig dizilerinin kalitesi için dikkat edilmesi gereken bir takım durumlar anlařıldı. Bunlar; İnsert dizinin her iki ucunun da ikiden fazla kez dizayn edilmesi olarak bilinen “Dual End (Çift yön)” gereksinimi, “Noisy” terimi ile ifade edilen, jelde oluřan bir hata yüzünden bilgisayar programında okunabilmesine karřın orijinal diziyi belirlemede yetersiz olan fragmentlerin atılması, Sampling Without Replacement (Yerine koymadan örnekleme) uygulaması ile o ana kadar dizilemesi yapılmıř hiçbir dizi ile hibridize olmayan klonların ayırđ edilmesidir.

İyi tasarlanmıř ve fragment birleřmesi problemi için pratik algoritmalara sahip, kullanımı kolay bilgisayar programları birçok özelliėi bünyesinde bulundurması arařtırmacılar tarafından arzu edilmektedir. Myers tarafından ideal ve kullanım kolay bir programın ne gibi özelliklere sahip olması gerektiėi ortaya konulmuřtur (Myers, 1994).

Dizilerde en sık karřılařılan ve ařılması gereken sorun hatalardır ve genellikle %1 civarında olurken bu rakam bazen %5’in üzerine de çıkabilir. Kullanacaėımız programın bu yüksek hata oranlarını mümkün olduėu kadar kullanılabilir řekilde sunması istenmektedir. Böylece daha fazla dizinin kullanılabilir olması amaçlanmaktadır. Ancak hataların varlıėı, fragmentler arasındaki üst üste binmelerin ve baėlanma bölgelerindeki çoklu eřleřmelerin düzenlenmesini dikkatle yapılmasını gerektirmektedir. Bir arařtırıcının bir fragment birleřtirme programından beklentileri genom çalıřmasının en hızlı ve en doėru řekilde yapılması için oldukça önemlidir (Myers, 1994)..

Arařtırıcının beklentilerini en iyi karřılayacak fragment birleřtirme programının istenilen bazı özelliklere sahip olmalıdır. Eldeki dizilerin oryantasyonunun yani hangi yöne doėru dizilendiėinin program tarafından bilinmesi fragment birleřtirme programı için bir avantaj olacaktır. EST ön iřleme ařaması uygun bir programla yapılmalı, ancak fragment birleřtirme programlarının algoritmaları tarafından vektör parçalarının taranabilmesi mümkün olmalıdır. Aynı zamanda hangi fragmentin nerden geldiėini diėer fragmentlerle nasıl üst üste bindiėini belirten bilgileri vermelidir. Fragment birleřtirmede en iyi çözümlenmenin program tarafından yapılmıř olduėunun gösterilmesi için program tarafından alternatif birleřtirme sonuçları da karřılařtırılabilir (Myers, 1994). Bu çalıřmada

Phrap ve CAP3 programı fragment birleřtirmek amacıyla kullanılmıřtır.

1.4 GenBank

GenBank, Amerikan Ulusal Saęlık Enstitüleri (NIH National Institutes of Health) kampüsünde bulunan, Ulusal Tıp Kütüphanesi (NLM National Library of Medicine) bünyesinde çalışan, Ulusal Biyoteknoloji Merkezi (NCBI National Center for Biotechnology Information) tarafından oluşturulup dağıtımı yapılan, biyolojik ve bibliyografik veri desteęi saęlayan, nükleotid ve protein dizlerinin toplandıęı kamuya açık bir veritabanıdır.

NCBI; GenBank'ı arařtırıcıların doğrudan gönderdikleri dizi bilgileri sayesinde oluřturmaktadır. Dięer veri kaynaklarını ise, EST dizilerinin ve dięer dizileme merkezlerinin gönderdięi diziler oluřturmaktadır. Amerika Birleřik Devletleri Patent ve Marka Ofisi de (USPTO United States Patent and Trademark Office) konu edilen patentlerden dizi bilgisi saęlamaktadır. Ayrıca, dięer veritabanları da katkı saęlamaktadır. İngiltere' deki EMBL (European Molecular Biology Laboratory) [2] ve Japonya'daki DDJB (DNA Data Bank of Japan) [1] GenBank DNA veribankasının uluslararası uzun süreli iřbirlikleri sonucunda, dizi bilgileri alışveriři saęlanmakta ve geniř kapsamlı bir dizi bilgisi oluřturulmaktadır. NCBI dizi bilgilerini; Internet, FTP ve dizi karřılařtırma servisleri sayesinde ücretsiz olarak sunmaktadır (Benson ve ark. , 2001). Bilinen bu veritabanlarının dışında, Ensembl (European Molecular Biology Laboratory- European Bioinformatic Institute ve Sanger Enstitüsünün ortak bir çalışması) [14], UCSC (University of California Santa Cruz) [15], GO (Gen Ontoloji) [12], KEGG (Kyoto Encyclopedia of Genes and Genomes) [16] veritabanları da yaygın olarak kullanılmaktadır.

1.4.1 GenBank Kayıtları Ve Kısımları

Her Genbank kaydı; dizinin kısa bir tanımı kaynak organizmanın taksonomik sınıflandırması, bilimsel adı, bibliyografik referanslar ve kodlama bölgeleri ile dięer biyolojik özelliklerini (transkripsiyon öğeleri, tekrarlanan diziler, mutasyon veya modifikasyon yerleri) içeren özellik çizelgesi gibi verileri içerir. Kodlama bölgeleri için protein translasyonu da özellik çizelgesinde bulunmaktadır.

Genbank dağıtımındaki dosyalar, kabaca taksonomik sınıflandırmaya (bakteriler, virüsler, primatlar, kemiriciler) karşılık gelen kısımlara bölünmüşlerdir. Son yıllarda biyolojinin bazı bölümlerine özel EST dizileri, genom dizileri, yüksek verimli genomik çıktılar gibi yeni kısımlar eklenmiştir. Şu anda 17 tane spesifik veri alanı bulunmaktadır. NCBI'nın ftp sunucusuna gönderilen yeni Genbank sürümleri; dosya transferindeki tutarlılığın sağlanması için, EST ve primatlar gibi bazı kısımlar çoklu dosyalara bölünmüşlerdir. (Benson ve ark. , 2001) Bu dosyalar; EST' ler, STS' ler (Dizi işaret bölgeleri), GSS' ler (Genom haritalama dizileri), HTG'lerdir (yüksek verimli genomik veriler).

1.4.2 Dizi belirleyicileri ve Erişim numaraları

Her GenBank dizi kaydı, o dizi için özel ve dizi içeriğinde bir değişme olsa bile sabit kalan bir erişim numarasına sahiptir. Farklı kaynaklardaki özel dizileri belirlemek ve dizi içeriğindeki değişimleri takip etmek için, NCBI her diziye belirleyici bir "gi" numarası vermektedir. Dizideki bir değişim halinde, yeni dizi versiyonuna yeni bir "gi" numarası verilecektir. Bu gi numaraları, Erişim (ACCSESSION) alanını takip eden Nucleotide ID (NID) kısmında bulunmaktadır.

1.4.3 EST'ler ve NCBI

EST'lerin kolay ulaşılmaları ve bir gen keşif aracı olarak kullanılmaları için, diğer genom verilerine de ulaşılabilen bir veritabanında bulunmaları gerekmektedir. Bir EST oluşturulduğunda, NCBI tarafından yönetilen NIH dizi veritabanı olan Genbank' a yüklenir. Bu nedenle NCBI da, 1992 yılında sadece EST'ler için kullanılan bir veritabanı oluşturulmuştur. Bir EST dizisi Genbank' a geldiğinde hemen incelenmekte ve dbEST veritabanına aktarılmaktadır [4].

Bir genin kopyası olan mRNA birçok kere eksprese edildiğinden, bu mRNA'ların dizilenmesi ile oluşturulan EST'lerin sayısında da buna bağlı olarak fazlalık olabilmektedir. Yani aynı EST'nin birden fazla kopyaları oluşabilmektedir. Bu gibi fazlalık EST'lerin bulunması ve dizilerin üst üste binmesi yüzünden, belirli bir EST arandığı zaman birbirine benzeyen dizilerden arama yapmak durumunda kalınmaktadır. Tüm bu EST dizilerini incelemek araştırmacılara zaman kaybettirmektedir. Bu gibi fazlalık

ve dizi akışması durumlarını özmek için NCBI arařtırcıları 'unigene' adında bir veritabanı oluşturmuşlardır. Bu veritabanı Genbank dizilerini otomatik olarak tarayarak, onları gen olma ihtimali fazla olan aynı veritabanında kopyası veya herhangi bir parçası bulunmayan sadece birbirinden farklı dizilerin bulunduğu veriye, yani unigene dönüřtürmektedir. Böylece kolaylıkla karşılaştırma yapılabilir.

2. MATERYAL VE METOD

2.1 MATERYAL

2.1.1 Bitki Materyali

Yalova Atatürk Bahçe Kùltürleri Merkez Arařtırma Enstitüsü Mùdùrlüğü tarafından yürütölen ıslah programında geliřtirilen Gemlik G20 kùltür çeřidi bitki materyali olarak kullanılmıřtır.

2.1.2 Kullanılan Kitler

RNA izolasyonu için RNeasy kiti (Qiagen,), mRNA izolasyonu için Oligotex mRNA (Qiagen,), cDNA kùtùphanesi oluřturmak için CloneMiner cDNA Library Construction Kit (Invitrogen) kullanıldı. Protein ekspresyonu ve fonksiyonel analiz için çoklu hedefi olan vektör içersine hazırlanan kùtùphane E. Coli içersine yüksek verimde aktarılır.

2.1.3 Programlar

2.1.3.1 Phred Ve Phrap

Vektör dizilerin uzaklařtırılması ve örtüřen dizilerin belirlenip birleřtirilmesi (fragment assembly) iřlemleri için Phred ve Phrap unix paket programları kullanılmıřtır. Washington Üniversite'sinden Phil Green ve grubunun oluřturduđu, EST analizinde kullanılan, Phred vektör ve adaptör diziler ile düşük deđerli dizilerin uzaklařtırılmasını sađlarken bu çalıřmada en son sürümü, 3/19/99 versiyonu kullanılan Phrap programı ise örtüřen dizilerin belirlenip birleřtirilmesini sađlar. Phrap programı içersinde aynı zamanda “cross_match” adında ek bir program daha bulunmaktadır. “cross_match” vektör dizilerin yok edilmeden önce maskelenmesini sađlar ve phred ile atılmayan vektör kalıntılarının tamamen uzaklařtırılmasını sađlar (Ewing ve Green 1998; Ewing ve ark. , 1998).

2.1.3.2 CAP3

Fragment assembly işlemi için alternatif bir program olarak kullanılmış ve sonuçları değerlendirilmiştir. Huang ve arkadaşları tarafından 1992 de geliştirilen fragment assembly programı CAP3, ulaşılabilirliği, kullanım kolaylığı ve sonuçlarının diğer fragment birleştirme programlarına (Phrap) göre daha güvenilir olması sebebiyle bu çalışmada ve pek çok çalışmada tercih edilmiştir (Huang ve ark., 1999).

2.1.3.3 Consed / Autofinish

Consed / Autofinish, Washington Üniversite'sinden Phil Green'in ekibinde bulunan David Gordon tarafından görüntülemek, düzenlemek ve dizi birleşimini sonuçlandırmak için phrap programı ile birlikte tasarlanmış bir programdır (CAP3 ile de uyumludur).

Kullanıcının yorumunun da dahil olduğu primerleri ve kalıpları seçmek için, gereken durumlarda ek dizileme reaksiyonlarını gerçekleştirmek için ve bölme yaparak birleştirme işleminin doğruluğunu kontrol etme özelliği ile forward-reverse parçalar hakkında bilgi edinmek için sonuçlandırma işlemlerine olanak sağlamaktadır (Gordon ve ark., 1998). Fragment birleştirme işlemi tamamlandıktan sonra sonuçların doğruluğunu kontrol etmek amacıyla kullanılmıştır.

2.1.3.4 BLAST2GO

Ana Conesa ve ekibi tarafından geliştirilen Blast2GO (B2G) fonksiyonel analiz çalışmaları ve gen veya protein dizilerinin (karakterizasyonu yapılmamış) analizi için kapsamlı bir biyoinformatik araçtır. Bu çalışmada elde edilen contig dizilerinin fonksiyonları hakkında daha fazla fikir sahibi olunması için kullanılmış bir programdır (Conesa ve ark., 2005).

2.1.3.5 VecScreen

NCBI tarafından geliştirilen VecScreen programı EST dizisinden istenmeyen dizilerin uzaklaştırılıp uzaklaştırılmadığını kontrol etmek için,

<http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html> internet adresinden elde edilen dizilerin yüklenerek programın çalıştırılmasıyla (online) kullanılan bir uygulamadır. Bir dizinin vektör ya da adaptör dizisi içerip içermediğini grafiklerle ifade eder.

2.1.3.6 Sequin

NCBI tarafından geliştirilen Sequin programı bir EST projesinde tüm çalışmalar tamamlandıktan sonra birleştirilmiş ve bir gen olma ihtimali olabilen contig dizilerinin GenBank veritabanına kaydedilmesini sağlayan programdır.

2.2 METOD

2.2.1 RNA İzolasyonu

RNA izolasyonu için Qiagen RNeasy kit kullanıldı. Sıvı azot içerisinde toz haline getirilen 100 mg yaprak ve olgunlaşmamış zeytin tanesi örnekleri RNA izolasyonu için kullanıldı. İzolasyon Qiagen RNeasy kitin protokolüne göre yapıldı.

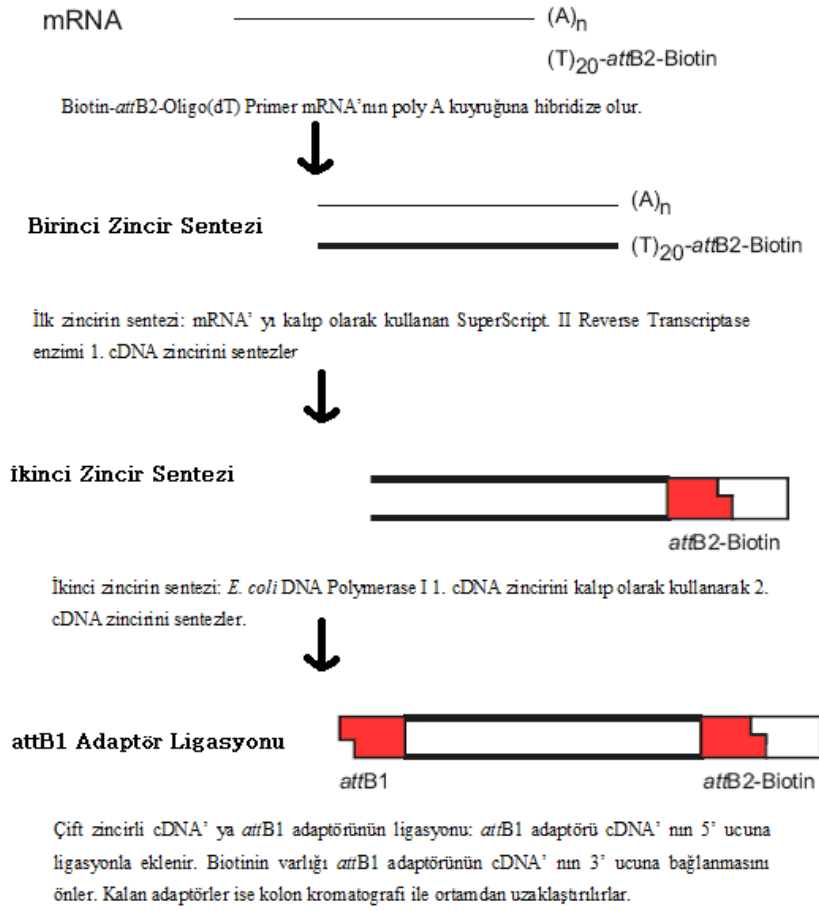
2.2.2 mRNA İzolasyonu

Elde edilen mRNA miktarının yeterli olması için başlangıç RNA miktarı 10 farklı izolasyon bir araya getirilerek arttırıldı. mRNA izolasyonu için Oligotex mRNA Spin-Column kiti (Qiagen,) kullanıldı. Kit protokolü uygulanarak total RNA 'dan yaklaşık 3 µg mRNA izole edildi. Elde edilen bu 3 µg mRNA cDNA kütüphanesinin oluşturulması için başlangıç materyali olarak kullanıldı.

2.2.3 cDNA Kütüphanesinin Oluşturulması

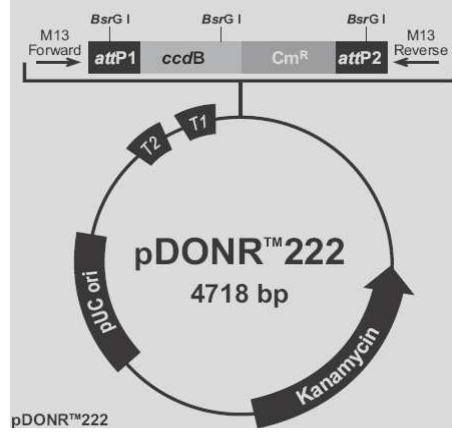
Başlangıç materyali olarak yaprak ve zeytin meyvesinden elde edilen mRNA'lar ayrı ayrı kullanılarak iki farklı kütüphane İnvitrogen CloneMiner™ ticari kiti kullanılarak oluşturuldu. Protokole göre önce Biotin-attB2-Oligo (dt) primeri mRNA'nın polyA kuyruğuna hibridize oldu ve birinci cDNA sentezi Super Script II Reverse Transcriptase

enzimi ile yapıldı. Karşı zincirin sentezi için E.coli Polymerase I enzimi kullanıldı ve çift zincirli cDNA'ya attB1-adaptörü ligasyonla eklendi (Şekil 2).



Şekil 2.1.: cDNA kütüphanesinin oluşturulma aşamaları.

Daha sonra CloneMiner kiti için geliştirilen vektör pDONRTM222 (Şekil 3) ve hazırlanan cDNA rekombinantı, BP Clonase enzim karışımı kullanılarak kitin protokolüne göre birleştirildi. Konakci hücre E.coli DH10BTM içersine vektör-cDNA rekombinantı CellJect Duo kullanılarak elektoporasyon yöntemi ile alındı. Elde edilen cDNA kütüphanesi SOC mediumda 1 saat 37 C inkübe edildi ve freezing media ile -80 C de saklandı.



Şekil 2.2.: pDONR vektör

2.2.4 Kütüphanenin Karakterizasyonu

Kütüphaneden LB agar medium üzerinde bulaştırılan koloniler tek tek rastgele seçilerek, 384 kuyucuklu plateler içine alındı ve -80 C de plateler içinde saklandı. Platelerden rastgele seçilen altmış klondan alkali lizis yöntemi ile plasmid DNA'sı izole edildi (Sambrook et al., 1989, Feliciello and Chinalli 1993). İzole edilen DNA Bgl1701 restriksiyon enzimi ile kesildi ve insört boyutunu belirlemek için %1'lik agaroz jel üzerinde yürütüldü. İnsört büyüklükleri ve kaliteleri gözlemlendikten sonra 3840 adet klon dizileme için Genom Dizileme Merkezi, WUSTL gönderildi. M13 universal primerleri kullanılarak PCR çoğaltımı için kalıp olarak kullanılan klonlar Genom Dizileme Merkezinde ABI 3730 kapiler sequencer (PE Applied Biosystems, Foster City,CA) kullanılarak otomatik high-throughput teknolojisi ile dizileme gerçekleştirildi.

2.2.5 Vektör Uzaklaştırma

Phred programı tam olarak “base calling” adı verilen dizileme, vektör ve adaptör dizilerinin uzaklaştırılması işlemlerini yapar. Phred, kromatogram dosyalarındaki baz izlerini takip ederek base calling yapan, dizideki bazlara quality değerleri atayan bir programdır.

Öncelikle oluşturulan base calling ve quality değerleri Phrap birleştiricisi tarafından contig dizisi oluşturulmasında kullanılmak üzere Phred programı ile çıktı haline getirildi. Phrap programı içerisinde bulunan “cross_match” adında ek bir program ise vektör dizilerin yok edilmeden önce maskelenmesini sağlar ve phred ile atılmayan vektör kalıntılarının tamamen uzaklaştırılmasını sağlar.

Cross-Match Smith-Waterman algoritması üzerine kurulmuştur. Program çok hassas ve verimli bir karşılaştırma algoritması çalıştırır ve iki dizi arasındaki benzerlikleri tanımlamaya çalışır. Ayrıca hedef diziden, vektör kontaminasyonunu çıkartmak için ek bir bilgisayar süreçlerine ihtiyaç vardır. Cross Match; Partigene gibi dizi temizlenmesi EST kümelenmesi, dizileme ve fonksiyonel bildirim gibi farklı işleri yapan biyoinformatik programları arasında ilişkiyi sağlayan bir program ile diğer EST işleme paketlerine de entegre edilmiştir. (Chen ve ark. , 2007)

ABI 3730 kapiler sequencerden elde edilen, kromatogram formatındaki 3840 klonun FASTA formatına dönüştürülmesi ve düşük değerli olan dizilerin belirlenmesi için Phred programı ile dizilerin analizi yapıldı. 106 EST dizisi Phred sonuçlarına göre yetersiz olduğundan değerlendirmeden çıkarıldı. Kalan 3734 EST dizisi ise Phred programı ile kromatogram formatından FASTA formatına dönüştürülmüş olarak elde edildi. Program çalıştırılırken default değerler (phred skoru için 20 olan) yani programda kullanıcı tarafından hiçbir değişiklik yapılmadan kullanılmıştır. EST dizilerini vektör dizilerden tamamen kurtarmak için FASTA formatındaki 3734 dizi Phrap programı içindeki “cross_match” uygulaması ile analiz edildi. Böylece EST’ler fragment birleştirme işlemi için hazır hale getirildi.

2.2.6 Contig Oluşturma

Phrap programı ile fragmentlerin birleştirilmesi:

Phrap ise fragmentlerin birleşmesi için phred'in base calling skorlarını kullandı. Phrap, verilen pozisyondaki tüm bağımsız dizileri inceledi ve en yüksek skoru almış diziyi kullandı ve üst üste binmiş DNA dizilerinden, en yüksek kalitede olanları seçerek istatistik olarak hesaplanmış mozaikler oluşturdu. Bu, çoklu karşılaştırma metodlarıyla gelen karışık algoritma işlemlerini ve bu yöntemler kullanıldığı için oluşan dizilerdeki baz tutarsızlıklarını engellendi.

Phrap programının çalışma prensipleri şu aşamalardan oluşmuştur:

1-Dizi ve quality dosyalarının içeriğini okunarak, düşük kaliteli okuma sonuçları elenip okuma işleminin tamamlanmasını sağlandı.

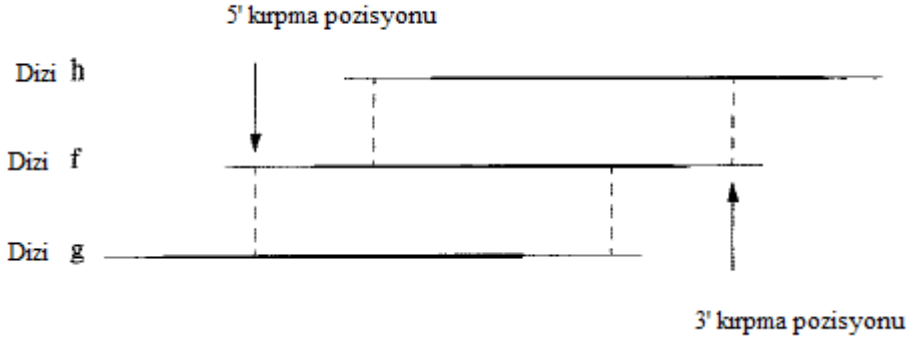
2-Okunmuş olan parçaların eşleşen bazları bulundu. Birbiriyle tam olarak eşleşen

- sonular ayrıldı. Eşleşen dizi çiftleri Smith-Waterman algoritmasıyla incelendi.
- 3-Vektörlerle eşleşen kısımlar bulundu ve bu bölgeler birleşme işlemine katılmaması için işaretlendi.
 - 4-Yakın olarak eşleşen parçalar bulundu ve birleştirildi.
 - 5-Dizilerin kendileriyle yaptıkları veya hatalı yaptıkları eşleşmeler engellendi.
 - 6-Eşleşmiş bölgeleri tanımlamak için daha önceden doğrulanmış okunma sonuçları kullanıldı ve bunlar da quality değerlerini tekrar hesaplamak için kullanıldı.
 - 7-Her eşleşme için LLR skoru hesaplandı. (LLR skor üst üste binen kısımların boylarının ve kalitelerinin değerlendirildiği bir ölçüdür)
 - 8-Tüm eşleşmeler içinde en iyi eşleşen dizi çifti bulundu. (yani en yüksek LLR skoruna sahip üst üste binme belirlendi.)
 - 9-Contiglerin düzenlenmesi azalan LLR skorlarda en uygun eşleşen kısımlar kullanılarak sağlandı. (Greedy algoritma ile)
 - 10- Okunma sonuçlarından yüksek kaliteli olarak çıkan parçalardan bir mozaik gibi contigler oluşturuldu.
 - 11-Contigler için birleşen kısımlar okundu; uyuşmayanları ve yanlış birleşmiş olabilecek olanları çizelge haline getirilip contiglerin LLR skorları ayarlandı [6] .

CAP3 programı ile fragmentlerin birleştirilmesi:

- 1) Örtüşen dizi çiftleri hızlı bir şekilde tanımlandı
- 2) Düşük kaliteli bölgeler kırıldı:

Şekil 2'de dizilerin 5' ve 3' kırılma pozisyonlarının hesaplanması gösterilmektedir. Diğer diziler ile güçlü benzerlik gösteren bölgeler en uyumlu birleşme bölgelerinin başlangıç ve bitiş kısımlarının her bir dizi parçası için potansiyel bir örtüşme ile hesaplanması sonucu belirlenmiştir. Verimlilik için, Smith ve Waterman (1981) algoritması ile hesaplama dinamik programlama matrisinde köşegenlerin kesiştiği bir bölgeyi kısıtlamaktadır (Pearson ve Lipman 1988; Green 1996). Oysa, CAP3 programına ait potansiyel olarak örtüşen her bir dizi parçasını hızlı bir şekilde bulabilen yöntem, parçalar için köşegen bölgelerini belirleyebilmektedir. Köşegen bölgesi her iki yöne de kullanıcı tarafından belirlenebilen kesin baz numaraları ile genişletilebilmektedir. Lokal bir birleşme hesaplaması köşe bölgelerinin genişlemesi üzerinden gerçekleştirilmiştir (Huang ve ark., 1999).



Şekil 2.3.: 5' ve 3' kırma bölgelerinin f dizisinde hesaplanması. F dizisi g ve h dizileri ile yüksek lokal benzerlik göstermektedir. Bir çift kırık hat benzerliğin başlangıç ve bitiş pozisyonlarını gösterir. Kalın bir hat dizilerin yüksek kaliteli bölgesini gösterir.

3) Örtüşmelerin hesaplanması ve değerlendirilmesi

Diziler arasındaki örtüşmeler hesaplandıktan sonra yanlış olan örtüşmelerin belirlenmesine geçildi. Öncelikle 5' ve 3' uçlarındaki zayıf bölgelerden uzaklaştırılmış temiz diziler elde edilmişti. İki dizi arasındaki örtüşme, maksimum benzerlik skoru (Huang 1992) ile dizilerin global alignmentı olarak tanımlandı. Örtüşmenin iki düşük kaliteli dizinin lokal alignmentı yerine iki temiz dizinin en iyi global alignmentı olarak tanımlanması yanlış örtüşmelerin belirlenmesi için uygun olacaktır. İyi bir global alignment örtüşmenin yanlış olduğunu belirten dizilere ait benzer olmayan bazı iyi bölgeleri de gösterebiliyorken, iyi bir lokal alignment yalnızca benzer bölgeleri gösterir.

Potansiyel örtüşme ile her bir dizi parçası için, köşegen bölge daha önceden hesaplanmış en iyi lokal alignmentın başlangıç konumuna merkezlendi. Lokal alignment için köşegen genişliği hesaplanmasında iki yönde köşegen genişletme boyutu kullanıldı. Global alignment ile maksimum benzerlik skoru köşegen üzerinden hesaplandı. Doğrusal düzlemde köşegenin hesaplanabilmesi için parçala-fethet tekniği (Chao ve ark., 1992) kullanıldı. Lokal alignment hesaplamasında, eşleşme ve fark skorları baz quality değerleri ile ağırlıklandırıldı. Global alignment diziler arasındaki örtüşmedir. Uzunluk, benzerlik skoru ve örtüşmenin tanımlanma yüzdesi bu alignmentlarla sırasıyla belirlendi.

Her bir örtüşme beş ölçüm ile değerlendirildi. Eğer örtüşme bu beş ölçümün herhangi biri için yeterli olmazsa bu örtüşme için contig oluşumu düşünülemez. İlk üç ölçüm, uzunluk

için minimum gereksinimlere göre örtüşme yeterliliğini, tanımlama yüzdesini ve benzerlik skorunu ölçmektedir. Dördüncü ölçüm, örtüşen bazların yüksek quality değerlerinin farklarını inceler. Beşinci ölçümde ise; Örtüşmenin içerdiği iki bölgenin dizileme hatası oranları ile ilgili örtüşmenin farklı oranları incelenmektedir (Huang ve ark., 1999).

4) Contiglerin oluşturulmasında kalıpların kullanımı 4 aşamada gerçekleşti;

-Başlangıç olarak diziler düzenlendi, bunun için örtüşme skorlarının en aza doğru sıralanmasına dayalı greedy yöntemi kullanıldı (Huang 1996a).

-Mevcut düzenlemenin kalitesi kontrol koşulları tarafından değerlendirildiği için mevcut düzenin her bir bölümüne ait koşul yeterliliğine dair bilgi toplandı.

-Mevcut düzenlemenin koşulları sağlayamayan herhangi bir bölümü olduğunda koşullara uygun hale getirilecek şekilde düzeltildi. Bu düzeltmeden sonra 2. Aşamaya geri dönüldü.

-Contigler kalıpların bağlanması ile oluştu.

(Huang ve ark., 1999)

5) Aligmentlerin (eşleşmelerin) oluşturulması ve consensus dizilerin (contig olacak olan geçici dizilerin) elde edilmesi (Huang ve ark., 1999)

Phred ve cross match programları ile fragment birleştirme işlemine hazır hale getirilen EST dizilerine Phrap ve CAP3 programları ile fragment birleştirme işlemi uygulandı. CAP3 programı default değerleri ile çalıştırılırken Phrap programının EST örtüşme değerleri 15,20,30 baz çifti şeklinde değerlerle denendikten sonra en uygun olarak (değer artıkça contig ve geri kalan singlet sayısının değişmemesi bakımından) 20 baz çiftilik örtüşme değeri kabul edilerek çalışıldı. 3734 EST dizisinin 2228 tanesi yaprak dokusundan izole edildiği için Yaprak EST'leri olarak adlandırılmıştır. Ve geri kalan 1506 EST de meyve dokusundan izole edildiği için meyve EST'leri olarak adlandırılmıştır. 3734 EST bir arada CAP3 ve Phrap'la çalıştırıldığı gibi meyve ve yaprak olarak ayrı ayrı da çalıştırılmış, sonuçlar kaydedilmiştir.

2.2.7 EST Dizilerinin Kontrolü

Fragment birleştirme işleminden önce vektör ve adaptör dizilerden Phred ve cross match programı ile tamamen kurtulduğu düşünülen EST dizilerinin bu istenmeyen dizilerden

kurtulup kurtulmadığını kontrol etmek için VecScreen programı ile sonuçlar değerlendirildi.

VecScreen bir nükleotid dizisindeki vektör kökenli segmenti tanımlamaya yarayan hızlı bir sistemdir. NCBI veritabanlarında bulunan dizilerdeki vektör kontaminasyonunu en aza indirmek amacıyla VecScreen' i geliştirmiştir. Bir dizi GENBANK' a dahil edilmeden önce VecScreen programı ile vektör dizisinden uzaklaştırılıp uzaklaştırılmadığı kontrol edilir.

VecScreen programı verilen herhangi bir dizinin vektör taşıyıp taşımadığını vektör dizilerine ait veritabanı ile karşılaştırma yaparak belirlemektedir. Bunun için BLAST algoritmasını kullanır. Eğer bir dizi içerisinde herhangi bir vektöre ait bir segment taşıyorsa VecScreen programı bu vektör kontaminasyonunun yerini ve karakterizasyonunu yapmaktadır. [5]

Yapılan kontrollerde EST dizilerimizin temiz olduğu anlaşıldıktan sonra fragment birleştirme işlemini başlatmak için bir engel kalmamıştır. CAP3 ve Phrap programı ile ayrı ayrı yapılan fragment birleştirme işleminin sonuçları da contig dizilerinin Consed programında incelenmesi ile kontrol edilmiştir (Şekil 2.4.).

Consed / Autofinish, Washington Üniversite'sinden Phil Green'in ekibinde bulunan David Gordon tarafından görüntülemek, düzenlemek ve dizi birleşimini sonuçlandırmak için phrap programı ile birlikte tasarlanmış bir programdır (CAP3 ile de uyumludur).

Consed programı ile sonuçlandırma işlemi, temelde 3 stratejiye bölünmüştür:

Ek veri aktarımı veya önemli düzenlemelerin yapılmasına veya diğer anomalilerin tespit edilmesine karar vermek amacıyla dizi birleşiminin ve verilerin görüntülenmesi;

Ek dizi verilerinin belirlenmesi;

Dizi birleşimindeki veya consensus dizideki hataların düzeltilmesi.

Süreç için birkaç tekrarlama gerektirse bile sonuçlandırma için en etkili olan programdır (özellikle büyük projelerde) (Gordon ve ark., 1998).

CAP3 programında “.contig” uzantılı dosyalar birleştirilmiş dizileri (contig dizilerini), “.singlet” uzantılı dosyalar hiçbir birleştirme işlemine dahil olmayan EST dizilerini, “.ace” uzantılı dosyalar ise Consed programında contig dizilerinin incelenmesi için kullanılacak dosyaları oluşturmaktadır. Phrap programında da aynı dosyalar bulunmakta ve aynı amaç için kullanılmaktadır. Consed programı yalnızca Linux tabanlı bilgisayarlarda çalışabilen

bir programdır. Bu programla beraber bilgisayara FASTA formatındaki EST dizilerinin bulunduğu bir klasörü ve fragment birleştirme işlemine ait programların output dosyalarını kaydetmek ve çalıştırmak yeterli olmuştur. Daha sonra hangi contig dizisinin hangi EST dizisini içerdiği kolaylıkla gözlemlendi. Zeytin EST'leri için CAP3 ve Phrap ile oluşturulan contig dizileri bu programla incelendiğinde Phrap programına ait contiglerin hatalı sonuçlar verebiliyorken CAP3 sonuçlarında bu durum gözlemlendi. Ve bu yüzden BLAST analizleri CAP3 sonuçları ile elde edilen contig dizileri ile yapıldı.

2.2.8 Blast Yöntemi İle Fonksiyonların Belirlenmesi

Genbank üzerinden en çok yapılan arama, dizi bilgisi benzerlik sorgulamasıdır. NCBI BLAST programlarıyla diziler arasındaki benzerlikleri bulunmasında yardımcı olmaktadır. BLAST ailesi programları, sorgulanan dizi ve veritabanındaki dizileri en iyi şekilde benzeşim yaptıran programlardır. BLAST araması NCBI web sitesinden E-posta servisi (blast@ncbi.nlm.nih.gov) ile veya FTP ile dağıtılan ve kendi başına çalışan programlar aracılığıyla yapılmaktadır. Vektör ve adaptör dizilerinden kurtulmuş ve anlamlı hale getirilmiş diziler daha önce fonksiyonu belirlenmiş ve veritabanına kaydedilmiş dizilerle BLAST programı kullanılarak karşılaştırılmıştır. Böylece dizilerin fonksiyonları hakkında tahmini değerlere ulaşılmıştır.

Bu çalışmada gen olması ihtimali olan, CAP3 programı aracılığıyla elde edilen contig dizileri için BLAST analizleri yapılmıştır. CAP3 contiglerinin BLASTX analizleri için BLAST2GO programı kullanılırken BLASTN analizleri NCBI'nın web sitesinde 69 meyve, 205 yaprak ve 299 toplam (tüm EST'lerden elde edilen contigler) contigler için tek tek yapılmıştır. Phrap contigleri için ise NCBI web sitesi kullanılarak BLASTN analizi yapılmıştır. Veritabanında kayıtlı olan zeytin EST'leri ile örtüşen contig dizileri BLASTN sonuçlarına göre belirlendi. Ve BLASTX analizi sonuçları BLAST2GO programının aracılığıyla gerçekleştirilip değerlendirilmiştir. Elde edilen grafikler incelenmiştir. Tüm BLAST analizleri ve BLAST2GO programı default değerlerine göre çalıştırılmıştır. (Her bir dizi için en fazla benzerlik gösteren 20 blast hit alınacak şekilde, blast e-value: 1e-3 olacak şekilde, Blast program olarak BLASTX seçilerek, Blast analizini GenBank'a kayıtlı dizilerle yapabilmek için: Blast mode olarak Qblast-NCBI seçilmiştir.)

Temel olarak B2G, girilen contig dizilere benzer dizileri bulmak için uzak (QBLAST-NCBI) Blast programını kullandı. Program belirlenen her hit dizi için GO terimlerini ayıkladı ve yüklenen dizilere ya da diziyeye ait verilen GO analizleri ile ilgili değerlendirmeleri yaptı. GO analizi gen Ontoloji ilişkilerinin yapısını görüntüleyebilmekte ve Enzim kodlarını KEGG haritaları üzerinde aydınlatmaktadır (Conesa ve ark., 2005).

Blast2GO kullanımı ile Blast analizi 4 adımdan oluşmuştur: Blast analizi, haritalama, analiz (annotation), istatistiksel analiz ve görselleştirme.

Genomik veri değerlendirmenin en önemli yönlerinden biri bireysel dizileri birbiri ile ilişkilendirmek ve biyolojik fonksiyonları ile ilgili bilgileri ifade etmektir. Otomatik işlevsel açıklama bu sorunu çözmek için etkili bir yaklaşımdır. Fonksiyonel analiz, dizi grupları arasındaki işlevsel farklılıkların değerlendirilmesi için çok yararlı olabilecek büyük gen gruplarının fizyolojik anlamlarının belirlenmesi ile fonksiyonel sınıflardaki genlerin kategorizasyonunu sağlar. Gen Ontoloji (GO) GO Konsorsiyumunda (Ashburner ve ark., 2000) geliştirilmiştir ve bu tür analizler için en uygun taslağı sağlamaktadır ayrıca biyoloji en geniş kapsamda kapalı ve kendi yönetimindeki doğrudan asiklik grafik (DAG) yapısı nedeniyle biyolojik içerik çerçevesinde görselleştirme sağlar. Farklı geliştirme ekipleri tarafından GO kullanımı ile dizileri analiz etmek için çeşitli yazılımlar geliştirildi. Çeşitli masaüstü ve web uygulamaları elektronik olarak, bilinmeyen dizilerin GO terimlerini atamak için benzerliğe dayalı olarak (Martin ve ark., 2004; Groth ve ark., 2004; Khan ve ark., 2003; Zehetner, 2003) veya gen analizleri çerçevesinde genomik verileri analiz etmek için kullanılabilir (Al-Shahrour ve ark., 2004; Doniger ve ark., 2003). Ancak, GO temelli analizleri gerçekleştirmek için çalışırken zayıf karakterize olmuş organizmalarda bir takım engellerle karşılaşıldı. Genel olarak, bu araçlar high-throughput dizilerinin analizleri için tasarlanmamıştır. Bunların depolanması ve görselleştirme yetenekleri sınırlıdır veya girişi yapılacak veri olarak sadece gen veya prob tanımlayıcıları kabul edilebilir, onları daha önceden veritabanlarında bulunan diziler ile kısıtlayabilir. Blast2GO (B2G), evrensel bir GO açıklama, canlandırma ve istatistikler çerçevesinde ortaya çıkan bu sınırlamalara uygun bir çözüm sağlamak için model olmayan türlerin genomik araştırmalarına gelişmiş bir fonksiyonel analiz oluşturmaktadır. B2G (1) otomatik ve high-throughput dizi analizine izin veren ve (2) analiz temelli veri değerlendirme için işlevselliğini entegre eden bir tasarıma sahiptir.

Bu çalışmada, B2G fasta formatında girilen dizilerin homologlarını bulmak için BLASTX programını kullandı. Program belirlenen her hite ait GO terimlerini haritalama ile mevcut analizlerin birleşimi için aldı. Analizde son olarak girilen dizi için GO terimleri atandı. Fonksiyonel analiz aşamaları GO ilişkilerinin grafiksel olarak oluşturulması ve ilgili alanlarda renklerle yapılan vurgulamalarla gözlemlenmiştir. B2G bir Java uygulaması olan Java Web Start ile sunulmuştur. Bu platformdan bağımsız olarak bir de Internet bağlantısı gereksinimi vardır (Conesa ve ark., 2005).

2.2.9 Sonuçların GenBank Veritabanına Gönderilmesi

Sonuçlar GenBank'a iki aşamada kaydedildi. Öncelikle işlenmiş 3734 EST dizisinin kaydı yapıldı. Daha sonra, CAP3 programı ile bu EST dizilerinden üç farklı aşamada elde edilen contig dizilerinin (69 adet meyve dokusuna ait EST'lerden üretilen contigler, 205 adet yaprak dokusuna ait EST'lerden elde edilen contigler ve toplam 3734 dizinin bir arada oluşturduğu 299 adet contig) kaydı yapıldı.

NCBI tarafından geliştirilen Sequin programı bir EST projesinde tüm çalışmalar tamamlandıktan sonra birleştirilmiş ve bir gen olma ihtimali olabilen contig dizilerinin GenBank veritabanına kaydedilmesini sağlayan programdır.

NCBI kendi başına çalışan, çoklu-platform özeliğine sahip ve NCBI ile internet üzerinden bağlantı kurabilen Sequin adında bir gönderim (submit) programı geliştirmiştir. Sequin; basit dizileri olduğu kadar BankIt ve diğer gönderim programlarının sahip olmadığı, uzun diziler ve parçalı diziler işleme özelliğine sahiptir. Sequin; uygun bir düzeltme, gelişmiş not ekleme yeteneği ve birçok doğrulama gibi geliştirilmiş kalite kontrol özelliklerine sahiptir. Ayrıca filogenetik, populasyon, mutasyon çalışmalarından gelen dizi bilgilerinin gönderimini kolaylaştırmak ve bu verilerin eşleşmiş dizilerle birleştirilmesini sağlamak üzere de tasarlanmıştır. Sequin ayrıca dizi bilgilerinin düzeltilmesi ve güncelleştirilmesi hatta dizi analizi yapma özelliklerine de sahiptir. Örnek olarak Sequin, girdi olarak FASTA ve ASN1 biçimlerini kullanan herhangi bir analiz aracıyla birlikte çalışabilmektedir. Bunlara ek olarak Sequin; büyük kayıtlarla (5.6 Mb boyutunda *Escherichia coli* genomu) çalışabilme ve tablolar aracılığıyla verilerin notlarını okuma yeteneğine sahiptir. PC, MAC ve Unix platformlarında çalışan

versiyonları 'ncbi.nlm.nih.gov' FTP adresinden ücretsiz olarak edinilebilir (Benson ve ark. , 2004).

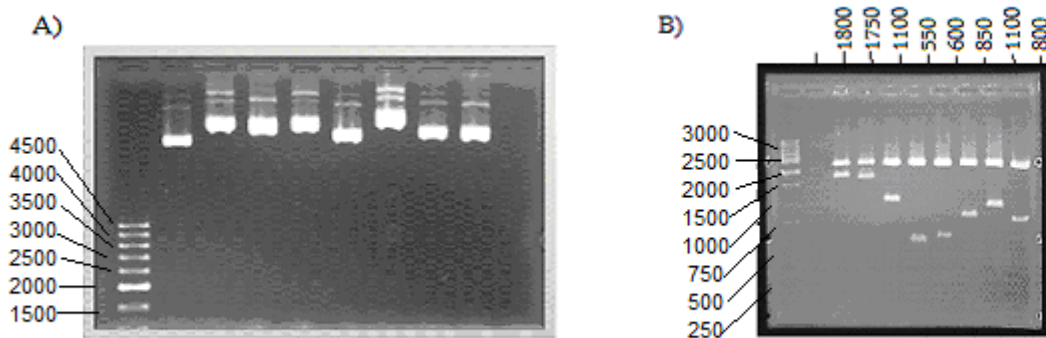
Elde edilen contig dizileri GenBanka gönderilmeden önce EST dizileri fragment birleştirme işleminden önceki şekilleriyle batch-sub@ncbi.nlm.nih.gov adresine dbEST veritabanına kaydedilmek üzere istenilen formata göre gönderildi. Gönderilen EST'ler için atanan accession numaraları yine mail yoluyla veritabanından elde edildikten sonra, consed programından elde edilen bilgilere göre hangi accession numarasına sahip hangi EST dizisinin hangi contiglerde yer aldığına dair bilgileri içeren bir dosya hazırlanmış ve hazırlanan bu dosya ile sequin programına kaydedilen contig dizileri için oluşturulan seq formatlı bir dosya gb-sub@ncbi.nlm.nih.gov adresine gönderildi. Bu şekilde contig dizileri de GenBanka kaydedilmiş oldu [13].

3. SONUÇLAR

3.1 CDNA KÜTÜPHANELERİNİN KARAKTERİZASYON SONUÇLARI

Farklı konsantrasyonlar ile oluşturulan titrasyon sonucunda sayılan klonilere göre yaprak dokusundan hazırlanan kütüphane yaklaşık 2.4×10^6 klondan oluştuğu ve meyve dokusundan elde edilen kütüphane ise yaklaşık 2.2×10^5 klondan oluştuğu bulunmuştur. Oluşturulan yaprak kütüphane için alkali lysis yöntemi ile 60 adet plasmid DNA izolasyonu yapılırken zeytin için 35 adet yapıldı (Şekil 9a). Klonların insört büyüklüğü yaprak dokusu için, 200 baz çifti ile 2500 baz çifti arasında değişirken ortalama insört büyüklüğü 1.6 kb olarak belirlendi (Şekil 9b). Zeytin meyvesi için insört büyüklüğü en az 70 baz çifti ile en fazla 1500 baz çifti aralığında değişirken ortalama büyüklüğün 1.1 kb olduğu belirlendi. Elde edilen sonuçlara göre yaprak kütüphanesi sadece % 0,14 oranında nonrekombinant klonlara sahipken meyve kütüphanesi % 5 nonrekombinant klona sahipti.

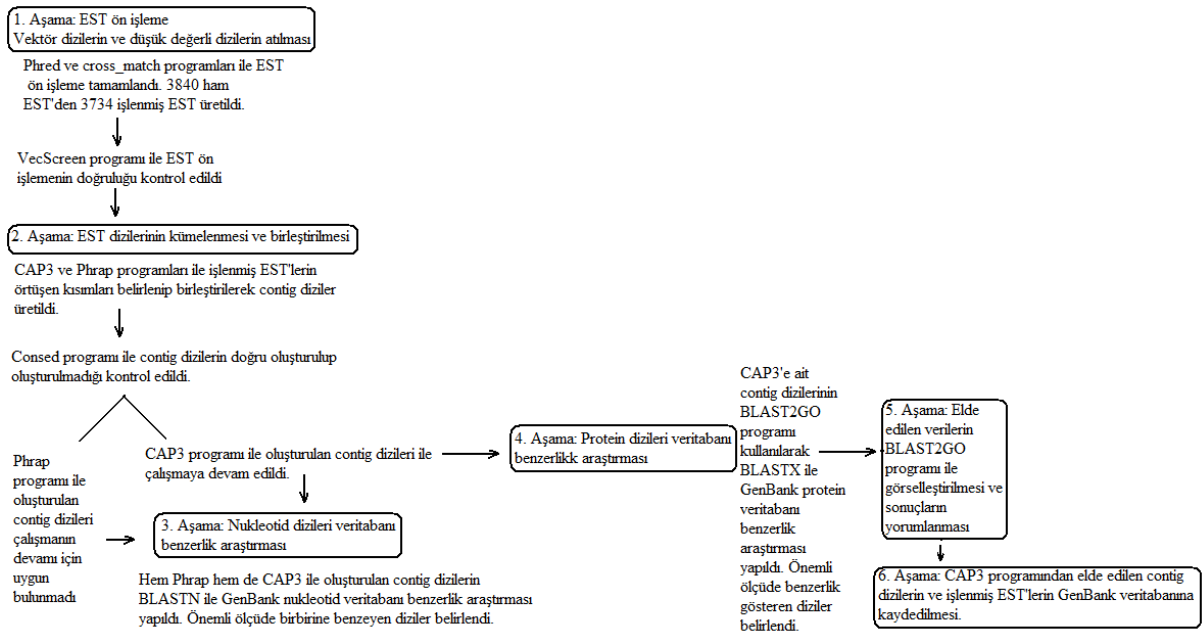
cDNA kütüphanelerinin, yapılan karakterizasyon çalışmaları sonucunda insört büyüklüğü ve klon sayısı açısından genom analizlerine uygun olduğu gözlemlendi. Yaprak kütüphanesinden 2304 klon, zeytin meyvesinden 1536 klon olmak üzere toplam tesadüfi seçilen 3840 adet klon EST koleksiyonu oluşturmak üzere dizin analizine gönderildi.



Şekil 3.1.: Plasmid DNA'ların jelde görünümü. A) Manuel plasmid DNA izolasyonu sonrası görülen DNA'lar B) Plasmid DNA'ların Bsp1407I restriksiyon enzimi ile kesimi

3.2 BİYOİNFORMATİK ANALİZLERİ

Phred ve Phrap (Phrap programının içindeki cross_match) program paketi ile vektör fragmentleri uzaklaştırılan EST dizileri için, Phrap ve CAP3 programları ayrı ayrı kullanılarak fragment birleştirilmesi (fragment assembly) uygulaması gerçekleştirildi. Elde edilen birleştirilmiş EST diziler (contigler) hakkında daha fazla bilgi sahibi olmak amacıyla NCBI web sitesindeki nükleotid dizileri ile benzerlik araştırılması yapıldı. Bu çalışmada yapılan biyoinformatik analizlerin özeti Şekil 3.2.'de görülmektedir (Şekil 3.2.).



Şekil 3.2.: EST analizindeki genel aşamaları. 1. Ham EST dizilerinin kesilen veya maskelenen vektör kontaminasyonu, düşük değerlikli diziler ve tekrar bölgeleri bakımından kontrol edilmesi. Ayrıca, düşük kaliteli, singleton olan ve kısa diziler de ayrılır. Bu çalışmada EST ön işleme Phred ve Cross_match programları ile yapılmıştır. Vecscreen programı ile EST ön işleme aşamasının doğruluğu kontrol edilmiştir. 2. Kümelenen ve birleştirilen EST'lerin contig (birleşmiş dizi) oluşturması. (contig = tahmini transkript) Hem CAP3 hem Phrap programı ile EST birleştirme yapılmıştır. Consed programı ile oluşturulan tüm contigler incelendiğinde çalışmaya devam etmek için CAP3 programının sonuçları daha uygun bulunmuştur. 3. Homolog olan ve benzer fonksiyona sahip olan dizileri belirlemek için DNA veritabanı benzerlik araması uygulanır. BLASTN ile GenBank veritabanı benzerlik araştırması hem CAP3 hem de Phrap programından elde edilen contig dizileri için yapılmıştır. 4. CAP3 programından elde edilen contig dizileri için protein veritabanı benzerlik araması

uygulanır. BLASTX ile GenBank protein veritabanı benzerlik araştırması yapıldı. Ayrıca BLASTX analizleri sırasında BLAST2GO programı kullanılmıştır. 5. Sonuçların görselleştirilmesi ve yorumu ile fonksiyonel anlamlandırmanın yapılması için analiz genişletilir. 6. CAP3 programından elde edilen contig dizilerin ve EST ön işleme aşamasını tamamlamış EST'lerin GenBank veritabanına kaydedilmesi için submit işlemleri yapıldı.

Phrap ve CAP3 programları ayrı ayrı bu 3 gruba uygulandı ve iki program sonuçları arasında farklılar gözlemlendi. Phrap programı ile toplam contig sayısı 924 ve singlet sayısı 1302 olarak bulundu (Çizelge 1). Aynı EST dizileri CAP3 programı ile fragment birleştirilmesi yapıldığında toplam contig sayısı 299 iken singlet sayısı 2368 olarak belirlendi (Çizelge 2).

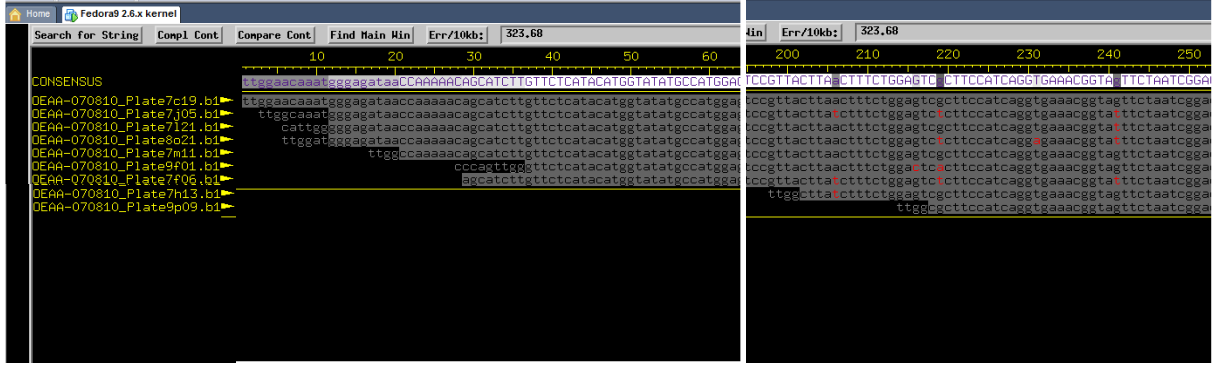
Çizelge 3.1.: PHRAP programı ile biyoinformatik analizleri sonucu kütüphaneden elde edilen değerler.

	YAPRAK	ZEYTİN (MEYVE)	TOPLAM
EST Sayısı	2228	1506	3734
Contig Sayısı	335	550	924
Singlet Sayısı	987	485	1.302
En Uzun Contig	2597	1909	2558
En Kısa Contig	393	461	393

Çizelge 3.2.: CAP3 programı ile biyoinformatik analizleri sonucu kütüphaneden elde edilen değerler.

	YAPRAK	ZEYTİN (MEYVE)	TOPLAM
EST Sayısı	2228	1506	3734
Contig Sayısı	205	69	299
Singlet Sayısı	1.591	887	2.368
En Uzun Contig	1924	1909	2018
En Kısa Contig	514	461	514

Çalışmalar tamamlandıktan sonra, Consed programı ile birleştirilmiş EST'ler incelendiğinde CAP3 programı ile oluşturulan contiglerin Phrap programı ile oluşturulan contiglere göre daha hatasız olduğu gözlemlendi. Bu yüzden CAP3 programına ait sonuçların Gemlik zeytin bitkisine ait cDNA kütüphanesini en hatasız şekilde yansıttığına karar verildi (Şekil 3.3.).



Şekil 3.3.: Consed sonuçları: Meyve dokusundan elde edilen EST'lerin CAP3 programı ile birleştirilmesi sonucu 69 contig üretilmiştir. 9. Contigi meydana getiren 9 EST'nin birleşme şekli Consed programı ile şekildeki gibi gözlenmiştir.

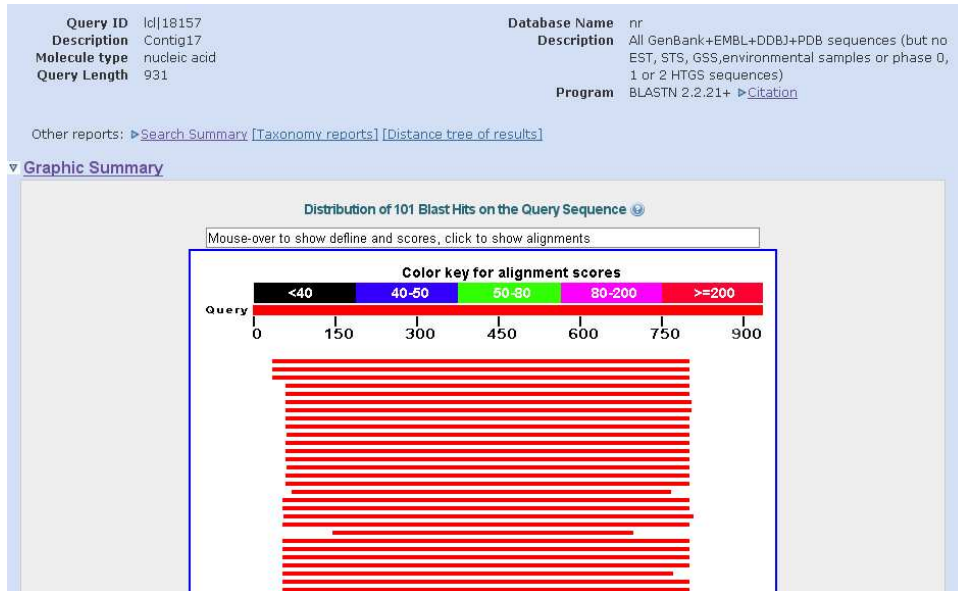
3.2.1 BLAST Analizi Sonuçları

Elde edilen contiglerin GenBank veritabanı ile karşılaştırılabilmesi için BLASTN ve BLASTX programlarından yararlanıldı. BLAST programlarını çalıştırmak ve elde edilen sonuçların grafiklerini sağlamak amacıyla BLAST2GO otomatik BLAST programı kullanıldı. BLAST2GO programı sayesinde hem BLAST programları ile karşılaştırma yapıldı hem de elde edilen sonuçlar grafiksel olarak gösterildi. Benzerlik araştırması, birleştirilmemiş ham EST dizilerinin meyve, yaprak ve toplam (elde edilen bütün EST diziler) olarak 3 ayrı gruba ayrılmasından sonra yapıldı. BLAST analizleri okunurken, bu analiz sonuçlarının verdiği değerler dikkate alınır. Birbirleri ile BLAST programı kullanılarak karşılaştırılan iki dizinin birbirlerine benzeyen dizi uzunluklarının değerini kesirle ifade eden değer "Query coverage" olarak bilinir. Örneğin; İki diziden veritabanına ait olmayan dizinin uzunluğu 100 nükleotidse ve bu dizinin veritabanındaki bir diziyile kesintisiz olarak benzeyen nükleotidlerinin sayısı 50 nükleotid uzunluğundaysa bu dizi için Query coverage değeri %50 olarak ifade edilmektedir. BLAST sonuçlarını değerlendirirken karşımıza çıkan farklı değer ve tanımlar mevcuttur. Örneğin; Query coverage değerinde ifade edilen benzer uzunlukların birbirlerine benzeme oranlarını veren değer "Max identity"dir. "Accession Number", veritabanındaki dizilerin tanınmasını sağlayan her bir dizi için farklı olan numaradır. "Description" terimi, contiglerin veritabanında benzerlik gösterdiği dizinin fonksiyonuna ve ait olduğu organizmaya ait bilgiler içeren kısımdır. Son olarak "E-value" değeri, NCBI'da BLAST programları tarafından karşılaştırılan iki dizinin bir pozisyonda aynı nükleotide sahip olma şansını veren değerdir.

BLASTX analizleri için BLAST2GO programı kullanılırken BLASTN analizleri

<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi?PAGE=Nucleotides> web sitesinden her bir contig için tek tek yapıldı. Şekil 3.4.'de BLASTN analizlerinin sonuçlarından alınan bir örnekte meyve dokusuna ait EST'lerden CAP3 programı ile oluşturulan contiglerden "Contig 17"'nin sonuçları görülmektedir (Şekil 3.4.).

A)



B)

Descriptions

Legend for links to other resources: [U](#) UniGene [G](#) GEO [G](#) Gene [S](#) Structure [M](#) Map Viewer

Sequences producing significant alignments:
(Click headers to sort columns)

Accession	Description	Max score	Total score	Query coverage	E value	Max ident	Links
AY462246.1	Rehmannia glutinosa ascorbate peroxidase (APX) mRNA, complete	762	762	81%	0.0	84%	U G G G G G G G G G G G G G G G G G G G G G G G G G
EU870515.1	Picrorhiza kurrooa ascorbate peroxidase (apx) mRNA, complete cds	681	681	81%	0.0	82%	
EF031031.1	Acanthus ebracteatus ascorbate peroxidase mRNA, complete cds	640	640	81%	3e-180	81%	
AY078080.1	Capsicum annuum putative ascorbate peroxidase mRNA, complete	608	608	79%	2e-170	81%	
X81376.1	C.annuum mRNA for cytosolic ascorbate peroxidase	590	590	79%	3e-165	81%	
U15933.1	Nicotiana tabacum ascorbate peroxidase (APX) mRNA, complete cd	575	575	79%	1e-160	80%	
D85912.1	Nicotiana tabacum mRNA for cytosolic ascorbate peroxidase, compl	564	564	79%	4e-157	80%	
AK320564.1	Solanum lycopersicum cDNA, clone: LEFL1010DB09, HTC in leaf	562	562	79%	2e-156	80%	
AK322885.1	Solanum lycopersicum cDNA, clone: LEFL1045AA09, HTC in leaf	544	544	79%	3e-151	79%	
DQ222519.1	Solanum tuberosum clone 121A05 ascorbate peroxidase-like mRNA	544	544	78%	3e-151	79%	
DQ096286.1	Lycopersicon esculentum cytosolic ascorbate peroxidase 1 (APX1) r	544	544	79%	3e-151	79%	
DQ002888.1	Capsicum annuum cytosolic ascorbate peroxidase (APX1) mRNA, cc	544	544	79%	3e-151	79%	
AY974805.1	Lycopersicon esculentum ascorbate peroxidase (APX) mRNA, compl	544	544	79%	3e-151	79%	
AB041343.1	Solanum tuberosum APx mRNA for ascorbate peroxidase, complete	544	544	78%	3e-151	79%	
AK246586.1	Solanum lycopersicum cDNA, clone: FC16AB03, HTC in fruit	539	539	79%	1e-149	79%	
DQ096287.1	Lycopersicon esculentum cytosolic ascorbate peroxidase 2 (APX2) r	539	539	79%	1e-149	79%	
FJ532356.1	Solanum lycopersicum cultivar Micro-Tom cytosolic ascorbate perox	514	514	74%	5e-142	79%	
AF159630.1	Fragaria x ananassa cytosolic ascorbate peroxidase (APX) mRNA, A	504	504	79%	4e-139	78%	
AF022121.1	Fragaria x ananassa cytosolic ascorbate peroxidase (APX-c) mRNA,	504	504	79%	4e-139	78%	
EU547804.3	Camellia sinensis ascorbate peroxidase mRNA, complete cds	496	496	80%	7e-137	78%	
AF159629.1	Fragaria x ananassa cytosolic ascorbate peroxidase (APX) mRNA, A	492	492	79%	1e-135	78%	
AB434928.1	Capsicum chinense CcAPX mRNA for ascorbate peroxidase, partial	471	471	58%	3e-129	81%	
AF159628.1	Fragaria x ananassa cytosolic ascorbate peroxidase (APX) mRNA, A	469	469	79%	9e-129	78%	
AF159627.1	Fragaria x ananassa cytosolic ascorbate peroxidase (APX) mRNA, A	469	469	79%	9e-129	78%	
AF039953.1	Fragaria x ananassa cytosolic ascorbate peroxidase (ApX-SC) mRNA	469	469	79%	9e-129	78%	
AF159631.1	Fragaria x ananassa cytosolic ascorbate peroxidase (APX) mRNA, A	464	464	79%	5e-127	77%	

Çizelge 3.3.: Meyve dokusuna ait EST'lerden CAP3 programı ile üretilen 69 contigin <http://www.ncbi.nlm.nih.gov/blast/Blast.cgi?PAGE=Nucleotides> web sitesinde yapılan BLASTN analizlerinin sonuç çizelgesi.

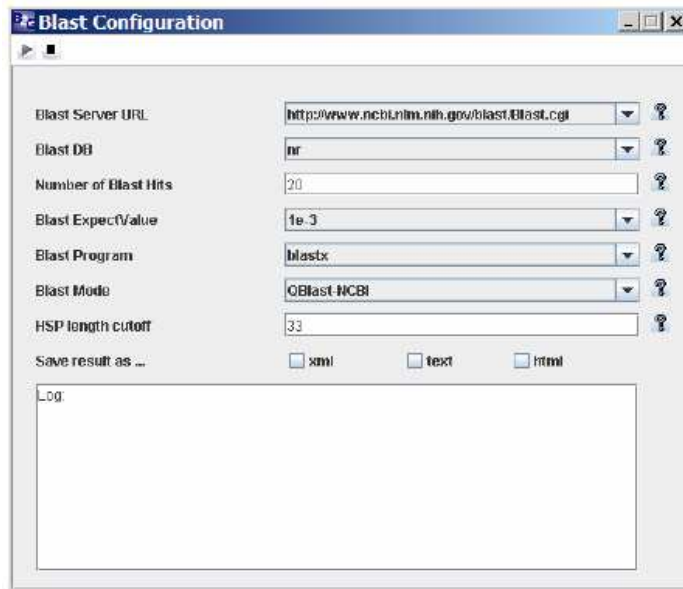
Contig No	Accession	Description	Query coverage	E value	Max ident
Contig 1	XM_002262882.1	PREDICTED: Vitis vinifera hypothetical protein LOC100249242 (LOC100249242), mRNA	4%	0.054	87%
Contig 2	AC203552.6	Medicago truncatula clone mth2-171k22, complete sequence	36%	9e-30	78%
Contig 3	XM_002272313.1	PREDICTED: Vitis vinifera hypothetical protein LOC100264565, transcript variant 2 (LOC100264565), mRNA	49%	3e-15	70%
Contig 4	EF506503.1	Olea europaea putative cytochrome P450 mRNA, partial cds	28%	0.0	98%
Contig 5	AY518204.1	Suaeda maritima subsp. salsa CAX (cax) mRNA, complete cds	21%	1e-24	78%
Contig 6	AC150798.3	Medicago truncatula chromosome 2 clone mth2-33b11, complete sequence	11%	9e-08	77%
Contig 7	BT012880.1	Lycopersicon esculentum clone 113981F, mRNA sequence	71%	2e-141	80%
Contig 8	EF554358.1	Vitis vinifera GDP-mannose-3',5'-epimerase mRNA, complete cds	71%	9e-177	83%
Contig 9	AC215367.2	Solanum lycopersicum chromosome 2 clone C02HBa0044L14, complete sequence	24%	8e-12	84%
Contig 10	XM_002281739.1	PREDICTED: Vitis vinifera hypothetical protein LOC100242675 (LOC100242675), mRNA	11%	3e-15	83%
Contig 11	AY619952.1	Nicotiana tabacum putative UDP-glucuronate decarboxylase 3 mRNA, complete cds	81%	0.0	84%
Contig 12	XM_002304452.1	Populus trichocarpa tubulin, beta chain (TUB4), mRNA	81%	2e-174	81%
Contig 13	XM_002304398.1	Populus trichocarpa predicted protein, mRNA	52%	3e-114	85%
Contig 14	AF259982.1	Ipomoea batatas putative aspartic protease mRNA, complete cds	43%	3e-66	78%
Contig 15	AK322569.1	Solanum lycopersicum cDNA, clone: LEFL1039BE10, HTC in leaf	9%	1e-14	84%
Contig 16	DQ250022.1	Petunia integrifolia subsp. inflata S-RNase-binding protein (SBP1) mRNA, complete cds	25%	4e-21	74%
Contig 17	AY462246.1	Rehmannia glutinosa ascorbate peroxidase (APX) mRNA, complete cds	81%	0.0	84%

Contig 18	CU225525.1	Populus EST from leave	56%	3e-55	74%
Contig 19	AY083162.1	Olea europaea subsp. europaea beta-glucosidase (bglc) mRNA, complete cds	98%	0.0	94%
Contig 20	AF429430.1	Olea europaea polyubiquitin OUB2 mRNA, complete cds	87%	0.0	92%
Contig 21	X97131.1	Forsythia x intermedia mRNA for EF-1-alpha	69%	0.0	89%
Contig 22	AJ843980.1	Plantago major mRNA for hypothetical protein	35%	2e-45	84%
Contig 23	AK329671.1	Solanum lycopersicum cDNA, clone: LEFL3153B15, HTC in root	50%	3e-70	79%
Contig 24	AC139571.2	Mus musculus BAC clone RP24-73K9 from chromosome 15, complete sequence	8%	8e-12	97%
Contig 25	AY818399.1	Manihot esculenta chloroplast latex aldolase-like protein mRNA, partial cds; nuclear gene for chloroplast product	76%	0.0	81%
Contig 26	XM_002277840.1	PREDICTED: Vitis vinifera hypothetical protein LOC100264510 (LOC100264510), mRNA	7%	2e-09	86%
Contig 27	DQ335783.1	Medicago truncatula cytochrome P450 monooxygenase CYP72A59 (CYP72A59) mRNA, complete cds	3%	7e-05	90%
Contig 28	AK325906.1	Solanum lycopersicum cDNA, clone: LEFL2002DE10, HTC in fruit	46%	2e-100	83%
Contig 29	AK322165.1	Solanum lycopersicum cDNA, clone: LEFL1034BD04, HTC in leaf	41%	2e-89	81%
Contig 30	AK320385.1	Solanum lycopersicum cDNA, clone: LEFL1008CE05, HTC in leaf	40%	2e-74	82%
Contig 31	FN357297.1	Schistosoma mansoni genome sequence supercontig Smp_scaff000006	3%	0.20	96%
Contig 32	BT012698.1	Lycopersicon esculentum clone 113569R, mRNA sequence	83%	0.0	82%
Contig 33	XM_002283204.1	PREDICTED: Vitis vinifera hypothetical protein LOC100251643 (LOC100251643), mRNA	24%	6e-35	79%
Contig 34	AK322020.1	Solanum lycopersicum cDNA, clone: LEFL1032BH11, HTC in leaf	32%	2e-30	82%
Contig 35	XM_002263142.1	PREDICTED: Vitis vinifera hypothetical protein LOC100244978 (LOC100244978), mRNA	27%	8e-23	88%
Contig 36	AB236867.1	Panax ginseng cab mRNA for chlorophyll a/b binding protein, complete cds	86%	1e-178	82%
Contig 37	XM_002263632.1	PREDICTED: Vitis vinifera hypothetical protein LOC100263206 (LOC100263206), mRNA	62%	2e-96	78%
Contig 38	EF506512.1	Olea europaea putative metallophosphatase/diphosphonucleotide phosphatase	28%	2e-126	96%

Contig	Accession	Description	Identity	E-value	Score
		1 mRNA, partial cds			
Contig 39	AJ011621.1	<i>Antirrhinum majus</i> mRNA for squamosa promoter binding protein-homologue 3	5%	8e-08	86%
Contig 40	BT013631.1	<i>Lycopersicon esculentum</i> clone 132433R, mRNA sequence	22%	7e-45	83%
Contig 41	AC215482.2	<i>Solanum lycopersicum</i> chromosome 2 clone C02SLm0064E19, complete sequence	55%	2e-70	81%
Contig 42	U82976.1	<i>Citrus sinensis</i> pectinesterase mRNA, complete cds	29%	6e-42	79%
Contig 43	XM_002269228.1	PREDICTED: <i>Vitis vinifera</i> hypothetical protein LOC100259768 (LOC100259768), mRNA	61%	1e-127	81%
Contig 44	AF452012.1	<i>Petunia x hybrida</i> aquaporin-like protein (PIP2;1) mRNA, complete cds	70%	2e-141	77%
Contig 45	XM_002269734.1	PREDICTED: <i>Vitis vinifera</i> hypothetical protein LOC100251163 (LOC100251163), mRNA	14%	3e-14	77%
Contig 46	XM_002321566.1	<i>Populus trichocarpa</i> anion exchanger family protein, mRNA	83%	4e-168	80%
Contig 47	AF249917.1	<i>Perilla frutescens</i> ADP-glucose pyrophosphorylase large subunit gene, complete cds	95%	0.0	82%
Contig 48	AK246885.1	<i>Solanum lycopersicum</i> cDNA, clone: FC06CB09, HTC in fruit	21%	7e-23	77%
Contig 49	CT009651.1	<i>Medicago truncatula</i> chromosome 5 clone mtel-70c15, COMPLETE SEQUENCE	4%	6e-05	90%
Contig 50	EF506513.1	<i>Olea europaea</i> putative ribulose-1,5-bisphosphate carboxylase/oxygenase activase mRNA, partial cds	68%	0.0	95%
Contig 51	AK319177.1	<i>Solanum lycopersicum</i> cDNA, clone: LEFL1094CD10, HTC in leaf	11%	9e-15	81%
Contig 52	AP011522.1	<i>Solanum lycopersicum</i> DNA, chromosome 8, clone: C08SLm0119G07, complete sequence	2%	0.18	100%
Contig 53	XM_002264712.1	PREDICTED: <i>Vitis vinifera</i> hypothetical protein LOC100261543 (LOC100261543), mRNA	32%	1e-54	80%
Contig 54	EF666999.1	<i>Salvia miltiorrhiza</i> putative aldo/keto reductase 2 mRNA, complete cds	56%	4e-72	77%
Contig 55	AJ299254.1	<i>Nicotiana tabacum</i> mRNA for putative DNAJ protein	66%	6e-93	81%
Contig 56	BT014212.1	<i>Lycopersicon esculentum</i> clone 133389F, mRNA sequence	75%	8e-140	80%
Contig 57	AF384040.1	<i>Mentha x piperita</i> farnesyl diphosphate synthase mRNA, complete cds	75%	5e-171	81%
Contig 58	AM113862.1	<i>Plantago major</i> partial mRNA for glycosyltransferase (qual gene)	74%	0.0	85%
Contig 59	AJ011840.2	<i>Catharanthus roseus</i> mRNA for 1-deoxyxylulose 5-phosphate synthase	46%	9e-48	75%

Contig 60	AK325009.1	Solanum lycopersicum cDNA, clone: LEFL1090AD06, HTC in leaf	56%	2e-63	75%
Contig 61	XM_002265202.1	PREDICTED: Vitis vinifera hypothetical protein LOC100246653 (LOC100246653), mRNA	16%	8e-19	77%
Contig 62	XM_002279756.1	PREDICTED: Vitis vinifera hypothetical protein LOC100245436 (LOC100245436), mRNA	62%	4e-153	79%
Contig 63	AY333925.1	Antirrhinum majus ACC oxidase AC01 mRNA, partial cds	48%	0.0	82%
Contig 64	XM_002307260.1	Populus trichocarpa predicted protein, mRNA	39%	4e-39	74%
Contig 65	AK324895.1	Solanum lycopersicum cDNA, clone: LEFL1088AF05, HTC in leaf	36%	6e-16	80%
Contig 66	AK247621.1	Solanum lycopersicum cDNA, clone: LEFL1067BB05, HTC in leaf	42%	3e-33	73%
Contig 67	AK321700.1	Solanum lycopersicum cDNA, clone: LEFL1027DE05, HTC in leaf	66%	3e-150	82%
Contig 68	XM_002323542.1	Populus trichocarpa Ca ²⁺ antiporter/cation exchanger, mRNA	8%	9e-15	86%
Contig 69	DQ191635.1	Solanum tuberosum serine/threonine protein kinase SAPK8-like protein mRNA, complete cds	67%	5e-145	83%

Çizelge 3.3.'de görüldüğü gibi elde edilen sonuçların büyük ve karmaşık oluşu nedeniyle bu sonuçların grafiklerle kolayca anlaşılır şekilde sınıflandırılması için BLAST2GO programından tüm BLASTX analizlerinde default parametreleri kullanılarak yararlanılmıştır (Şekil 3.5.).



Şekil 3.5.: BLAST2GO default değerleri: Her bir dizi için en fazla benzerlik gösteren 20 blast hit alınacak şekilde, blast e-value: 1e-3 olacak şekilde seçilmiştir.

3.2.2 Gen Ontoloji Çalışmaları Sonuçları

BLAST2GO elde edilen BLAST sonuçları gen ontoloji sınıflarına göre grafiklendirilmektedir. Gen ontoloji (GO) bir gen veya gen ürününü tanımlayarak sahip olduğu fonksiyon hakkında bilgi veren bir oluşumdur. GO içerisinde yer alan gen ürünleri ve/veya genler hiyerarşik bir yapı içerisinde görevlerine göre sıralanırlar. Bu hiyerarşik yapının en üstündeki 3 temel sınıf; moleküler fonksiyon (Molecular Function-MF-), hücre bileşen (Cellular Component-CC-) ve biyolojik proses (Biological Process)'dir. Bu çalışmada BLAST2GO programı ile elde edilen sonuçlara ait grafikler verilmiştir (Şekil 3.8., Şekil 3.9., Şekil3.10., Şekil 3.14., Şekil 3.15., Şekil 3.16., Şekil 3.19., Şekil 3.20., Şekil 3.21.).

3.2.2.1 Yaprak Dokusuna Ait Contiglerin BLAST Çalışması Sonuçları

Yaprak dokusuna ait 2228 EST dizisinden 637 tanesi CAP3 programı analizleri sonucu birleşme işlemine katılarak kendi aralarında 205 contig oluşturmuşlardır. Bu 205 yaprak contig dizisinden 10 contig veritabanındaki çeşitli zeytin (*Olea europaea*) dizileri ile benzerlik göstermiştir (Çizelge 3.4.). Buna göre yaprak dokusunda görülen ve zeytin bitkisinde daha önce belirlenmiş olan bu 10 protein; oksidoreduktaz ailesine ait, NAD⁺ veya NADP⁺ için akseptör olarak CH-OH grupları üzerinde etkili olan mannitol sentezleyici mannitol dehidrogenaz, fotosentez reaksiyonları sırasında su molekülünden elektronun açığa çıktığı fotosistem II evresinde görevli polipeptit “photosystem II 10 kDa polypeptide”, bitki büyümesi, tuz basıncı ve ozmatik basınçla ilgili olan FMN (flavin mononucleotid) kullanarak hidroksi asitlerin katalizasyonunu ve oksidasyonunu gerçekleştiren peroksimal bir protein olan glikolat oksidaz domaini “glycolate oxidase-like FMN-binding domain protein” (fotosentez miktarını azaltan bir protein), yağ transferi proteinleri (LTPs) olarak bilinen, yağ asidi grupları ile birlikte fosfolipitlerin hücre membranından taşınmasında sorumlu olan, açıl (acyl) gruplarını da bağlayabilen “plant lipid transfer protein”, Rubisco adıyla bilinen Calvin döngüsünde karbon fiksasyonunu kataliz eden, yaprak dokusunda en fazla bulunan protein “ribulose-1,5-bisphosphate carboxylase/oxygenase activase”, İki glikoz molekülünü β 1->4 bağı ile bağlayan glukozidaz enzimi “beta-glucosidase”, bitkilerde depolayıcı vakuolar membran proteini “tonoplast intrinsic protein”, ökaryotik hücrelerde iki izopeptit formu tek bir monomer olarak bulunan, 76 ubiquitin protenin birleşiminden oluşmuş, hücreler arası

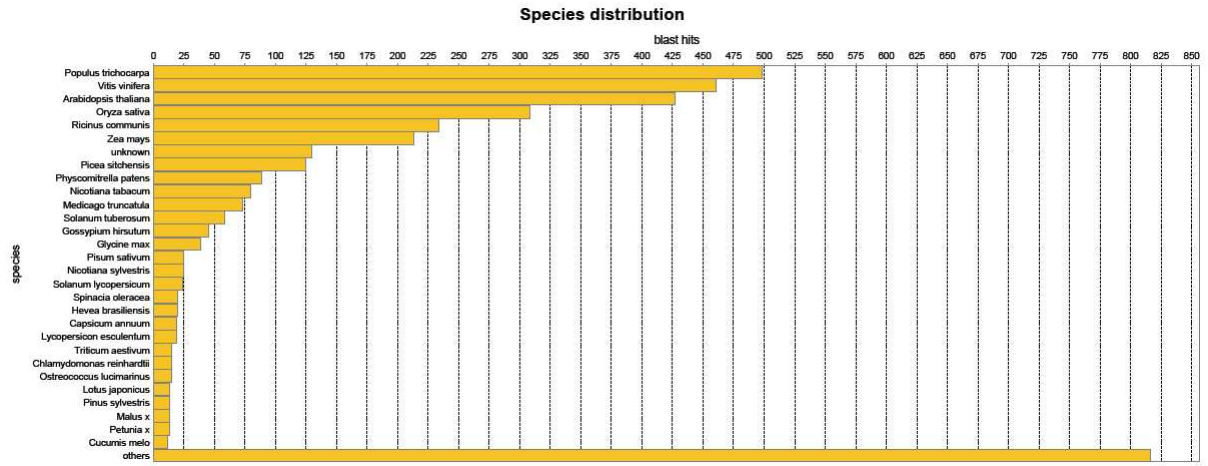
sinyalleri iletmek ve geniş protein ailelerini bağlamak için fonksiyonel “polyubiquitin OUB2”, daha önce zeytinde belirlenmiş bazı diziler, glukoneogenesis içinde küçük molekkülerden glukozun sentezlenmesinde görev alan enzim “glyoxisomal malate dehydrogenase” şeklinde sıralanmaktadır (Çizelge 3.4.).

Çizelge 3.4.: CAP3 programından elde edilen yaprak dokusuna ait contig dizilerinin BLAST sonuçlarına göre GenBanktaki zeytin dizileri ile benzerlik gösteren contigler

Contig no	Veritabanında eşleştiği zeytin dizisi	Query coverage	Max ident	Contig uzunluğu	İçerdiği EST sayısı
Contig 1	<i>Olea europaea</i> putative mannitol dehydrogenase 1 (MTD1) mRNA, partial cds	%71	%84	1069	2
Contig 7	<i>Olea europaea</i> photosystem II 10 kDa polypeptide mRNA, partial cds	%30	%99	724	4
Contig 16	<i>Olea europaea</i> putative glycolate oxidase-like FMN-binding domain protein mRNA, partial cds	%34	%99	1899	7
Contig 79	<i>Olea europaea</i> putative plant lipid transfer protein mRNA, partial cds	%25	%100	914	5
Contig 102	<i>Olea europaea</i> putative ribulose-1,5-bisphosphate carboxylase/oxygenase activase mRNA, partial cds	%43	%98	1797	19
Contig 130	<i>Olea europaea</i> subsp. <i>europaea</i> beta-glucosidase (bglc) mRNA, complete cds	%92	%96	1850	6
Contig 146	<i>Olea europaea</i> tonoplast intrinsic protein (tip) mRNA, complete cds	%87	%86	911	2
Contig 148	<i>Olea europaea</i> polyubiquitin OUB2 mRNA, complete cds	%85	%87	1216	4
Contig 159	<i>Olea europaea</i> 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 192	<i>Olea europaea</i> putative glyoxisomal malate dehydrogenase mRNA, partial cds	%46	%97	1383	2

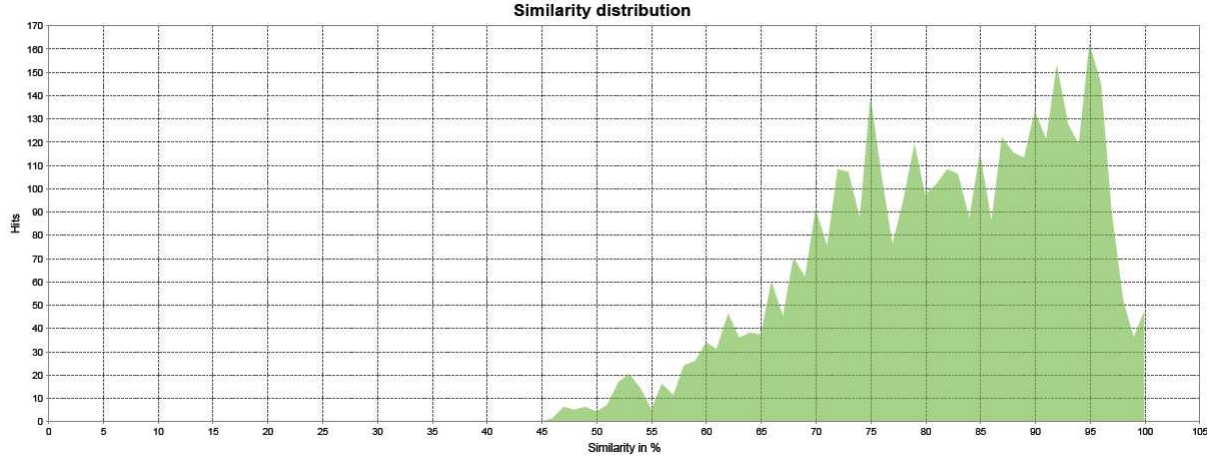
Şekil 3.6.’de BLAST2GO programının “Species distribution” olarak grafiksel verdiği karşılaştırma sonucu, yaprak dokusuna ait 2228 EST dizisinden elde edilen 205 contigin BLASTX analizlerine göre benzerlik gösterdiği diğer organizmaları sıralamaktadır. “blast hits” olarak adlandırılan parametre, her bir contig için GenBank veritabanından alınan, ve eldeki diziye en çok benzeyen ilk 20 diziyi vermektedir. 205 contig için; $205 \times 20 = 4100$ blast hits bulunmaktadır. “Species” parametresi ise, *Olea Europaeae* cv. *Gemlik* yaprak dokusuna

ait contig dizileri ile benzer dizilere sahip diğer organizmaları göstermektedir. Buna göre; yaprak dokusu contiglerinde *Olea Europaeae cv. Gemlik* organizması ile en fazla benzerlik gösteren organizma “*Populus trichocarpa*” yani “Batı Balsam Kavağı” olmuştur. İkinci sırada, “*Vitis vinifera*” (Üzüm) ve üçüncü sırada “*Arabidopsis thaliana*” yer alır. *Populus trichocarpa* ve *Vitis vinifera* bitkileri *Olea europaeae* bitkisi gibi dikotileden bitki olduğu için ve GenBankta kayıtlı dizi sayıları çok fazla olması nedeniyle benzerliğin en fazla olduğu düşünülmektedir. Buna karşılık *Arabidopsis thaliana* bitkisi monokotil bir bitki olduğu halde benzerliğin yüksek çıkması da zeytin bitkisi ile benzerliğin düşük çıkması aynı şekilde GenBanktaki dizi sayısı çok fazla veya çok az olması ile açıklanabilir. (Şekil 3.6.).



Şekil 3.6.: Yaprak için Species distribution: CAP3 programından elde edilen yaprak dokusuna ait contig dizilerinin BLASTX sonuçlarına göre GenBanktaki diğer organizmalarla olan benzerliği

Şekil 3.7.’de BLAST2GO programının “Similarity distribution” olarak verdiği sonuç, yaprak dokusuna ait 2228 EST dizisinden elde edilen 205 contigın BLASTX analizlerine göre benzerlik gösterdiği GenBank veritabanına kayıtlı blast hitlerle örtüşen nükleotid dizilerinin nükleotid sayısı bakımından benzerlik ortalamasını göstermektedir. “Hits” olarak adlandırılan değer blast hitlerini, “similarity” olarak adlandırılan değer örtüşen nükleotid sayısının tüm contig uzunluğuna oranını “%” cinsinden vermektedir (Şekil 3.7.).

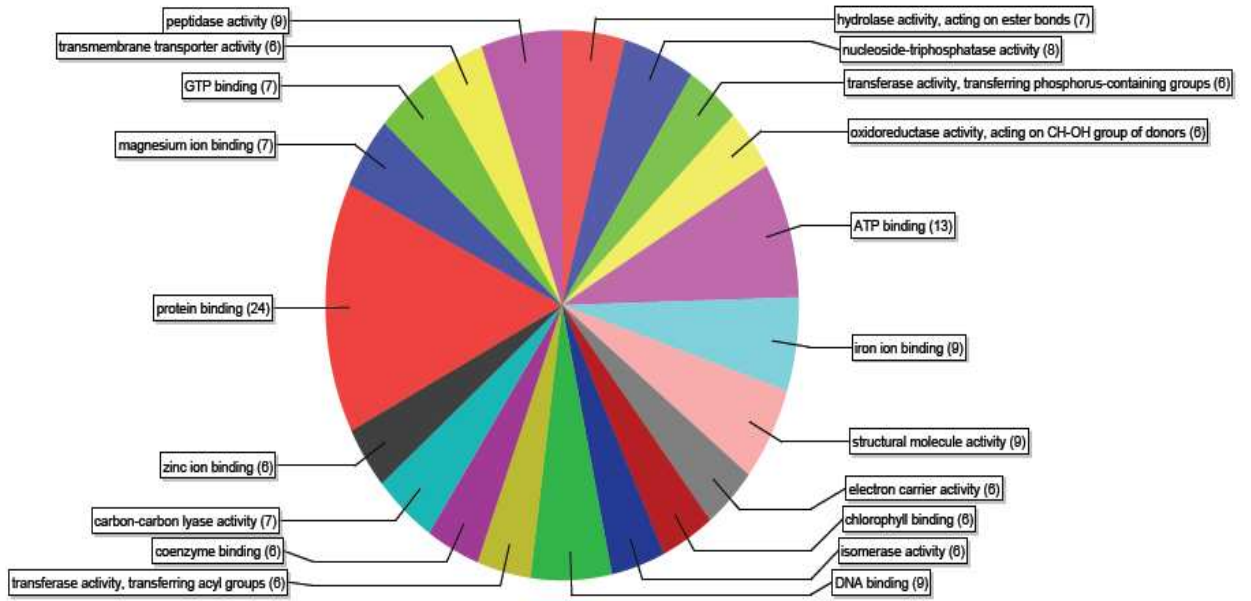


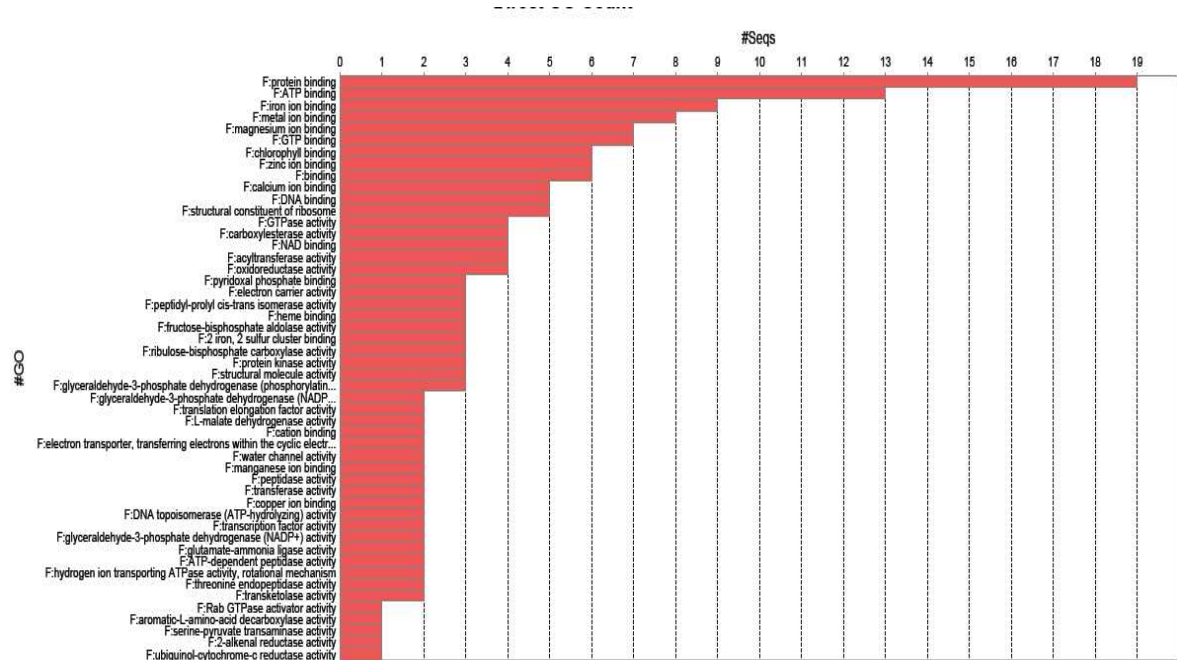
Şekil 3.7.: CAP3 programından elde edilen yaprak dokusuna ait contig dizilerinin BLASTX sonuçlarına göre GenBanktaki diğer nükleotid dizileri ile olan benzerlik oranı

Default değerleri ile çalıştırılan BLAST2GO programı ile yapılan BLASTX analizleri sonuçlarına göre contigler; ifade ettikleri gen ontoloji terimlerine göre sınıflandırılmıştır. Şekil 3.8.'deki dairesel grafikte yaprak dokusuna ait 2228 EST dizisinden CAP3 programı ile oluşturulan 205 contigin BLASTX analizlerinden elde edilen gen ontoloji değerlerinin moleküler fonksiyonlarına göre sınıflandırılması gösterilmiştir. Buna göre 205 contig içerisinde en fazla görülen fonksiyon 24 contigde belirlenmiş olan “protein binding” (protein bağlayıcı) aktivitesidir. Protein bindingi 13 contigle “ATP binding” (ATP bağlayıcı) aktivitesi izler. Daha sonra dokuzar contigle; “DNA binding” (DNA bağlayıcı), “Structural molecule activity” (yapısal molekül aktivitesi), “İron ion binding” (demir iyonu bağlayıcı), “Peptidase activity” (protein peptidlerini parçalayan peptidaz enzimi) şeklinde fonksiyonlara sahip gen ontoloji terimleri yer alır. Belirlenen diğer fonksiyonlar ise şunlardır; 8 contig “Nucleoside-triphosphatase activity” (Purin ve thiamine metabolizmasında yer alan hidrolaz grubundan nükleozid trifosfatları hidroliz eden enzim aktivitesi) için, yedişer contig “Carbon carbon lyase activity” (iki karbon atomu arasındaki bağı parçalayan enzim aktivitesi), “Hydrolase activity, acting on ester bonds” (yağ moleküllerinde ester bağları üzerinde etkili olan, ester bağlarını hidroliz eden enzim aktivitesi), “GTP binding” (Güanozin trifosfat bağlayıcı), “Magnesium ion binding” (magnezyum iyonu bağlayıcı) için, altışar contig “Coenzyme binding” (koenzim bağlayıcı), “Transferase activity transferring acyl groups” (açıl gruplarının transferini sağlayan enzim aktivitesi), “Chlorophyl binding” (klorofil bağlayıcı), “Electron carrier activity” (elektron taşıyıcı aktivite), “Zinc ion binding” (çinko iyonu bağlayıcı), “Oxidoreductase activity acting on CH-OH group of donors” (CH-OH gruplarının elektron

transferini sağlayan enzim aktivitesi), “Transferase activity transferring phosphorus containing groups” (fosfor iyonunun transferini sağlayan enzim aktivitesi), “Transmembrane transporter activity” (hücre içi ve hücre dışı taşımayı sağlayan enzim aktivitesi) için ve son olarak da 5 contig de “İsomerase activity” (izomerlerin birbirine dönüşümünü sağlayan enzim aktivitesi).

Şekil 3.8.’deki ikinci grafik ise şekil 3.8.’deki dairesel grafiğin açılımını oluşturmaktadır. Dairesel grafikte her bir dilimde gösterilen Gen Ontoloji terimine dahil olan bir contigin bu terime ait olmasını sağlayan fonksiyonunu veren çizelge şeklinde contig sayıları ile birlikte sıralanmıştır (Şekil 3.8.).

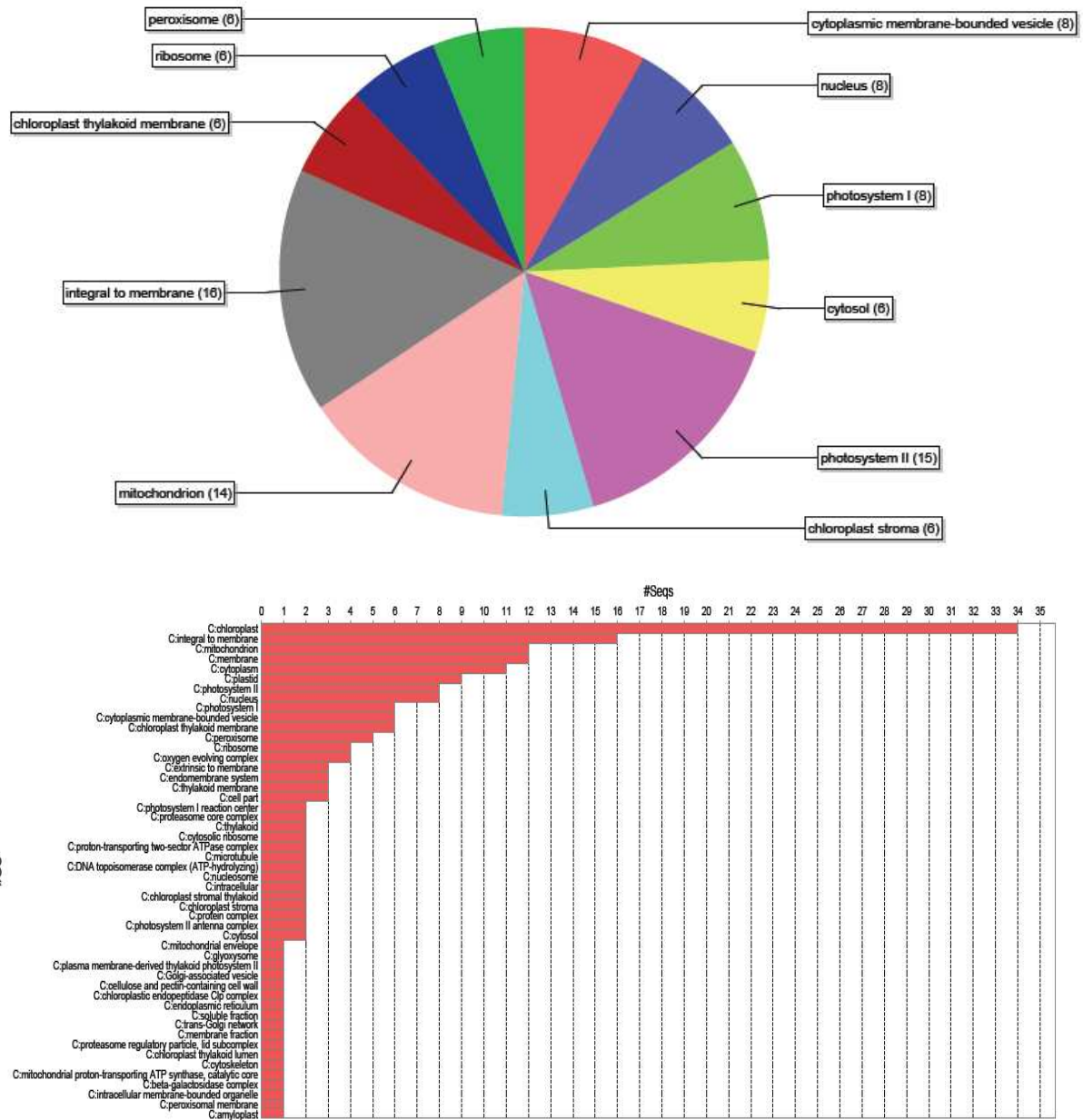




Şekil 3.8.: Yaprak dokusuna ait contiglerin moleküler fonksiyonlarına göre sınıflandırılması.

Şekil 3.9.'daki dairesel grafik Şekil 13'deki moleküler fonksiyon sınıflandırmasında olduğu gibi yaprak dokusuna ait 205 contigın hücresel bileşenlerine göre sınıflandırılmasını göstermektedir. Buna göre 205 contig içerisinde en fazla görülen hücresel bileşen, on beşer contigde belirlenmiş olan “integral to membrane” (hücre zarı bileşenleri) ve “photosystem II” (fotosentezde su molekülünün parçalandığı, kloroplastta gerçekleşen reaksiyonların meydana geldiği yer fotosistem II)'dir. İkinci sırada ise, 14 contigle “mitochondrion” (mitokondri) bulunur. Daha sonra sekizer contigle; “cytoplasmic membrane-bounded vesicle” (hücrearası taşıma ve depolama yapılabildiğini sağlayan hücre zarındaki küçük bir kesecik), “nucleus” (nükleus), “photosystem I” (fotosentezde güneş ışığından kimyasal enerjinin üretildiği, kloroplastta gerçekleşen reaksiyonların meydana geldiği yer fotosistem I), altışar contigle “chloroplast stroma” (kloroplast stroma), “cytosol” (sitoplazma), “chloroplast thylakoid membrane” (kloroplast tilakoid membran), “ribosome” (ribozom), “peroxisome” (peroksizom) şeklinde hücresel bileşenler sıralanmaktadır.

Şekil 3.9.'daki ikinci grafik ise yaprak dokusuna ait 205 contig için bulunan hücresel bileşen GO terimlerinin tamamını vermektedir. Bir bakıma dairesel grafiğin açılımını oluşturmaktadır. Dairesel grafikte daha geniş kapsamlı verilen fonksiyonlar bu kısımda ayrıntılı olarak yine sahip oldukları contig sayıları ile birlikte verilmiştir (Şekil 3.9.).

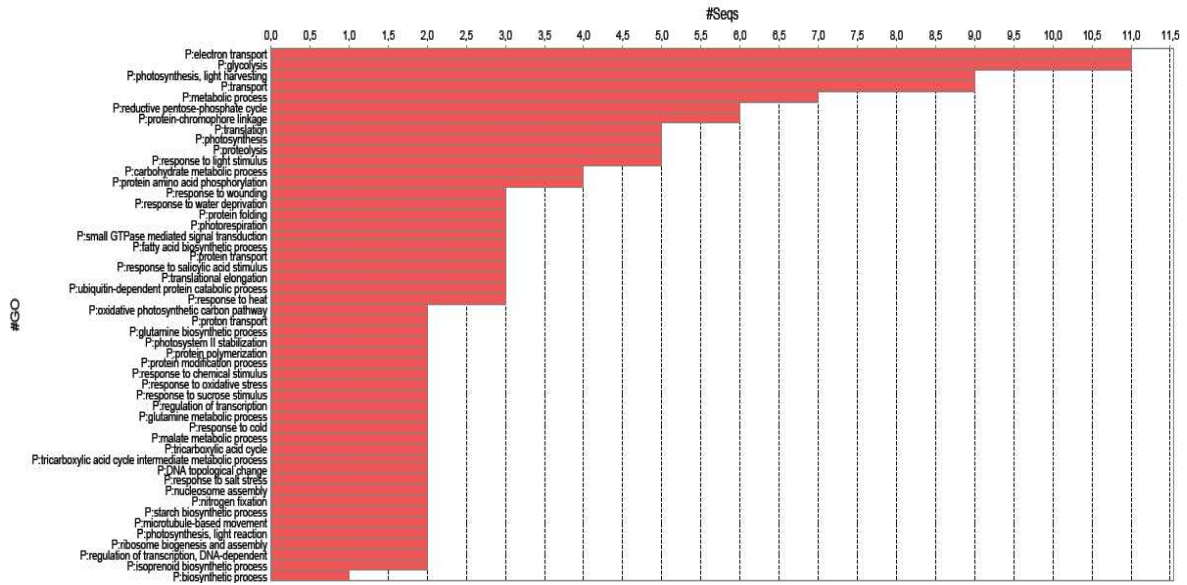
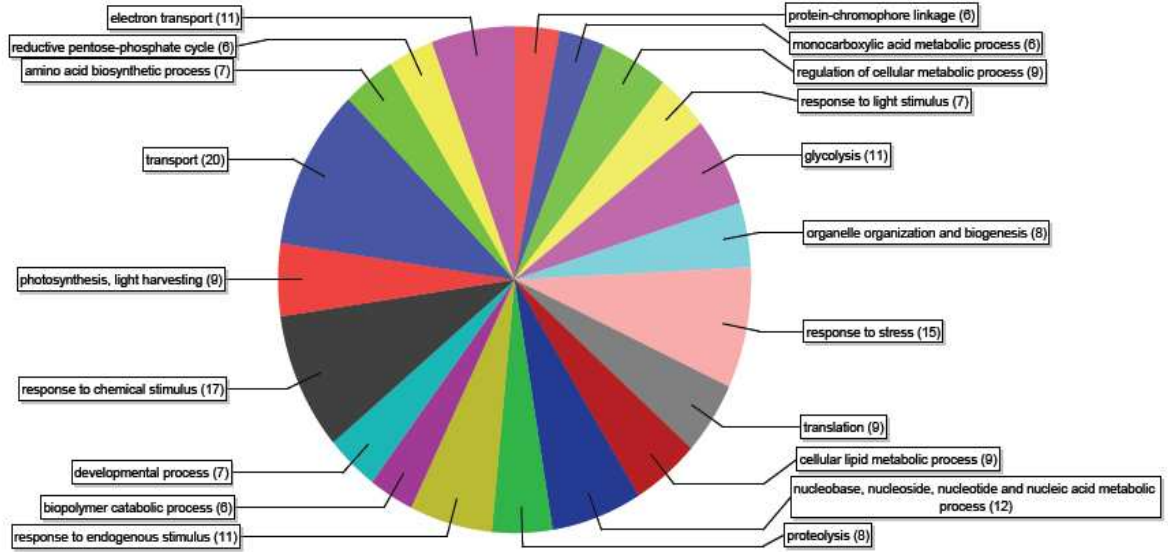


Şekil 3.9.: Yaprak dokusuna ait contiglerin hücre bileşenlerine göre sınıflandırılması.

Şekil 3.10.'daki dairesel grafikte yaprak dokusuna ait belirlenmiş biyolojik proseslerin sınıflandırılması gösterilmiştir. Buna göre 205 contig içerisinde en fazla görülen biyolojik proses, 20 contigde belirlenmiş olan “transport” (taşıma)’dur. İkinci sırada ise, 17 contigle “response to chemical stimulus” (organizmada kimyasal bir uyarana sonucu bir aktivitenin veya bir aşamanın değişimi) bulunur. Daha sonra 15 contigle “response to stress” (biyotik veya abiyotik stresten sorumlu proseslere ait diziler), 12 contigle; “nucleobase, nucleoside,

nucleotide and nucleic asit metabolic proses” (nükleobaz, nükleozit, nükleotid ve nükleik asit metabolizma prosesleri), on birer contigle “glycolysis” (glikolizis: solunum sırasında glikozun pürivik aside parçalanması), “response to endogenous stimulus” (organizmanın kendi içinde ortaya çıkan bir uyarın sonucu bir aktivitenin veya bir aşamanın deęiřimi), “electron transport” (elektron tařınması), dokuzar contigle “cellular lipid metabolic process” (organik çözücülerde çözünebilen suda çözünemeyen, lipit içeren, kimyasal reaksiyon ve sentez yolları), “translation” (translasyon), “regulation of cellular metabolic process” (her hücredeki kimyasal maddelerin dönüşümü ile kimyasal reaksiyonlar ve sentez yollarının hücrede oluşmasını, oluşma oranını ve oluşma periyodunu düzenleyen proseslerin hepsi), “photosynthesis, light harvesting” (fotosistem reaksiyon merkezleri arasında ışık fitonlarından absorbe edilen enerjinin absorpsiyonu ve transferi), sekizer contigle “organelle organization and biogenesis” (bir hücrede bir organelin oluşumunun, düzenlenmesinin ve yok edilmesinin hücresel düzeyde gerçekleşmesini sağlayan proseslerdir), “proteolysis” (peptit bağlarının yıkılması ile bir proteinin aktif formunun veya doğal yapısının bozulmasını sağlayan kimyasal prosesler), yedişer contigle “amino acid biosynthetic process” (amino asitlerin oluşmasını sağlayan kimyasal prosesler), “developmental process” (yaşam kısımlarının ilerleyerek tamamlandığı biyolojik prosesler), “response to light stimulus” (bir ışık uyarını sonucu bir hücredeki veya organizmadaki bir aktivitenin veya bir aşamanın deęiřimi), altışar contigle “protein-chromophore linkage” (kovalent veya kovalent olmayan bağlarla kromoforun proteine bağlanması), “monocarboxylic acid metabolic process” (monokarboksil gruplarını ve anyon gruplarını bulunduran kimyasal prosesler), “reductive pentose-phosphate cycle” (mezofil hücrelerinde malik aside dönüřtürülebilmesi için 4 karbonlu molekülün oluşturulması amacıyla CO₂ ile 3 karbonlu fosfoenol piruvat (PEP) molekülünün kombinasyonunu yapmak. pentoz-fosfat döngüsü (Calvin döngüsü) olarak da bilinir), “biopolymer catabolic process” (biyopolimerlerin yıkımından sorumlu kimyasal reaksiyonlar) şeklinde biyolojik prosesler sıralanmaktadır.

Şekil 3.10.’daki ikinci grafik ise yaprak dokusuna ait 205 contig için bulunan biyolojik proses GO terimlerinin ayrıntılarını vermektedir (Şekil 3.10.).



Şekil 3.10.: Yaprak dokusuna ait contiglerin biyolojik proseslerine göre sınıflandırılması

3.2.2.2 Zeytin (Meyve) Dokusuna Ait Contiglerin BLAST Çalışması Sonuçları

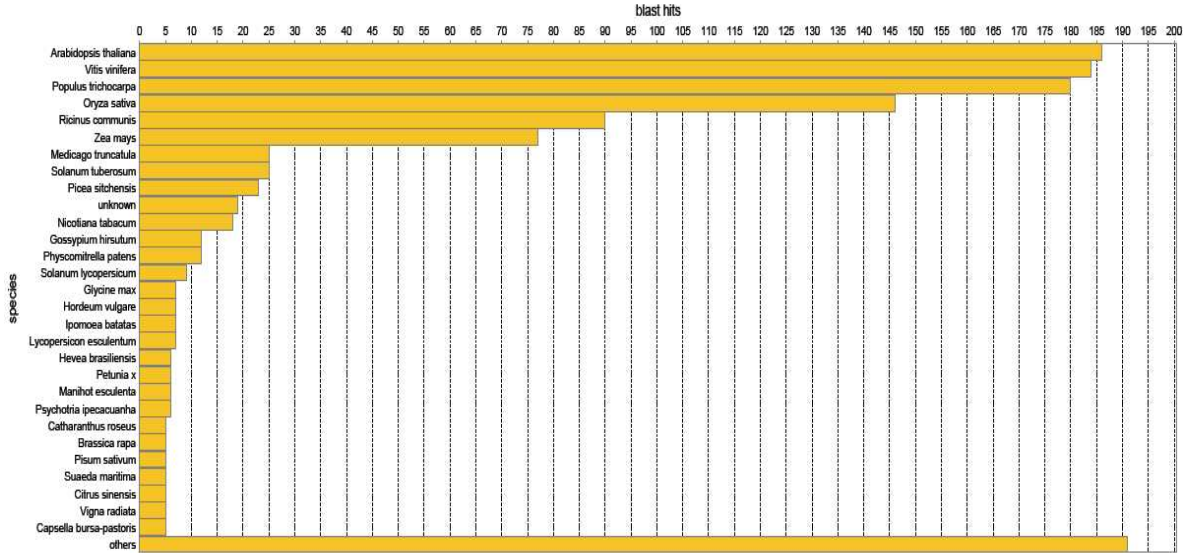
Meyve dokusuna ait 1506 EST dizisinden 619 tanesi CAP3 programı analizleri sonucu birleşme işlemine katılarak kendi aralarında 69 contig oluşturmuşlardır. Bu 69 meyve contig dizisinden 4 contig veritabanındaki çeşitli zeytin (*Olea europaea*) dizileri ile benzerlik göstermiştir (Çizelge 3.5.). Buna göre meyve dokusunda görülen ve zeytin bitkisinde daha önce belirlenmiş olan bu 4 protein; genel olarak monooksijenaz reaksiyonu katalizleyen bir enzim olarak “cytochrome P450”, iki glikoz molekülünü $\beta 1 \rightarrow 4$ bağı ile bağlayan glukozidaz

enzimi “beta-glucosidase”, sarı acı baklada görülen metalofosfataza benzer ADP ve GDP gibi difosfonükleotidlerin pirofosfat bağlarını kesen enzim “metallophosphatase/diphosphonucleotide phosphatase 1”, Rubisco adıyla bilinen Calvin döngüsünde karbon fiksasyonunu kataliz eden, yaprak dokusunda en fazla bulunan protein “ribulose-1,5-bisphosphate carboxylase/oxygenase activase”, şeklinde sıralanmaktadır (Çizelge 3.5.).

Çizelge 3.5.: BLAST sonuçlarına göre GenBanktaki zeytin dizileri ile benzerlik gösteren zeytin (meyve) dokusuna ait contigler

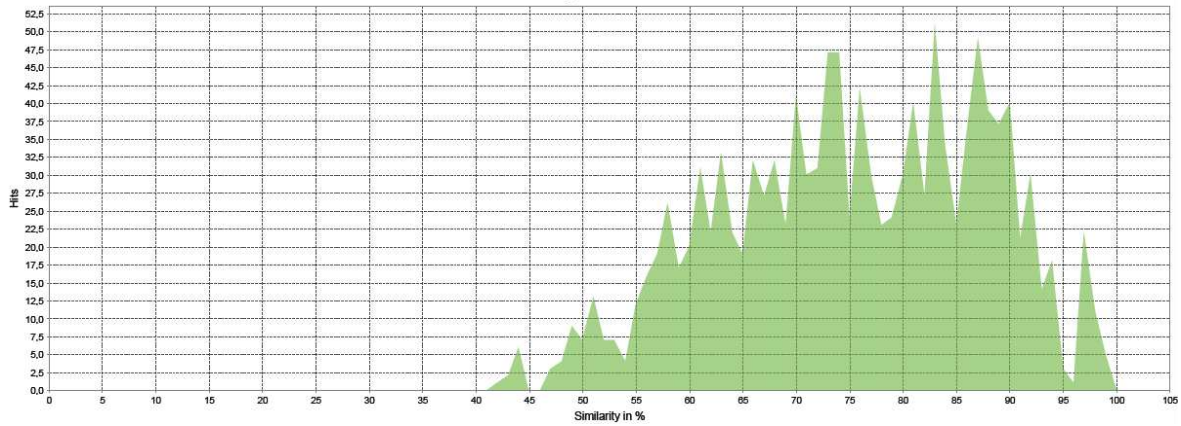
Contig no	Veritabanında eşleştiği zeytin dizisi	Query coverage	Max. ident	Contig uzunluğu	İçerdiği EST sayısı
Contig 4	Olea europaea putative cytochrome P450 mRNA, partial cds	%28	%98	1752	5
Contig 19	Olea europaea subsp. europaea beta-glucosidase (bglc) mRNA, complete cds	%99	%94	1279	6
Contig 38	Olea europaea putative metallophosphatase/diphosphonucleotide phosphatase 1 mRNA, partial cds	%28	%96	969	2
Contig 50	Olea europaea putative ribulose-1,5-bisphosphate carboxylase/oxygenase activase mRNA, partial cds	%68	%95	862	2

Şekil 3.11.’de BLAST2GO programı ile elde edilen “Species distribution” sonucu görülmektedir. Meyve dokusuna ait 1506 EST dizisinden elde edilen 69 contigın BLASTX analizlerine göre benzerlik gösterdiği diğer organizmalar sıralanmaktadır. Bu sonuç, $69 \times 20 = 1380$ blast hit için oluşturulmuştur. Buna göre; meyve dokusu contiglerinde *Olea Europaeae cv. Gemlik* organizması ile en fazla benzerlik gösteren organizma “*Arabidopsis thaliana*” olmuştur. İkinci sırada, “*Vitis vinifera*” (Üzüm) ve üçüncü sırada “*Populus trichocarpa*” (Batı Balsam Kavağı) yer alır. Yaprak dokusunda olduğu gibi benzer sonuçların meyve dokusunda da elde edilmesi sebebinin, *Populus trichocarpa* ve *Vitis vinifera* bitkilerinin *Olea europaeae* bitkisi gibi dikotileden bitki olması ve aynı zamanda monokotil olan *Arabidopsis thaliana* bitkisi gibi bu bitkilerin GenBankta kayıtlı dizi sayılarının çok fazla olması ile ilgili olduğu düşünülmektedir (Şekil 3.11.).



Şekil 3.11.: Meyve için Species distribution: CAP3 programından elde edilen meyve dokusuna ait contig dizilerinin BLASTX sonuçlarına göre GenBanktaki diğer organizmalarla olan benzerliği

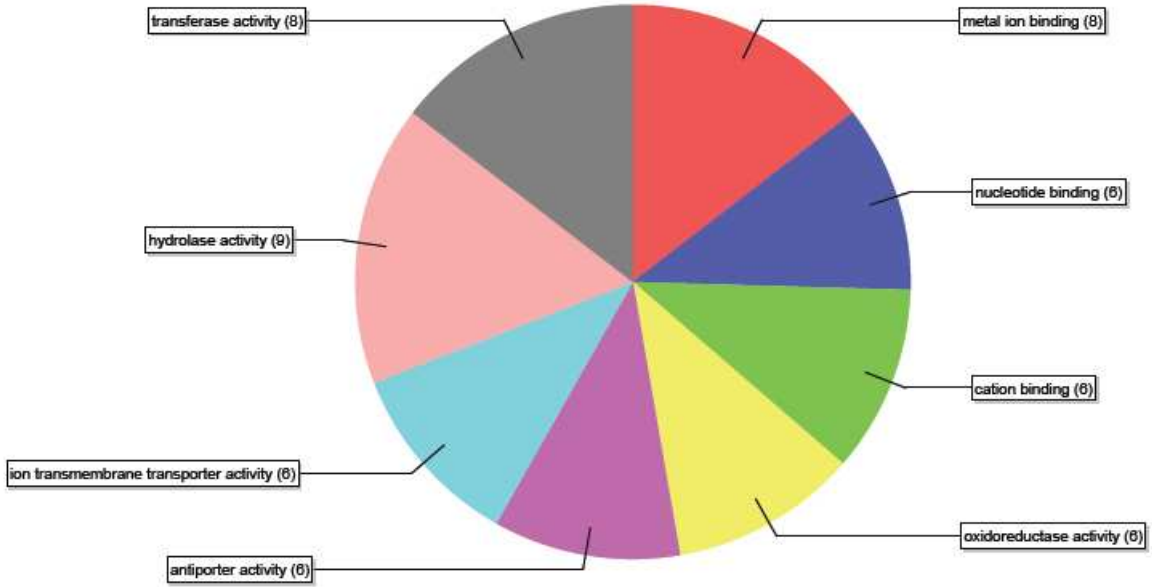
Şekil 3.12.'de BLAST2GO programının “Similarity distribution” olarak adlandırdığı sonuç, meyve dokusuna ait 1506 EST dizisinden elde edilen 69 contig BLASTX analizlerine göre benzerlik gösterdiği GenBank veritabanına kayıtlı blast hitlerle örtüşen nükleotid dizilerinin nükleotid sayısı bakımından benzerlik ortalamasını göstermektedir. “Hits” olarak adlandırılan değer 69 contigle örtüşen veri tabanına kayıtlı blast hitlerini, “similarity” olarak adlandırılan değer hitlerle contiglerin örtüştüğü nükleotid sayısının tüm contig uzunluğuna oranını “%” cinsinden vermektedir (Şekil 3.12.).

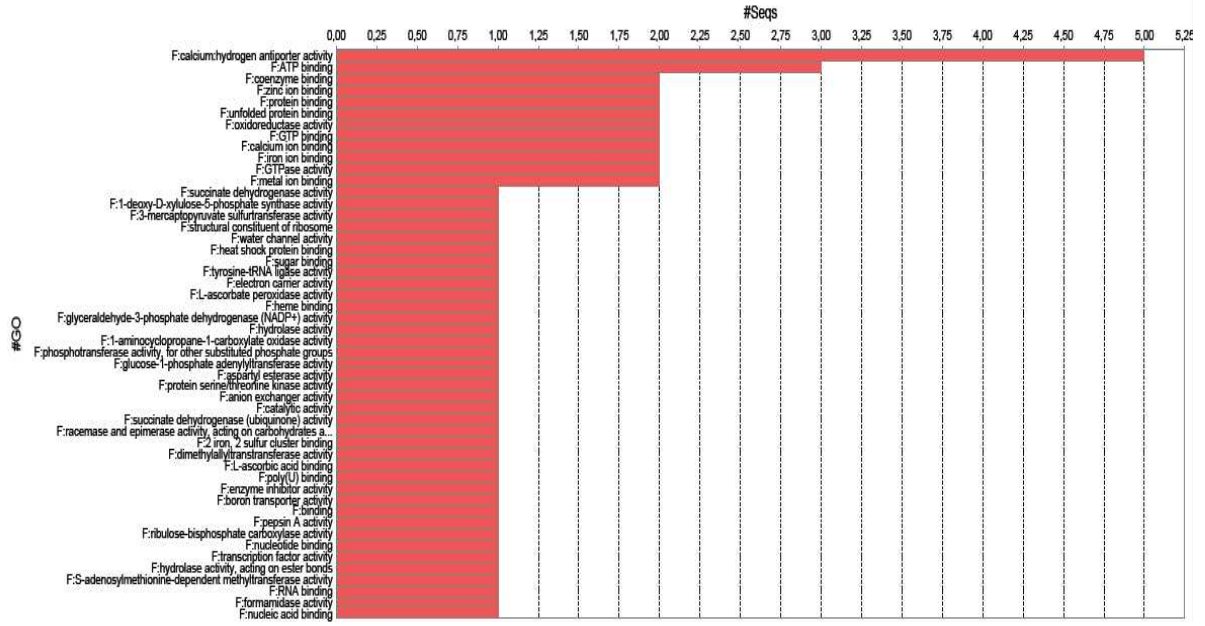


Şekil 3.12.: CAP3 programından elde edilen meyve dokusuna ait contig dizilerinin BLAST sonuçlarına göre GenBanktaki diğer nükleotid dizileri ile olan benzerlik oranı

Şekil 3.13'deki dairesel grafikte meyve dokusuna ait 1506 EST dizisinden CAP3 programı ile birleştirilen 69 contigin BLAST2GO programı default parametreleriyle BLASTX analizlerinden elde edilen gen ontoloji değerlerinin moleküler fonksiyonlarına göre sınıflandırılması gösterilmiştir. Buna göre 69 contig içerisinde en fazla görülen fonksiyon 9 contigde belirlenmiş olan “hydrolase activity” (hidroliz) aktivitesidir. Hidroliz aktivitesini sekizer contigle “Transferase” (Transferaz) aktivitesi ve “metal ion binding” (metal iyonlarını bağlayıcı aktivite) izler. Daha sonra altışar contigle; “İon transmembrane transporter activity” (hücre içi ve hücre dışı iyon taşımaya sağlayan enzim aktivitesi), “antiporter activity” (kemosmatik enerji dışında diğer enerji formları ile doğrudan bağlantılı olmayan, çift yönlü, hücre zarından karşılıklı aktif taşımaya yapıldığı mekanizmadır), “Oxidoreductase activity” (elektron transferini sağlayan enzim aktivitesi), “cation binding” (katyon bağlayıcı), “nucleotide binding” (nükleotid bağlayıcı) şeklinde sıralanmaktadır (Şekil 3.13.).

Şekil 3.14.'daki ikinci grafik ise meyve dokusuna ait 69 contig için bulunan moleküler fonksiyon GO terimlerinin tamamını vermektedir. Dairesel grafiğin daha ayrıntılı bir anlatımı olan diğer çizelgede fonksiyonlar ayrıntılı olarak sahip oldukları contig sayıları ile birlikte verilmiştir (Şekil 3.14.).

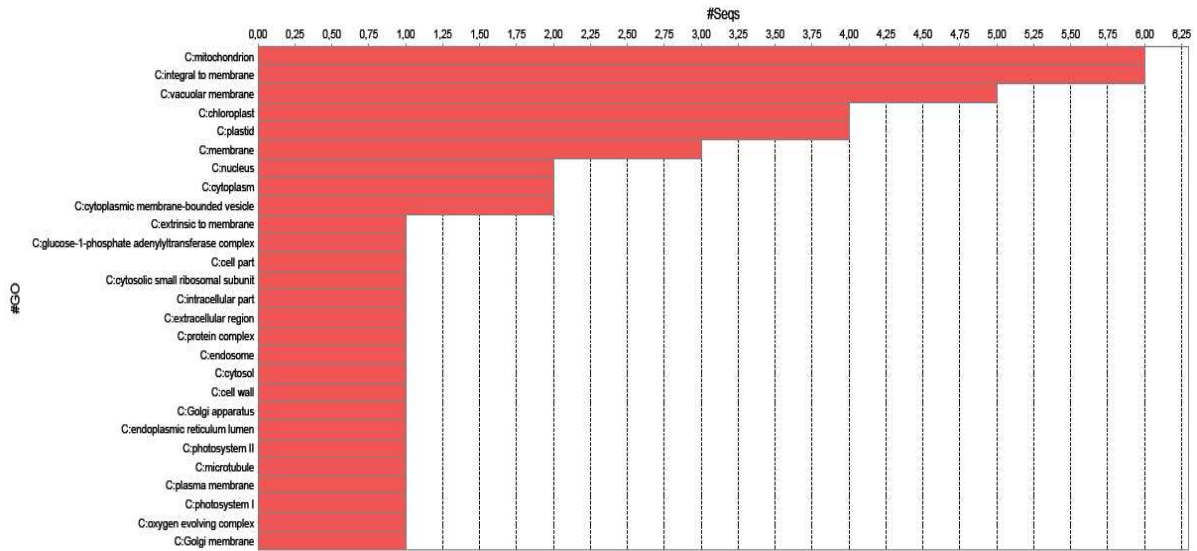
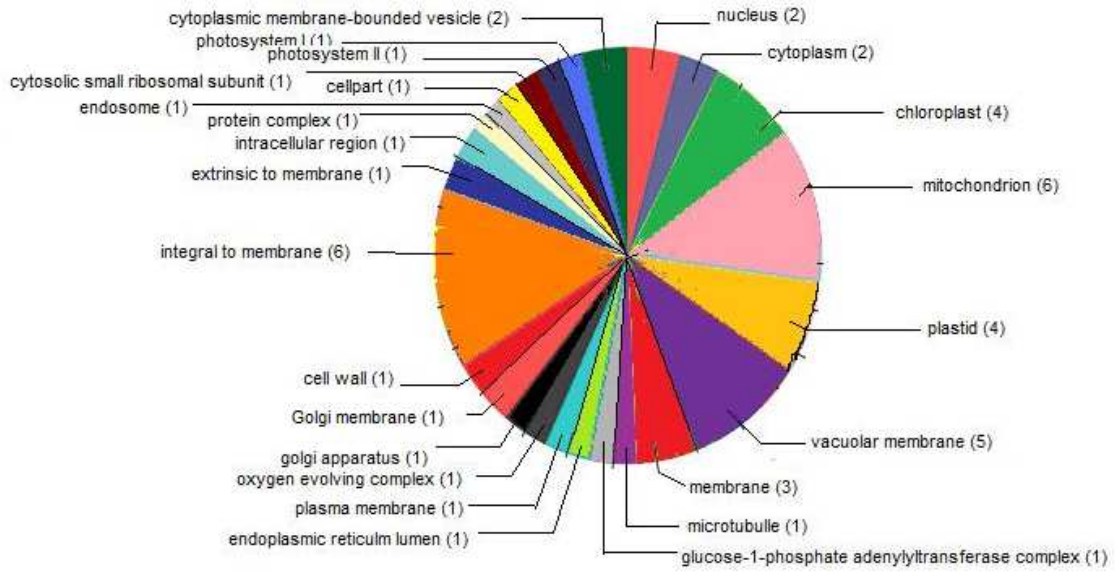




Şekil 3.14.: Meyve dokusuna ait contiglerin moleküler fonksiyonlarına göre sınıflandırılması

Şekil 3.15.’deki dairesel grafikte meyve dokusuna ait 1506 EST dizisinden CAP3 programı ile birleştirilen 69 contigin BLAST2GO programı default parametreleriyle BLASTX analizlerinden elde edilen gen ontoloji değerlerinin hücresel bileşenlerine göre sınıflandırılması gösterilmiştir. Contig sayısının yaprak ve toplam EST’lerden oluşan contig sayısından çok az olması ve oluşan hücresel bileşen sayısının da az olması sebebiyle bu kez dairesel grafik meyve contigleri için tüm hücresel bileşen GO terimlerini kapsamaktadır. 69 contigden 51 tanesinin hücresel bileşen fonksiyonu bulunmaktadır. En fazla görülen hücresel bileşenler diğer contig gruplarında da görüldüğü gibi “mitochondrion” (mitokondri) ve “integral to membrane” (hücre zarı bileşenleri) olmuştur. Bunun dışında “golgi apparatus oxygen evolving complex”, “microtubulle” ve “cytosolic small ribosomal subunit” gibi diğer hücresel bileşen grafiklerinde görülmeyen GO terimlerine de rastlanmaktadır.

Şekil 3.15.’deki ilk grafik ikinci grafikte birebir aynıdır. GO terimlerinin sahip oldukları contig sayıları ilk grafikte terimlerin yanındaki parantez içinde belirtilmiştir. İkinci grafikte ise çizelgenin “seqs” parametresinden bulunabilmektedir (Şekil 3.15.).

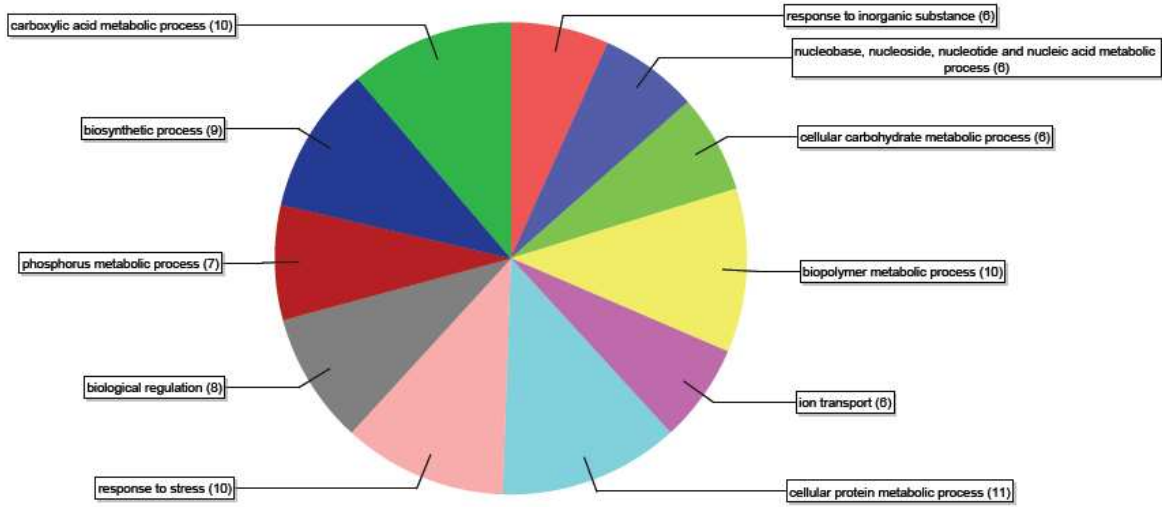


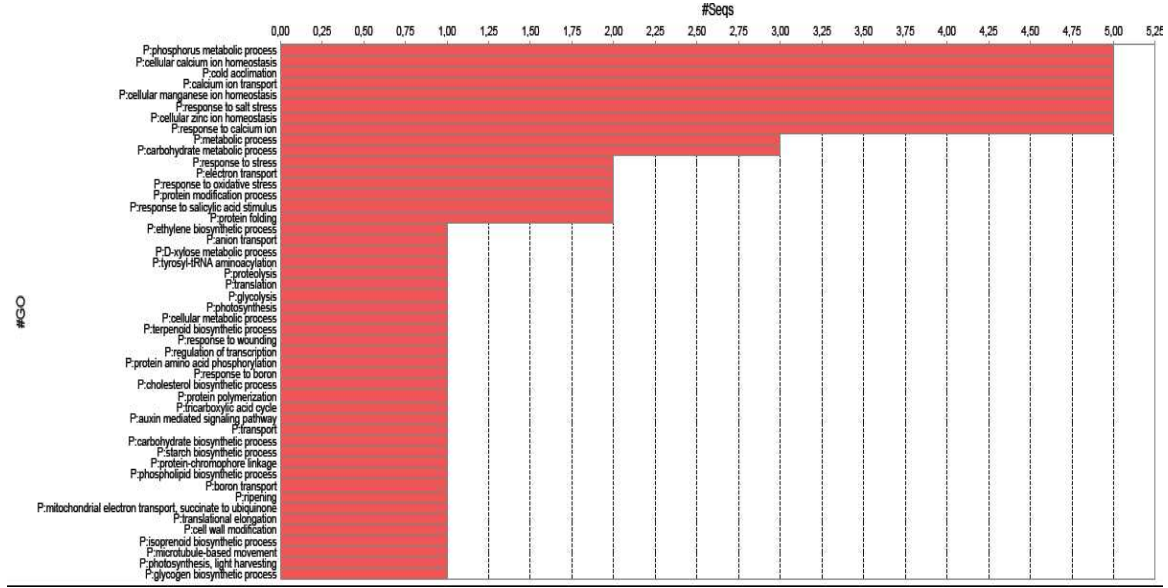
Şekil 3.15.: Meyve dokusuna ait contiglerin hücre bileşenlerine göre sınıflandırılması

Şekil 3.16.'daki dairesel grafikte meyve dokusuna ait 69 contigin BLAST2GO programı analizlerinden elde edilen gen ontoloji değerlerinin biyolojik proseslerine göre sınıflandırılması gösterilmiştir. 69 contig içerisinde en fazla görülen biyolojik proses, 11 contigde belirlenmiş olan “cellular protein metabolic process” (hücre içersindeki bir kimyasal reaksiyon için spesifik proteinin kullanıldığı prosesler)’dir. İkinci sırada ise, onar contigle “carboxylic acid metabolic process” (karboksil gruplarını ve anyon gruplarını bulunduran kimyasal prosesler), “response to stress” (biyotik veya abiyotik stresden sorumlu proseslere ait diziler), “biopolymer metabolic process” (polisakkaritler ve polipeptitler gibi

biyopolimerleri içeren kimyasal reaksiyonlar), 9 contig “biosynthetic process” (yeni bir biyo molekül sentezinde veya metabolizmanın ihtiyacı olan enerji üretiminde gözlenen reaksiyonlar), 8 contig “biological regulation” (biyolojik proseslerin oluşumunu, oluşma sıklığını, oluşma oranını, kalitesini ve fonksiyonunu düzenleyen reaksiyonlar), 7 contig “phosphorus metabolic process” (genellikle fosfat grupları halinde metal olmayan element fosfor veya fosfor içeren bileşikler taşıyan kimyasal reaksiyonlar), altışar contig “nucleobase, nucleoside, nucleotide and nucleic acid metabolic proses” (nükleobaz, nükleozit, nükleotid ve nükleik asit metabolizma prosesleri), “ion transport” (iyon taşınması), “cellular carbohydrate metabolic process” (karbonhidratları içeren kimyasal reaksiyonlar), “response to inorganic substance” (bir organizmada yada bir hücrede inorganik bir substratın uyarısı sonucu bir aktivitenin veya bir aşamanın değişimi) şeklinde bulunur.

Şekil 3.16.’daki ikinci grafik ise meyve dokusuna ait 69 contig için bulunan biyolojik proses GO terimlerinin tamamını vermektedir. Bir bakıma dairesel grafiğin açılımını oluşturmaktadır. Dairesel grafikte daha geniş kapsamlı verilen fonksiyonlar bu kısımda ayrıntılı olarak yine sahip oldukları contig sayıları ile birlikte verilmiştir (Şekil 3.16.).





Şekil 3.16.: Meyve dokusuna ait contiglerin biyolojik prosese göre sınıflandırılması

3.2.2.3 Kütüphanedeki Bütün EST' ler (Toplam) İçin Yapılan Biyoinformatik Analizlerden Elde Edilen Contiglerin BLAST Çalışması Sonuçları

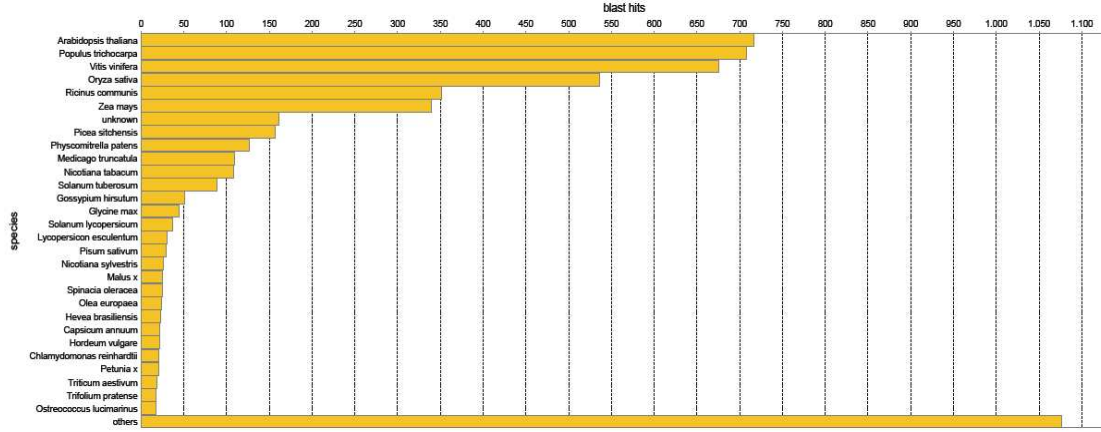
3734 EST dizisinin tamamı fragment assembly işlemi için CAP3 programı ile analiz edildi. Bu çalışmada 1366 EST fragmenti birleştirme işlemine dahil olarak 299 adet contig oluşturdu. Aynı fragment birleştirme programı (CAP3), aynı EST'ler için kullanılmasına karşın meyve dokusundan 69 contig, yaprak dokusundan 205 contig oluşmuşken iki dokuya ait EST'lerin toplamından $205+69=274$ contig yerine 299 contig oluşmuştur. 299 contig için yapılan BLAST analizi sonuçlarına göre bu contig dizilerinden 13 tanesi veritabanındaki çeşitli zeytin (*Olea europaea*) dizileri ile benzerlik göstermiştir. Bu benzerlik gösteren contiglerin hangileri olduğu ve hangi fonksiyonlarla eşleştiği çizelgeden incelenebilir. Buna göre bu 13 proteinden yalnızca bir polen kaynağında allerjen olarak tanımlanan ilk Cu/Zn super-oksit dismutaz olan "Cu/Zn super-oxide dismutase (ole e 5 allergen)" daha önce meyve ve/ veya yaprak dokularından elde edilen contiglerin BLAST analizlerinde rastlanmamıştır. Diğer 12 protein ise meyve ve yaprak dokularında da görülmüş olan BLAST sonuçları ile aynıdır (Çizelge 3.6.).

Çizelge 3.6.: BLAST sonuçlarına göre GenBanktaki zeytin dizileri ile benzerlik gösteren kütüphaneye ait contigler

Contig no	Veritabanında eşleştiği zeytin dizisi	Query coverage	Max. İdent.	Contig uzunluğu	İçerdiği EST sayısı
Contig 7	<i>Olea europaea</i> putative mannitol dehydrogenase 1 (MTD1) mRNA	%56	%86	895	2
Contig 14	<i>Olea europaea</i> photosystem II 10 kDa polypeptide mRNA, partial cds	%30	%99	724	4
Contig 24	<i>Olea europaea</i> putative glycolate oxidase-like FMN-binding domain protein mRNA	%22	%99	2819	9
Contig 85	<i>Olea europaea</i> putative plant lipid transfer protein mRNA	%25	%100	914	5
Contig 93	<i>Olea europaea</i> Cu/Zn super-oxide dismutase (ole e 5 allergen)	%79	%93	891	2
Contig 98	<i>Olea europaea</i> putative cytochrome P450 mRNA, partial cds	%28	%99	1756	8
Contig 111	<i>Olea europaea</i> putative ribulose-1,5-bisphosphate carboxylase/oxygenase activase mRNA, partial cds	%43	%98	1797	22
Contig 137	<i>Olea europaea</i> subsp. <i>europaea</i> beta-glucosidase (bglc) mRNA, complete cds	%93	%98	2018	13
Contig 155	<i>Olea europaea</i> tonoplast intrinsic protein (tip) mRNA, complete cds	%87	%86	911	2
Contig 157	<i>Olea europaea</i> polyubiquitin OUB2 mRNA, complete cds	%89	%91	1393	7
Contig 169	<i>Olea europaea</i> 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	<i>Olea europaea</i> putative glyoxisomal malate dehydrogenase mRNA, partial cds	%46	%97	1383	2
Contig 255	<i>Olea europaea</i> putative metallophosphatase/diphosphonucleotide phosphatase 1 mRNA, partial cds	%28	%96	969	2

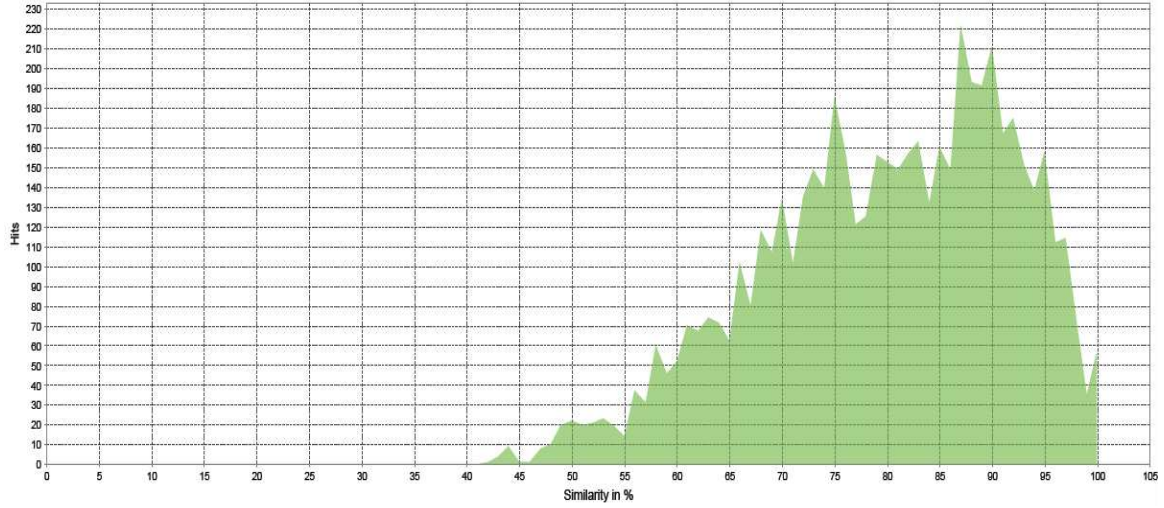
Şekil 3.17.'de BLAST2GO programı ile elde edilen “Species distribution” sonucu görülmektedir. Toplam 3734 EST dizisinden elde edilen 299 contigin BLASTX analizlerine göre benzerlik gösterdiği diğer organizmalar sıralanmaktadır. Bu sonuç, $299 \times 20 = 5980$ blast hit için oluşturulmuştur. Buna göre; toplam contiglerde *Olea Europaeae* cv. *Gemlik* organizması ile en fazla benzerlik gösteren organizma “*Arabidopsis thaliana*” olmuştur. İkinci

sırada, “*Populus trichocarpa*” (Batı Balsam Kavağı) ve üçüncü sırada “*Vitis vinifera*” (Üzüm) yer alır. Yaprak ve meyve dokusunda olduğu gibi benzer sonuçların bu toplam EST’lerden elde edilen contig dizilerinde de gözlemlenmesinin sebebi bu üç bitkiye ait GenBankta kayıtlı verinin çok fazla olması şeklinde düşünülmektedir (Şekil 3.17.).



Şekil 3.17.: Species distribution: CAP3 programından kütüphaneye ait contig dizilerinin BLASTX sonuçlarına göre GenBanktaki diğer organizmalarla olan benzerliği

Şekil 3.18.’de BLAST2GO programının “Similarity distribution” olarak adlandırdığı sonuç, yaprak ve meyve dokusuna ait toplam 3734 EST dizisinden elde edilen 299 contigin BLASTX analizlerine göre benzerlik gösterdiği GenBank veritabanına kayıtlı blast hitlerle örtüşen nükleotid dizilerinin nükleotid sayısı bakımından benzerlik ortalamasını göstermektedir. “Hits” olarak adlandırılan değer 299 contigle örtüşen veri tabanına kayıtlı blast hitlerini, “similarity” olarak adlandırılan değer hitlerle contiglerin örtüştüğü nükleotid sayısının tüm contig uzunluğuna oranını “%” cinsinden vermektedir (Şekil 3.18.).



Şekil 3.18.: CAP3 programından elde edilen kütüphaneye ait contig dizilerinin BLAST sonuçlarına göre GenBanktaki diğer nükleotid dizileri ile olan benzerlik oranı

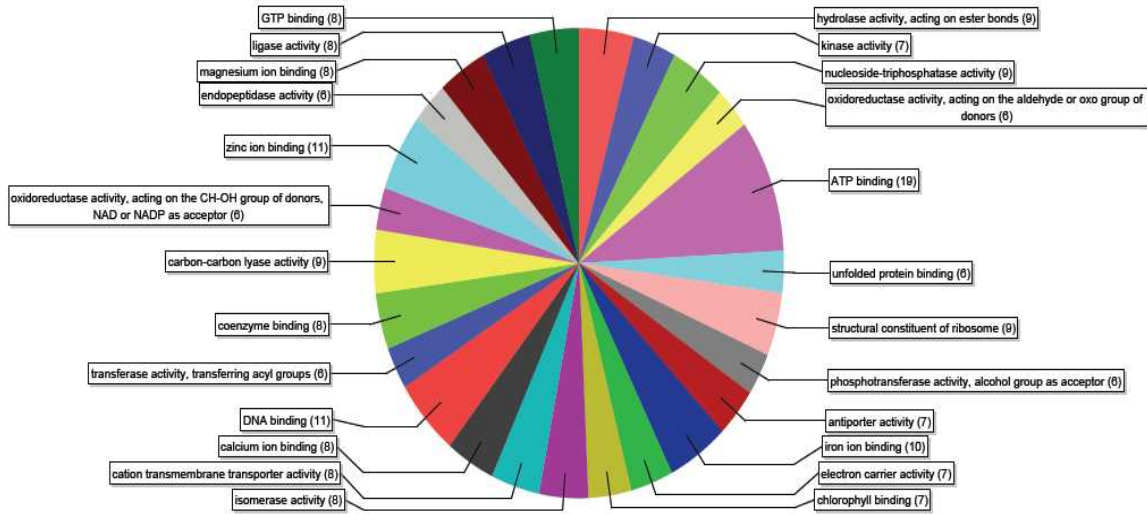
Şekil 3.19.'daki dairesel grafikte toplam 3734 EST dizisi için CAP3 programı ile oluşturulan 299 contigin BLAST2GO programı analizlerinden elde edilen gen ontoloji değerlerinin moleküler fonksiyonlarına göre sınıflandırılması gösterilmiştir. BLASTX sonuçlarına göre 299 contig içerisinde daha önce meyve ve/veya yaprak dokusunda da gözlemlenen moleküler fonksiyon terimleri sahip oldukları contig sayıları ile şu şekilde sıralanmaktadır;

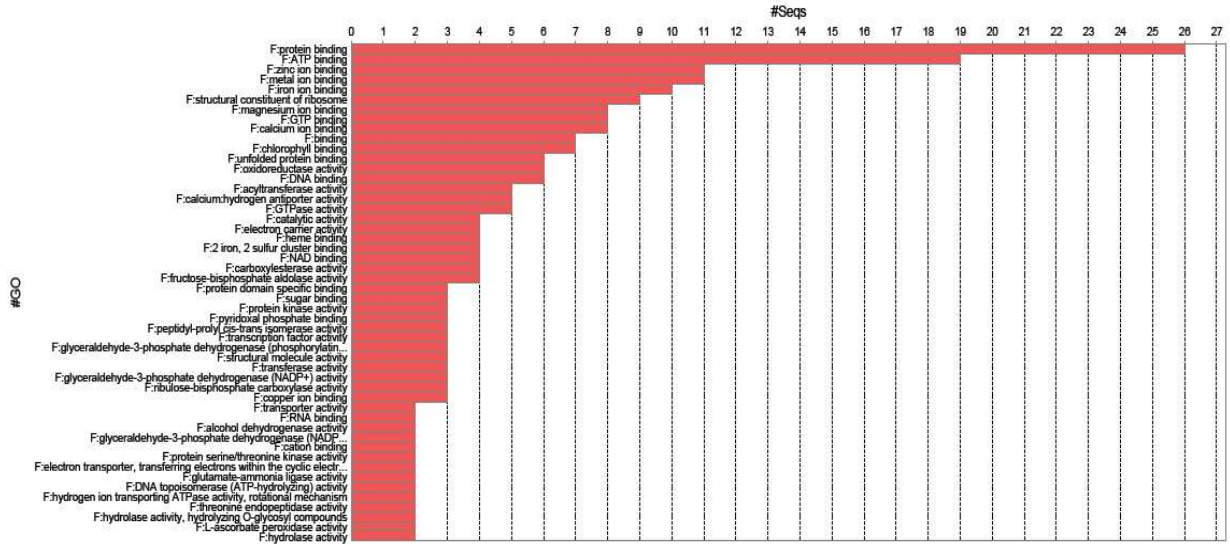
19 contig “ATP binding”,11 contig “DNA binding”,11 contig “zinc ion binding”,10 contig “İron ion binding”, 9 contig “Hydrolase activity, acting on ester bonds” , 9 contig “Nucleoside-triphosphatase activity”, 9 contig “Carbon carbon lyase activity”, 8 contig, “GTP binding”, 8 contig “Magnesium ion binding”, 8 contig “Coenzyme binding” , 8 contig “İsomerase activity”, 7 contig, “Electron carrier activity”, 7 contig “Chlorophyl binding”, 7 contig “antiporter activity”, 6 contig “Transferase activity transferring acyl groups”.

Bunların dışında 3734 EST dizisinin birleşiminden oluşmuş 299 contig içinde yer alan ve daha önce meyve veya yaprak dokusunda rastlanmamış toplam 68 contigde 10 farklı moleküler fonksiyon gözlenmiştir. Bu 10 farklı moleküler fonksiyonda contig sayıları şu şekilde dağılmaktadır; 9 contig “structural constituent ribosome” (bir molekülün bir ribozomun yapısal bütünlüğüne katılmasını sağlayan prosesler), 5 contig “Oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor” (CH-OH gruplarının NAD⁺ veya NADP'deki bir hidrojen veya elektronu azalttıkları redoks tepkimelerini katalizleyen enzimler), 8 contig “carbon transmembrane transporter activity” (hücre içi ve hücre dışı carbon moleküllerini taşımaya sağlayan enzim aktivitesi), 7 contig “kinase activity” (genellikle ATP'den bir substrat molekülüne fosfat gruplarının taşınmasını

katalizleyen enzimler), 6 contig “endopeptidase activity” (bir polipeptit zincirindeki içteki alfa-peptit bağlarının hidrolizini katalizleyen enzimler), 6 contig “Oxidoreductase activity, acting on the aldehyde or oxo group of donors” (oksijen ve aldehit gruplarının elektron transferini sağlayan enzim aktivitesi), 6 contig “phosphotransferase activity, alchole groups as acceptor” (bir bileşikten bir alkol grubuna fosfor taşıyan bir grubun transferini katalizleyen enzimler), 5 contig “unfolded protein binding” (belirlenmiş bir protein ile seçici etkileşim sağlayan enzimler), 8 contig “ligase activity” (genellikle nükleosid trifosfat oluşumunda birbirine bağlı difosfat bağları ile substratın ligasyonunu katalizleyen enzimler), 8 contig “calcium ion binding” (kalsiyum iyonu bağlayıcı enzimler).

Şekil 3.19.’daki ikinci grafik ise şekil 23deki dairesel grafiğin açılımını oluşturmaktadır. Dairesel grafikte her bir dilimde gösterilen Gen Ontoloji terimine dahil olan bir contigın bu terime ait olmasını sağlayan fonksiyonunu veren çizelge şeklinde contig sayıları ile birlikte sıralanmıştır (Şekil 3.19.).



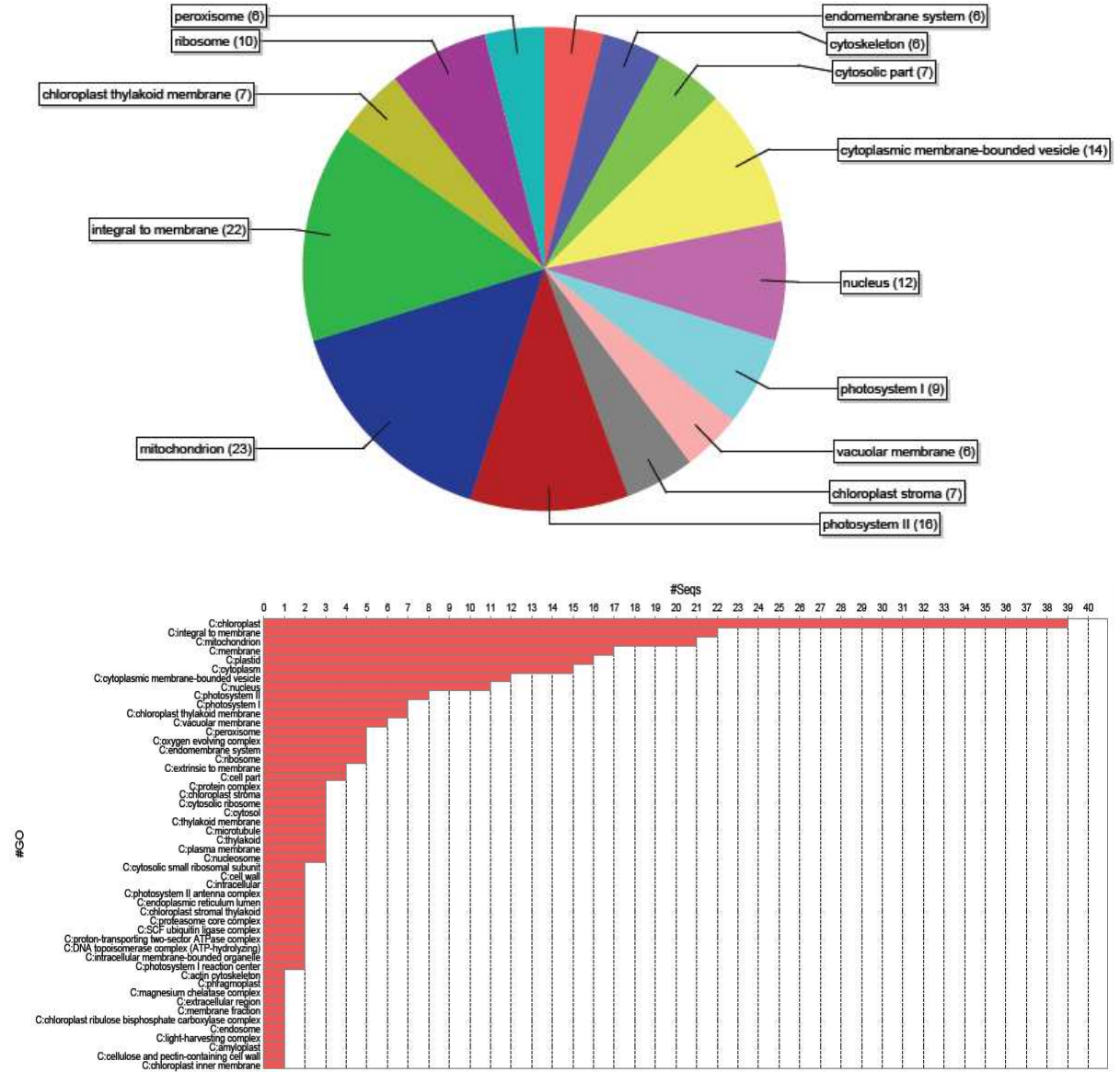


Şekil 3.19.: Kütüphaneye ait contiglerin moleküler fonksiyonlarına göre sınıflandırılması.

Şekil 3.20.'deki dairesel grafikte 299 contigin BLASTX analizlerinden elde edilen gen ontoloji değerlerinin hücresel bileşenlerine göre sınıflandırılması gösterilmiştir. Hücresel bileşen sınıflandırmasında daha önce meyve ve/veya yaprakta görülen hücresel bileşenler; 23 contig “mitochondrion” (mitokondri) için, 22 contig “integral to membrane” için, 16 contig “photosystem II” için, 14 contig “cytoplasmic membrane-bounded vesicle” için, 12 contig “nucleus” (nükleus) için, 10 contig “ribosome” (ribozom) için, 9 contig “photosystem I” için, 7 contig “chloroplast stroma” için, 7 contig “chloroplast thylakoid membrane” için, 6 contig “peroxisome” (peroksizom) için şeklinde sıralanmaktadır.

Daha önce meyve veya yaprak dokusunda gözlemlenmeyen 4 ayrı hücresel bileşen GO terimi toplam 25 contig içermektedir; 7 contig “cytosolic part” (herhangi bir sitosol oluşturuca parçayı, bir zar veya subselular partikülü içermeyen sitoplazma parçası), 6 contig “endomembrane system” (hücre içi taşımayı sağlayan zar yapılarının tamamı. Örneğin; endoplazmik retikulum, golgi aygıtı, vesiküller, hücre zarı ve nükleus zarı gibi.), 6 contig “cytoskeleton” (hücresinin iç formunu oluşturan sitoplazmadaki filament şeklindeki elementlerin tamamıdır. Herhangi bir deterjan uygulamasından sonra hücre zarı ve çözünebilir içerikler yok olduktan sonra geriye kalan yapılardır.), 6 contig “vacuolar membrane” (hücre sitoplazmasından içerik bakımından ayrılan, vakuolu çevreleyen çift katlı bir yağ tabakası).

Şekil 3.20.'daki ikinci grafik ise 3734 EST'nin tamamından oluşturulan 299 contig için bulunan hücresel bileşen GO terimlerinin açılımını oluşturmaktadır. Dairesel grafikte verilen hücresel bileşenlerin yaklaşık olarak her bir contig için fonksiyonel ayrıntısı bu kısımda verilmiştir (3.20.).



Şekil 3.20.: Kütüphaneye ait contiglerin hücre bileşenlerine göre sınıflandırılması.

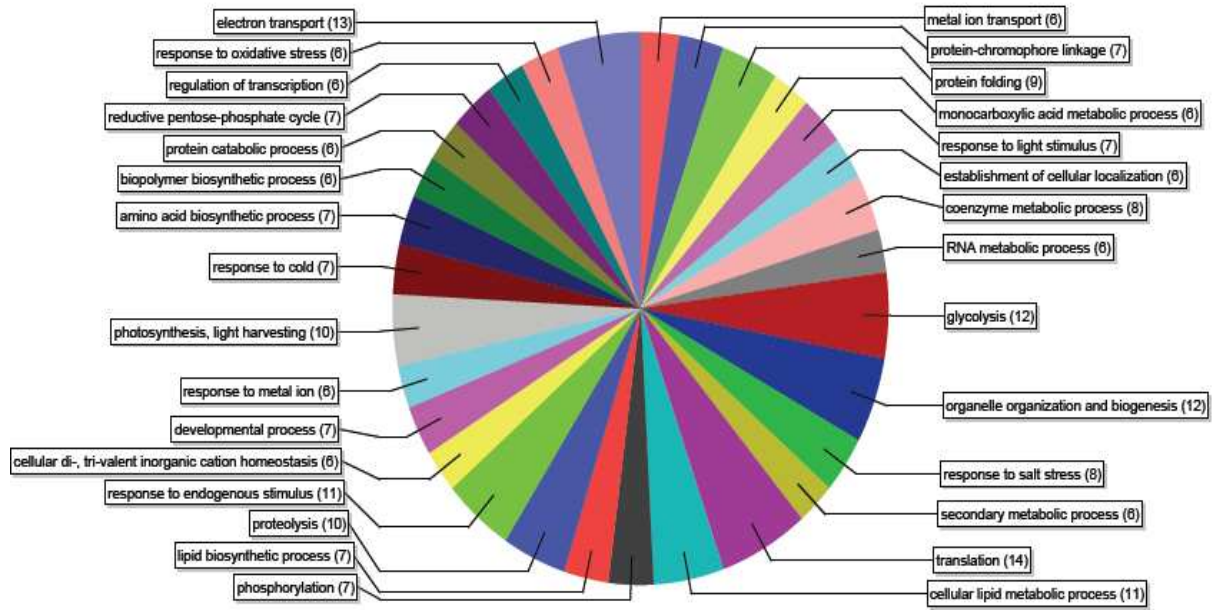
Şekil 3.21.'deki dairesel grafikte toplam 3734 EST dizisi için CAP3 programı ile oluşturulan 299 contigın BLAST2GO programı default parametreleriyle BLASTX analizlerinden elde

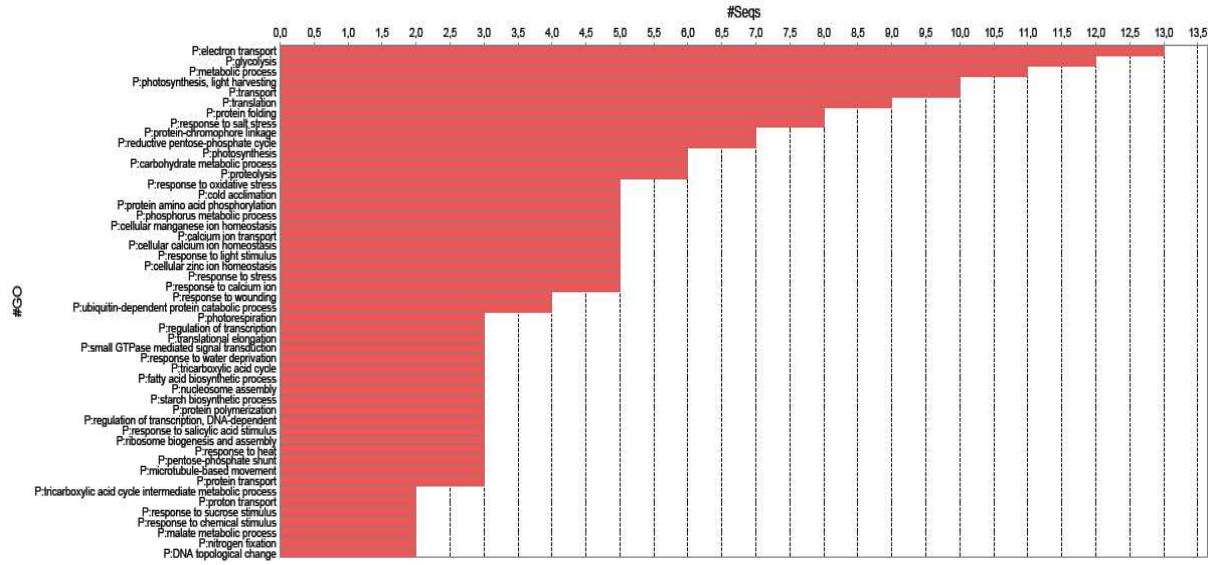
edilen gen ontoloji deęerlerinin biyolojik proseslerine gre sınıflandırılması gsterilmiřtir. Molekler fonksiyon ve hresel bileřen GO terimlerinde olduęu gibi biyolojik proses terimlerinde de meyve ve yaprak dokusuyla ortak olan ve olmayan GO terimleri bulunmaktadır. Buna gre 299 contig iin, meyve ve/veya yaprak dokusu contigleri ile ortak olan biyolojik proses GO terimleri řunlardır; 14 contig “translation”, 13 contig “electron transport”, 12 contig “glycolysis”, 12 contig “organelle organization and biogenesis”, 11 contig “response to endogenous stimulus”, 11 contig “cellular lipid metabolic process”, 10 contig “photosynthesis, light harvesting”, 10 contig “proteolysis”, 7 contig “response to light stimulus”, 7 contig “developmental process”, 7 contig “protein-chromophore linkage”, 7 contig “amino acid biosynthetic process”, 7 contig “reductive pentose-phosphate cycle”, 6 contig “monocarboxylic acid metabolic process”.

Meyve ve yaprak dokusundan farklı olan biyolojik proses terimleri ise; 9 contig, “protein folding” (Tek zincirli polipeptidlerin veya oklu alt niteleri olan komplekslerin kovalent ve kovalent olmayan bileřimlerine yardımcı prosesler), 8 contig, “response to salt stress” (tuz stresinden sorumlu proseslere ait enzimler), 8 contig “coenzyme metabolic process” (enzimatik bir reaksiyonu yrtmek iin, protein olmayan eřitli organik kofaktrleri gerektiren, bir substrat ve bir enzim iin ek olarak koenzimleri ieren kimyasal reaksiyonlar ve sentez yolları), 7 contig “lipid biosynthetic process” (lipitlerin oluřmasını saęlayan kimyasal prosesler), 7 contig “phosphorylation” (bir fosfat grubunu bir biyomolekle baęlayan proses), 7 contig “response to cold stress” (soęuk stresinden sorumlu proseslere ait enzimler), 6 contig, “biopolymer biosynthetic process” (biyopolimerlerin oluřumundan sorumlu kimyasal reaksiyonlar), 6 contig “response to oxidative stress” (sper oksit anyonlar, hidrojen peroksitler ve hidroksil radikalleri gibi reaktif oksijen trlerinin yarattıęı oksidatif bir stres sonucu bir hredeki veya organizmadaki bir aktivitenin veya bir ařamanın deęiřimi), 6 contig “protein catabolic process” (peptit baęlarının hidroliz edilmesiyle veya hidroliz edilmeden aktif kofigrasyonunun ya da doęal yapısının bozulmasıyla protein yıkımını saęlayan prosesler), 6 contig “response to metal ion” (bir metal iyonu uyarını sonucu bir hredeki veya organizmadaki bir aktivitenin ya da bir ařamanın deęiřimi), 6 contig “cellular di-,tri-valent inorganic cation homeostasis” (hre seviyesinde, divalent veya trivalent katyonların i dengesinin bakımını kapsayan prosesler), 6 contig “metal ion transport” (hreler arasında ya da hre iinden dıřına veya hre dıřından iine elektrik ykyle birlikte herhangi bir metal iyonunun doęrudan tařınması), 6 contig “RNA metabolic process”

(RNA moleküllerini içeren prosesler), 6 contig “secondary metabolic process” (büyüme ve hücre onarımı için gerekli olmayan, genellikle bir taksona özgü olan kimyasal değişiklikleri kapsayan proseslerdir), 5 contig “regulation of transcription” (RNA’yı kalıp olarak kullanan DNA veya DNA’yı kalıp olarak kullanan RNA sentezleri sırasında sentezin oluşumunu, oluşum oranı ve oluşum sıklığını düzenleyen proseslerdir), 5 contig “establishment of cellular localization” (bir protein kompleks veya bir organel gibi hücre içi bir molekülün spesifik yerinde veya hücre zarında taşınması veya onarılması için gereken prosesler) şeklinde sıralanmaktadır.

Şekil 3.20.’deki ikinci grafik ise 3734 EST’nin tamamından oluşturulan 299 contig için bulunan biyolojik proses GO terimlerinin ayrıntılarını vermektedir (Şekil 3.20.).





Şekil 3.21.: Kütüphaneye ait contiglerin biyolojik proseslerine göre sınıflandırılması.

3.3 SONUÇLARIN GENBANK VERİTABANINA KAYDEDİLMESİ

EST dizileri fragment birleştirme işleminden önceki şekilleriyle batch-sub@ncbi.nlm.nih.gov adresine dbEST veritabanına kaydedilmek üzere istenilen formata göre gönderildi (Yapılan bu çalışmayla ilgili yayın bilgileri, kütüphane bilgileri, araştırmacıya ait iletişim bilgileri ve EST dizi bilgilerinden oluşan format). Gönderilen EST'ler için GenBank tarafından belirlenen, GO242703'dan GO246436'ya kadar olan accession numaraları yine mail yoluyla elde edildi. Kaydedilen bilgiler 2009 yılı temmuz ayından itibaren erişime açıktır (Ek 1).

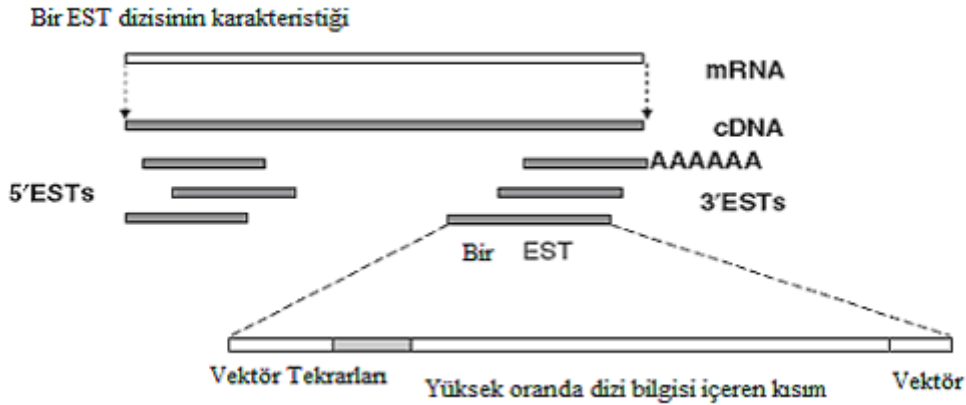
Consed programından elde edilen bilgilere göre contig dizilerinin içeriklerini anlatan yani hangi accession numarasına sahip hangi EST dizisinin hangi contiglerde yer aldığına dair bilgileri içeren bir dosya hazırlanmıştır (Ek 2). Hazırlanan bu dosya ile sequin programı aracılığıyla oluşturulan contig dizilerine ait seq formatlı bir dosya gb-sub@ncbi.nlm.nih.gov adresine gönderildi. Bu şekilde contig dizileri de GenBanka kaydedilmiş oldu.

4. TARTIŞMA

4.1 EST ANALİZLERİ

Genom çalışmaları sırasında genler hakkındaki bilgilere en hızlı ulaşma yolu olan EST'lerin doğru analiz edilmesi çalışma sonuçlarının güvenilirliği için son derece önemlidir. Bir organizmadan elde edilen cDNA klonlarının o organizmanın işlevsel genlerine ait bilgilere giden en kısa yol olması nedeniyle cDNA dizisinin EST analizleri ile incelenmesi uzun soluklu genom çalışmalarının doğrultusunu belirlemek bakımından oldukça önemlidir. Farklı bilgisayar programlarından yararlanılarak tüm EST analiz aşamaları tamamlandığında çalışılan organizmanın genomu ile ilgili ipuçları sağlanmış olur. Kesin bilgilere ulaşmak için bu ipuçları değerlendirilir.

EST üretimi sırasında sıklıkla hatalarla karşılaşmak mümkündür. mRNA'nın kısa bir kopyası olan EST dizisi uç kısımlarda yüksek oranda hata içermeye eğilimlidir. Genellikle EST dizilerinin kalitesi orta kısımlarda daha iyidir. EST'lerin işlenmesi sırasında vektör ve tekrar dizileri dizilerin sonlarında veya nadir olarak da ortalarında bulunmaktadır (Şekil 4.1.)



Şekil 4.1.: Bir EST dizisi genellikle vektör kontaminasyonu, bölünemeyen tekrar olasılıkları veya düşük kaliteli karmaşık bölgeler ile başlar ve biter (Nagaraj ve ark., 2006).

Genel olarak, EST dizilerinin okunması (baz adlandırılması) sırasında başlangıçta düşük değerler (Her 100 baz çiftinde % 20 veya %50'nin üzerinde), sonrasında aşamalı olarak daha iyi değerler, son kısımlarda ise bir kez daha düşük değerler alınır (Aaronson ve ark., 1996).

Bu hatalı veya düşük kaliteli, vektör veya adaptör parçalarından ham EST ler ön işleme yapılarak temizlenir ve bu işlemin kontrolü gerçekleştirilir. EST dizilerinin sadece cDNA dizisine ait nükleotidleri taşıdığından emin olduğunda, EST analiz aşamalarına devam edilir. İkinci aşamada eldeki işlenmiş EST'lerin bir birleri ile olan yakınlıklarına bakılır. Bir EST dizisi bir cDNA klonunu yansıtır ancak klon bir genin dizi bilgisinin tam olarak karşılamayabilir. Aynı genin farklı bölgelerinden eksprese olan ürünler klonlanmış olabileceği gibi, genin farklı boylarda fragmentleri klonlanmış olabilir. cDNA'ları ifade eden EST'ler biyoinformatik programları ile bir araya getirildiğinde bu gene ait dizi bilgisi ortaya çıkmaya başlayacaktır. İşlenmiş EST dizilerinin örtüşen baz dizileri belirlenip, birleştirildiğinde gen olma olasılığı olan contig diziler oluşur. Oluşturulan contiglerin fonksiyonel analizleri için daha önce yapılan genom çalışmalarından elde edilen dizi bilgileri ile elimizdeki dizi bilgilerinin karşılaştırılması yapılarak muhtemel genlerin tahimini fonksiyonları bu benzerlikler ışığında ortaya çıkarılır.

EST'ler değişken ve çoklu uygulamalara sahiptir. İnsan genomunun haritasını oluşturmak için ilk kez kullanımının (Wilcox ve ark., 1991) ardından, tek bir EST dizisinin gen kapsamında değerlendirilmesi (Aaronson ve ark., 1996) ile gen temelli markırlar ile haritalama yapılabildi (Boguski ve Schuler, 1995). Hızla artan genomik veriler ile, EST veritabanları; gen yapısının tahmini için (Kan ve ark., 2001; Jiang ve Jacob, 1998), alternatif kesmelerin (splicing) incelenmesi için (Brett ve ark., 2000; Mironov ve ark., 1999; Modrek ve ark., 2001; Kan ve ark., 2004), dokuya veya hastalığa özgü gen anlatımını ayırt edebilmek için (Schmitt ve ark., 1999) ve olası SNP'lerin keşif ve karakterizasyonu için (Buetow ve ark., 1999; Picoult-Newberg ve ark., 1999; Useche ve ark., 2001) kullanılmaktadırlar.

EST temelli gen anlatım protokolleri, geniş ölçekte yan anlatımı yapılan genlerin tanımlanmasında ve analizinde kullanılmaktadır (Ewing ve ark., 1999; Heyer ve ark., 1999). Ewing ve arkadaşları (Ewing ve ark., 1999) doku türleri ve organlara göre sınıflandırılmış dbEST'den gelen pirinç genomunun dijital gen anlatım profillerini üretmektedirler. Bu silico (sanal) differential display tekniği çeşitli dokularda farklı anlatım yapan (Hawkins ve ark., 1999) ve normal ile hasta bireyler (Schmitt ve ark., 1999) arasında farklı olan genlerin belirlenmesini sağlar. Son EST projelerinde Qiu ve arkadaşları (Qiu ve ark., 2003) DNA 'çubuk kodlama' uygulayarak, spesifik mRNA havuzlarından gelen EST'leri izlemek için farklı 6 baz çiftlik DNA dizilerini kullanarak mısır cDNA kütüphanelerini kategorize ettiler. EST'ler ayrıca proteomik alanında peptid tanımlanması ve tam genom dizi bilgilerinin içinde kısmen bulunan proteomların proteom karakterizasyonu için çok değerli kaynaklar olmaktadır

(Lisacek ve ark., 2001; Kim ve ark., 2003). EST sıralama stratejileri ile birlikte diğer genomik ve proteomik yöntemleri (transkripsiyonu profillemeye ve peptid parmak izi) gen ve allel belirlenmesi için kullanılmaktadır (Nelson ve ark., 2005; Mooney ve ark., 2004). Genel olarak, EST verilerinin kullanılabilirliği gen bulmada ve transkriptom analizinde kullanılan klasik çalışmalardan farklıdır (Nagaraj ve ark., 2006).

Bu çalışmada Gemlik zeytin bitkisinin genomu hakkında fikir sahibi olabilmek için bu organizmaya ait EST'ler ve bu EST'lerin değerlendirilmesi amaçlanmıştır. Amaca uygun olarak Gemlik zeytinine ait 3840 EST elde edilmiştir ve bu EST'lerin analizleri yapılmıştır. Çalışma sırasında EST analizleri aşamaları için kullanılan bilgisayar programları da değerlendirilmiştir. Böylece karşılaşılan sıkıntılar, seçilen programın çalışmaya uygun olup olmadığı anlaşılmıştır. Uygun programların seçimine bazen uygulamalar yapılarak, bazen de yapılmış uygulamalarla ilgili kaynaklar değerlendirilerek karar verilmiştir.

4.1.1 Est Analiz Programları

Büyük ölçekli genomların EST analizlerinde, dizilerin fonksiyonlarını anlamak ve bu fonksiyonlara göre dizileri gruplandırmak gibi aşamaların hızlı bir şekilde yapılması için otomatik analiz araçlarına ihtiyaç duyulmuştur. Bilgisayar programlarının art arda, threshold (eşik değeri) kriterlerine göre dikkatle parametrelenmesi, EST analiz aşamalarının bilgisayar tarafından otomatik olarak güvenilir sonuçlar verecek şekilde yapılmasını sağlar. Böyle bir protokol bir veritabanından veya ham kromatogram dizisinden gelen EST'leri, otomatik olarak temizleyen, kümeleyen, birleştiren, contig üreten, kavramsal olarak çevrim yapan ve çeşitli DNA ve protein benzerlik araştırmalarına dayalı olarak varsayılan fonksiyonu atayan bir prosedir. Bu proses sırasında araştırmacı sadece, aşamaları sıralayarak, art arda otomatik olarak ilerlemenin sağlanması için bilgisayara komut verir.

EST analizi aşamaları için farklı bilgisayar programları kullanılabilir. Bunun için, her aşamada seçilebilecek seçenekler belirlenmeli ve değerlendirilmelidir. Bazen farklı seçeneklerin kombinasyonu ile alternatif uygulamalar da yapılabilir. Örneğin, CAP3 veya Phrap gibi bazı güvenilir programlar EST analizinde ve EST analiz aşamalarında, her ikisinin de avantajlarını kullanabilmek için kullanılmaktadır. Çizelge 4.1.'de EST analizleri aşamaları, bu aşamalar için önerilen bilgisayar programları kombinasyonları ve bu programlara ulaşılabilecek internet adresleri (web siteleri) verilmiştir (Çizelge 4.1.).

Çizelge 4.1.: EST analiz aşamalarının karakteristiği (Nagaraj ve ark., 2006).

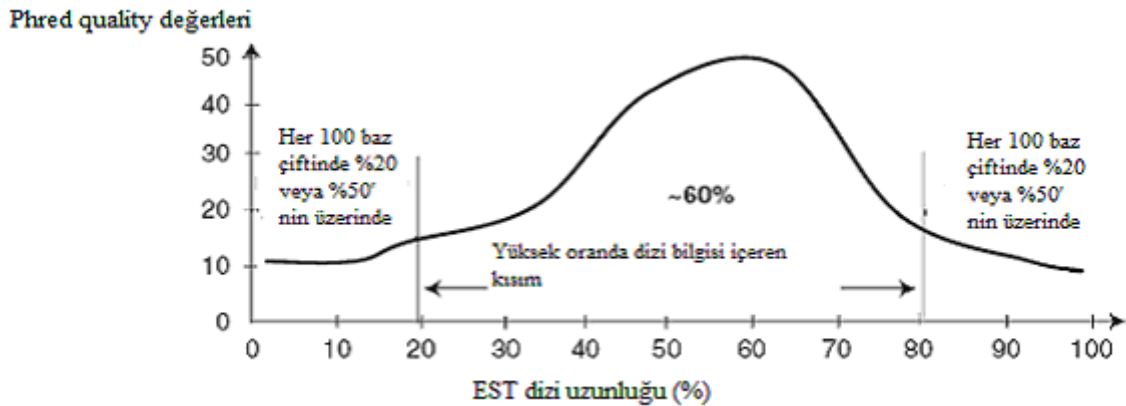
Kaynak Adı	Web site	Ön-İşleme programları	Kümeleme ve birleştirme programları	Fonksiyonel analiz programları
Edinburgh EST-Pipeline	http://zeldia.cap.ed.ac.uk/PartiGene/index.html	Phred; Cross_Match	CLOBB; Phrap	DECODER; ESTSCAN
EST Analysis Pipeline (ESTAP)	http://staff.vbi.vt.edu/estap/	Phred; Cross_Match	D2_cluster; CAP3	BLASTX
ESTAnnotator	http://genome.dkfz-heidelberg.de/menu/biounit/dev.shtml#estannotator	Phred; RepMask with UniVec data	CAP3	BLASTX
ESTIMA	http://titan.biotech.uiuc.edu/ESTIMA/	Bilinmiyor	BlastClust; CAP3	BLASTX
ESTweb	http://bioinfo.iq.usp.br/estweb/	Phred; Cross_Match		
Nematode.net	http://nematode.net/	Phred; Cross_Match	Phrap	BLASTX
NemaGene				
Clusters				
PipeOnline	http://bioinfo.okstate.edu/pipeonline/	Phred; Cross_Match	Phrap	BLASTX
The TIGR Gene Indices (TGICL)	http://www.tigr.org/tdb/tgi/	SeqClean; megaBLAST	CAP3; Paracel TranscriptAssembler®	DIANA-EST; ESTscan; Framefinder

Daha hızlı EST analizi aşamaları için seçilen kaynaklar akademik kullanım amacıyla serbestçe internet ortamından indirilebilir olması avantajdır. Ayrıca, bu programlar diğer EST analiz aşamaları için kullanılan programlara uyumlu olmalıdır. Örneğin; bir program için input dosyası olarak kullanılacak dosyaların formatı FASTA olması gerekiyor ise, bir önceki EST analiz aşamasının output dosyaları bu duruma uygun olarak FASTA formatında dosyalar olmalıdır.

4.2 EST ÖN İŞLEME

Başlangıç aşaması olan ön-işleme daha sonraki analiz etkinliğini artırmak için EST verilerindeki istenmeyen fragment parçalarının ve kalitesiz parçaların hepsini yok eder.

Vektör kontaminasyonu EST'lerde yaygındır ve genellikle vektörün bir bölümü de EST dizileri ile birlikte yer alır (Şekil 4.1.). Bu vektör parçalarının, EST'lerin birleştirilme aşamasında (fragment birleştirme) kümelenmesinden önce uzaklaştırılmış olması gerekmektedir. BLAST kullanımı ile UniVec ve EMVEC gibi vektör veritabanlarıyla EST'lerin karşılaştırılması (Altschul ve ark., 1997) veya Cross_Match (Smith ve Green, yayınlanmamış çalışma (Ewing ve Green, 1998)) gibi bir program ile ortadan kaldırılabilmek üzere vektör kontaminasyonları saptanabilir. EST verilerindeki tüm düşük kaliteli bölgeler NCBI veya nseg (Wan ve ark., 2003) içindeki DUST kullanılarak taranabilir ve maskelenebilir. Bu çalışmada zeytin EST'lerinin analizi için Phred programının (skor (q) = 20) çalıştırılması sonucu vektör kontaminasyonlarından kurtulunamamıştır. Ancak daha sonra phrap programı içersindeki cross-match uygulaması ile bu sorun giderilmiştir. Cross_match, EST dizilerindeki vektörleri bularak maskeleymiştir. Bunun için Phrap programı içersindeki vektör dosyasına bu çalışmada kullanılan pDONR 222 vektörünün baz dizisi FASTA formatında yazılmıştır. Cross-match uygulamasından sonra pDONR 222 vektörüne ait baz dizileri “X” şeklinde ifade edilmiştir. Bu “X” şeklinde ifade edilen kısımlar EST dizilerinden tamamen silindikten sonra zeytin EST'leri ön işleme aşamasını tamamlamıştır. Fakat düşük değerli dizilerin belirlenmesi ve kromatogram dosyalarının FASTA formatında elde edilmesi başta uygulanan Phred programı ile sağlanmıştır.



Şekil 4.2.: Phred quality skorları. Tahmini bir EST dizisi için dizi uzunluğunun fonksiyonu grafiklenmiştir. Bu çalışmada phred skoru 20 alınarak %1 hata yapma oranı kabul edilmiştir (Nagaraj ve ark., 2006).

Phred skorları (Ewing ve Gren, 1998) daha iyi dizi kalitesini karşılamak için en yüksekten başlayan bir sıralama yaparak dizi kalitesinin ölçümünü sağlar. Phred uygulamasında piklerin çevresinde her bir baz isimlendirmesi sırasında yapılan quality skor atamasında her bir baz isimlendirmesi için logaritmik olarak hata olasılıklarına bağlanılır. 4'den 40'a kadar olan quality skor çeşitleri ve hesaplanışı;

$$q = -10 \times \log_{10}(P) \text{ şeklindedir.}$$

q= Phred quality skor

P= Baz isimlendirmesinin hata oranı

Buna göre, Phred için, quality skor 30 olarak belirlenirse hata oranı 1/1000 olacaktır, skor 20 olursa bu oran 1/100, 10 olursa 1/10 olacaktır. En sık kullanılan skor 20 ve üzeri değerlerdir. Bu çalışmada da skor 20 olarak alınmıştır. EST dizilerinin bir sonraki aşamaya hazır hale gelip gelmediklerini anlamak için NCBI web sitesinde bulunan VecScreen programı ile EST'lerin işlenmiş halleri vektör dizileri içerip içermedikleri bakımından kontrol edilmiştir.

%5 kadarlık baz isimlendirmesi hataları (Aaronson ve ark., 1996), baz kekelemeleri (tekrarlanan bazlar, özellikle G ve T) gibi dizileme kısımları ve düşük kaliteli diziler EST'lerde sık gözlemlenen hatalardan bazılarıdır. Bu kısımlarda vektörlerden, bağlayıcılardan, adaptörden, kimerik dizilerden ve ayrıca genomik DNA parçalarından kaynaklanan kontaminasyonlar olabilir. Ayrıca, düşük kaliteli dizilerin bağlanması, tekrarlar (basit veya peşpeşe olan(tandem)), kısa dizi uzunlukları ve açıklama hataları analiz boyunca sorun çıkarabilir. Dahası, RNA düzenleme ve genomik çeşitler gibi doğal dizi çeşitleri ve SNP'ler yapay olarak oluşturulan ve doğal olarak meydana gelen dizilemelerin arasındaki yer değiştirmelerin ve verilen EST verilerindeki ekleme / silme olaylarının belirlenmesinde sorun çıkarabilirler (Nagaraj ve ark., 2006). Genom çalışmalarından EST'leri doğru şekilde kullanmak bir avantajdır. Çalışmanın hatasız olması en çok da EST'lerin işlenmesine bağlıdır. Bu yüzden EST ön- işleme aşamasına dikkat edilmelidir.

LINE'ler (Long interspersed nuclear elements), SINE'ler (Short interspersed nuclear elements), LTR'ler (Long Terminal Repeat) ve SSR'ler (Simple Sequence Repeats) gibi tekrar elementleri dizilerin birleşmesinde hata oluşumuna öncülük edebilir. Düşük kaliteli DNA

dizileri ve ara tekrarlar bakımından DNA dizilerinin taranması için analiz boyunca RepeatMasker (Ewing ve Gren, 1998) veya MaskerAid (Bedell ve ark., 2000) kullanımı ile tekrarların maskelenmesi yapılmalıdır. Poly (A) genomik diziler içinde kodlanmamaktadır bu yüzden kümeleme ve birleştirme proseslerinde yüksek kalitede EST elde etmek için geri kalan birkaç adenin kesilmelidir (genellikle 6-10) (Nagaraj ve ark., 2006) EST'ler için yalnızca ön işleme aşamasında kullanılan bilgisayar programları ve bu programların elde edilebileceği web siteleri Çizelge 4.2.'de görülmektedir (Çizelge 4.2.).

Çizelge 4.2.: EST ön işleme için kaynaklar.

Kaynak Adı	Web Adresi	Tanım
UniVec	http://www.ncbi.nlm.nih.gov/VecScreen/UniVec.html	Vektör veritabanı
EMVec	http://www.ebi.ac.uk/blastall/vectors.html	Vektör veritabanı
Phred/Cross_Match	http://www.phrap.org/	Baz isimlendirici/ vektör uzaklaştırıcı
Trimest	http://emboss.sourceforge.net/apps/#Apps	Poly (A) kuyruğu uzaklaştırıcı
Trimseq	http://emboss.sourceforge.net/apps/#Apps	Belirsiz uç kısımları uzaklaştırıcı
VecScreen	http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html	Vektör, bağlayıcı ve adaptör tanımlayıcı
Vector cleaning	http://www.aborygen.com/products/biOpen/tools-for-biOpen/vector-cleaning.php	Vektör temizleyici
Vectorstrip	http://emboss.sourceforge.net/apps/#Apps	Vektör dizileri arasındaki DNA dizilerini ayıklayıcı
Parcel	http://www.parcel.com/	EST ön-ışleme paketi
Lucy 2	http://www.complex.iastate.edu/download/Lucy2/index.html	Dizi kesimi ve görüntülenmesi
Dust	ftp://ftp.ncbi.nih.gov/blast/	Düşük değerlikli bölgelerin maskelenmesi
The TIGR Plant Repeat databases	http://www.tigr.org/tdb/e2k1/plant.repeats/	Bitki tekrar dizileri veritabanı
Masker Aid	http://blast.wustl.edu/masker/aid/	Tekrar dizileri maskeleyici
RepeatMasker	http://www.repeatmasker.org/	Tekrar dizileri maskeleyici

Parcel ve vektor cleaning ücretli diğer kaynaklar ücretsizdir.

4.3 EST BİRLEŞTİRME

EST ön işleme aşaması tamamlandıktan sonra işlenmiş EST dizileri ikinci aşama olan fragment birleştirme işlemine hazır hale gelmiş olur. Bu aşamada işlenmiş EST'ler öncelikle birbirleriyle örtüşen baz dizilerine göre gruplar oluştururlar. Bu kümeleşmenin amacı fazlalıkları azaltmak için tek bir genin aynı transkriptinden gelen ve üst üste binen EST'lerini tek bir küme içinde toplamaktır. Bir kütüphaneye ait olan EST topluluğu izole edildiği genomun dağılmış parçalarıdır. Her bir EST kümesi ise bu genomun herhangi bir genine ait bilgilerin sınıflandırıldığı gruptur. EST kümelemenin basit bir yolu bir çift dizi arasındaki benzerliğin ölçülmesi ile sağlanabilir. Daha sonra, bu mesafe ikili değerlere dönüştürülür ve önemli bir eşleşme olsun veya olmasın bağlanır, bu şekilde dizi çifti olarak kabul edilir veya birleştirilmiş olan kümelerden çıkarılır. Ptitsyn ve Hide EST kümelemesi için farklı aşamaları detaylı olarak açıklanmış ve EST kümeleme için 'sıkı' ve 'gevşek' olmak üzere iki yöntem önermişlerdir (Ptitsyn ve ark., 2005). Sıkı kümeleme yöntemi kapalıdır, Her bir kümedeki EST grubunun birbirleriyle olan ilgisi 5 've 3' uçlarından tek yönlü, hızlı dizilemenin yapılması esasına dayanır. Anlatımı yapılan genle dizi uzunluğu bakımından daha az benzerlik gösteren, daha kısa contigler sıkı kümeleme yöntemiyle oluşur. Buna karşılık, gevşek kümeleme sıra dışı ve düşük kaliteli EST dizi tekrarlarının tam bir dizi oluşturmak için birçok kez birleşmesi esasına dayanır ve contigler daha uzundur. Sonuç olarak, anlatımı yapılan gen verilerinin ve alternatif olarak kesilmiş transkriptlerin daha iyi kapsama alanı var, fakat kümeler içindeki paralogların (Benzer fakat özdeş olmayan fonksiyonları kodlayan homolog genler) varlığı riski bulunmaktadır. Örneğin; TIGR Gen Endeks (Lee ve ark., 2005) sıkı kümelemeyi kabul ederken stackPACK (Miller ve ark., 1999) gevşek kümeleme yöntemi kullanılarak tasarlanmış fragment birleştirme programlarıdır. Phrap (Ewing ve Green, 1998) ve CAP3 (Huang ve ark., 1999) ise, dizi kümeleme ve birleştirme işlemlerinde en yaygın kullanılan programlar arasında yer alırlar. Quackenbush ve arkadaşları (Liang ve ark., 2000) Phrap, CAP3 ve TIGR Assembler programlarını 118 000 sıçan EST veri setiyle karşılaştırmıştır. EST birleşiminde, birleşme ve algoritmalara bağlı kısımların doğruluğunda ve EST kümelemesi için araçların kritik değerlendirmesinde dizileme hatalarının etkilerini incelemişlerdir. Onların sonuçları göstermektedir ki, CAP3 tutarlı bir şekilde benzer programlara göre yüksek doğruluklu contig dizileri üretirken ve dizileme hatalarını etkin olarak kullanırken aynı zamanda gen ailesi üyelerinin belirlenmesi için de yüksek hassasiyeti bakımından daha iyidir. Wang ve arkadaşları tarafından yapılan daha yeni bir çalışmaya (Wang ve ark., 2004) göre ise, CAP3 EST kümeleme programından kaynaklanan iki tip hata gerçekleşmektedir. Buna

göre; aynı genden gelen EST'lerin aynı küme içinde yer almamaları (1.Tip hata) ve farklı genlerden gelen EST'lerin yanlış bir şekilde aynı küme içinde yer almaları (2.Tip hata) söz konusu hatalardır.

EST kümelene, birleştirme ve contig üretimi için ulaşılabilir farklı kaynaklar mevcuttur (Çizelge 8). Miller ve arkadaşları tarafından STACK dizi işleme eşleştirme ve contig bilgi bankası sistemi oluşturulmuştur (Miller ve ark., 1999). STACK ve diğer kaynaklar arasında en büyük fark, farklı gen anlatımlarının analizleri için EST verilerinin dokuya özel sınıflandırılmasının yapılmış olmasıdır. Bu sistemde 15 temel kategorideki doku ve bir hastalık kategorisi mevcuttur.

STACK, gevşek yöntemi kullanır, daha geniş bir gen kapsamı için ön işleme grubundaki EST'ler kontrol edilmemiş kümeleme stratejisine tabi tutulur. d2_cluster (baz çifti olmayan bir eşleşme algoritması) (Burke ve ark., 1999), Phrap ve CRAW programları STACK içinde sırasıyla kümeleme, birleştirme ve EST analiz eşleşmeleri için birleştirilmiştir. Sonra contigler klon-tanımlanması verilerini elde etmek amacıyla iyi kabul edilen tahmini geni sunmak için birleştirilir Çizelge 8'de EST dizilerini birleştirmek için kullanılacak programlar ve bu programlara ulaşılmasını sağlayabilecek web siteleri verilmiştir (Çizelge 4.3.) (Nagaraj ve ark., 2006).

Çizelge 4.3.: EST dizilerini birleştirmek ve contig oluşturmak için kullanılan programlar.

Kaynak Adı	Web sitesi
CAP3	http://genome.cs.mtu.edu/cap/cap3.html
CLOBB	http://zeldia.cap.ed.ac.uk/CLOBB/
CLU	http://compbio.pbrc.edu/pti
ESTate	http://www.ebi.ac.uk/~guy/estate/
ESTs aSSEmby using Malig	http://algggen.lsi.upc.es/recerca/essem/frame-essem.html
megaBLAST	ftp://ftp.ncbi.nih.gov/blast/
miraEST	http://www.chevreux.org/projects_mira.html
Paracel Transcript Assembler	http://www.paracel.com/
Phrap	http://www.phrap.org/
stackPACK	http://www.sanbi.ac.za/Dbases.html#stackpack
Xsact ve Xtract	http://www.ii.uib.no/~ketil/bioinformatics/

Paracel dışındaki kaynaklar akademik kullanım için ücretsizdir.

Fragment birleřtirme ařamasında kullanılan programlardan biri olan TIGR Gen endeksleyicinin alıřma prensiplerine bakılırsa; (TGI Kmelenme araları) (Pertea ve ark., 2003) yksek sıklıktaki bir yntem dizi benzerliklerini oluřturmak iin ilk kmelenme sreci sırasında uygulanır. Bu iřlem aynı zamanda birleřebilecek dizilerin tanımlanmasını ve baėlantı yerlerinin ayrılmasını daha sonra bireysel kmelerinin genellikle mRNA veya genomik dizi referanslı (Quackenbush ve ark., 2000) sanal transkriptleri veya geici contig dizileri oluřturmasını, daha uzun dizileri retmek iin birleřmesini saėlar. Contigler daha sonra ORF ve SNP gibi durumları belirlemek, microarrayler iin uzun oligonkleotidleri belirlemek, varsayılan sonular kullanılarak kontroll bir anlam oluřturmak, GO ve enzim numarası atamaları ve tamamlanmıř veya taslak genomlar ve kullanılabilir genetik haritalar iin haritalama yapmak iin deėerlendirilebilir. Elde edilen contig dizilerinin daha sonraki alıřmalara da kaynak olabilmesi iin GenBank'a kaydedilmesi doėru olacaktır (Nagaraj ve ark., 2006).

CAP3' geliřtiren arařtırmacılar tarafından (Huang ve ark., 1999) CAP3 ve PHRAP verileri iki ynden karřılařtırılmıřtır: Forward-reverse kalıpları ile dřk deėerlikli veriler ve forward-reverse kalıplarını iermeyen eřitli veriler. Dřk deėerlikli verilerin deėerlendirilmesinde en nemli řey (beř kat kapsamlı) CAP3 ve PHRAP ile EST'lerin kmeleřmelerinin mmkn olup olmayacaėını belirlemektir. İkinci tip verilerin deėerlendirilmesindeki en nemli řey ise contiglerin sayısı ve doėruluėu bakımından CAP3 ve PHRAP performanslarının deėerlendirilmesidir. Uygulamada kullanılan BAC veri setleri dřk deėerlikli verileri de iermektedir. Forward-reverse kalıpları her plasmid subklonunun ift ynl dizilemesi ile elde edilmiřtir. CAP3 programı ile yedi BAC setinin tamamında EST'ler arasında contig dizisi oluřturmak amacıyla kmeleřme oluřturulabiliyorken PHRAP ile altı veri setinde kmeleřme oluřturulabilmektedir. nk CAP3 ile forward-reverse kalıpları tarafından contigleri oluřturacak kısmi bir dzen saėlanır, bylece daha kolay ve hızlı kmeleřme oluřturulur. Her kmeleřmenin doėruluėu bu kmelerden oluřturulan contig dizisi ile kmeyi oluřturan EST dizi gruplarının karřılařtırılması ile kontrol edilir. Genom Dizileme Merkezindeki Shiao-Pyng Yang da (Washington niversitesi), CAP3 ve PHRAP'ın performansını eřitli boyutları kapsayan veri setleri zerinde deėerlendirmiřtir (S.-P. Yang, Pers. Comm. –yayınlanmamıř alıřma-). Bu veri setleri H NH0018C09 adlı gerek bir veri setinden istenen boyut kapsamında rastgele retilmiřtir. Gerek veri seti yeni kapiller makinelerde retilmiřtir. Phred tm diziler iin quality deėerlerini oluřturmak zere

kullanılmıştır. PHRAP'ın en son sürümü, 3/19/99 versiyonu, değerlendirilmede kullanılmıştır. Hiçbir forward-reverse kalıbı kullanılmamıştır.

Genom Dizileme Merkezindeki çalışmada kullanılan 16 veri seti için CAP3 ve PHRAP çalıştırılmış ve consensus diziler üretilmiştir. 167,358 baz çiftlik contig dizisi için CAP3 ve PHRAP programlarının performansı karşılaştırılmıştır. Bu veri setlerinde PHRAP genellikle CAP3'den daha uzun diziler üretirken, CAP3 oluşturduğu EST kümelerinde PHRAP'a göre daha az hata göstermektedir. Fakat CAP3 dizi sonlarında PHRAP'dan daha fazla hata oluşturmaktadır. Bu sonuca göre, CAP3'e ait her bir contig dizisinin 5' ve 3' uçları pek çok hata taşıyan düşük değerli uç bölgeleridir. Bu problem uç bölgelerin kesilip atılması ile giderilmiştir.

CAP3 EST birleştirme programının daha önceki benzerlerine göre çok sayıda yeni özelliği bulunmaktadır. Bu özellikler, dizi çiftlerinin hızlı tanımlanması, Dizinin 5' ve 3' kalitesiz uç bölgelerinin kesimleri, hesaplamaların değerlendirilmesi, contiglerin yapısındaki hataları düzeltmek için yapılan forward-reverse kalıplarının kullanımı ve contigler için consensus (EST kümesi) dizilerinin üretimi şeklinde sıralanabilir (Huang ve ark., 1999).

Birleştirme hatalarının düzeltilmesi amacıyla forward-reverse kalıplarının kullanımı da CAP3'ün farklı bir özelliğidir. Bu özelliği sağlayan algoritma sayesinde, kalıplardaki hatalar daha fazla ihmal edilir. Ancak düzeltme için yeterli sayıda kalıp tarafından destekleniyorsa algoritma geçerli birleştirmeye yönelik bir düzeltme yapmaktadır. Farklı dizi birleştirme programları farklı yaklaşımlar kullanmaktadır. CAP3, contigleri kurmak için yapılan kalıpları ve çoklu dizi eşleşmelerini oluşturmak ve quality değerlerini kullanmak bakımından diğer EST birleştirme programlarından ayrılmaktadır (Huang ve ark., 1999).

CAP3 contigler oluşturmak üzere çoklu dizi eşleşme metodu kullanmaktadır. Çoklu dizi eşleşme metodu quality değerleri tarafından büyüyen baz ağırlıklarına göre sıralanmaktadır. PHRAP ise contig üretmede farklı bir yaklaşım kullanmaktadır. "Dizi yüksek quality değerleri", dizilerin bir mozaığıdır. Mozaik yaklaşımında ve çoklu eşleşme yaklaşımında hatalar yer yer görülebilmektedir. Çoklu eşleşme ile ara sıra görülen bazların yanlış adlandırılması problemine mozaik yaklaşımda da rastlanmaktadır. (Huang ve ark., 1999).

Bir EST birleştirme programı kullanılacağı zaman CAP3 mü yoksa PHRAP'mı kullanılmasına karar verirken; quality değerleri bilinmiyorsa, kullanıcı contig dizilerin kuruluşunda gereksiz kısımları da kullanabilen CAP3'ü seçebilmektedir. Aksi takdirde,

kullanıcı PHRAP'ı seçebilir. PHRAP sık sık CAP3'ten daha uzun contigler üretmektedir, oysa CAP3 ortak dizilerde çoğunlukla PHRAP'tan daha az hata üretmektedir (Huang ve ark., 1999).

Bu çalışmada aynı EST dizisi verileri için önce Phrap daha sonra CAP3 programı kullanılmıştır. Consed programı ile her iki programın oluşturduğu contig dizileri incelenmiştir. Genom Dizileme Merkezinden Shiao-Pyng Yang'ın (Washington Üniversitesi) ve CAP3'ün tasarlayıcısı Huang ve arkadaşlarının gerçekleştirdiği PHRAP-CAP3 karşılaştırma uygulamalarında rastlanan sonuçlara bu çalışmada da rastlanmıştır. Phrap programı ile görülen hataların CAP3 programı ile görülmemesi, CAP3 içerisinde birleştirme hatalarının düzeltilmesi amacıyla forward-reverse kalıplarının kullanımını sağlayan algoritmanın varlığı ile açıklanabilir.

Sonuçların güvenilirliği bakımından değerlendirildiğinde CAP3 tercih edilen program olsa bile kullanım açısından da Phrap programının sahip olduğu avantajlar mevcuttur. Phrap programı Phred/Phrap paket programı şeklinde elde edildiğinden birden fazla uygulama için fonksiyoneldir. Bu paket içinde düşük değerli dizilerin uzaklaştırılması, vektör ve adaptör dizilerin atılması ayrıca kromatogram dosyalarının FASTA'ya dönüşümü için Phred programı ve Cross_match uygulaması bulunur. Böylece EST birleştirme aşamasından önce ham EST'nin en kolay şekilde işlenmesi sağlanmış olur. CAP3 ve Phrap programının uygulanmasına hazırlık olarak zeytin EST'lerinin işlenmesi Phred ve Cross_match ile sağlanmıştır. Phred/Phrap program paketi elde edilirken talep edilen Consed programı birleşmiş EST'lerin izlenmesine olanak verir. Bu da hataların anlaşılabilmesi için önemlidir. Hem CAP3 hem de Phred/Phrap/Consed programları akademik kullanım amacıyla ücretsiz olarak internet üzerinden indirilebilen dosyalardır. CAP3 programına sahip olmak birkaç dakika sürerken Phred/Phrap/Consed programları birkaç gün süren bir süreçle elde edilebilir.

4.4 EST FONKSİYONLARININ ANALİZLERİ

EST analiz aşamalarında üçüncü aşama EST'lerin fonksiyonel analizleridir. EST'lerin birleşmesi ile contigler (tahmini genler) belirlenir, olası fonksiyonları ise; bilinen, serbestçe kullanılabilir araç ve veritabanlarında veritabanı benzerlik araştırması analizlerinin yapılmasıyla bulunabilir. NCBI'daki farklı BLAST (Altschul ve ark., 1997) programları veritabanı benzerlik araştırmaları için kayıtlı dizi sayısının büyüklüğü sebebi ile araştırmacıların

ilk tercihidir. EST'lerin nükleotid dizileri veritabanına karşı araştırılmasında BLASTN ve protein veritabanlarına karşı araştırılmasında BLASTX kullanılır. BLASTX EST'lerden oluşan contig dizisini (query) protein veritabanı ile karşılaştırabilmek için protein ürünleri içinde altı okuma penceresine (ORF) çevirir. Ayrıca, Conserved Domain Veritabanı (CDD) ve COG (Orthologous Gruplarının Kümelemesi) (Tatusov ve ark., 2003) veritabanı (Marchler-Bauer ve ark., 2002) ile protein domainleri için RPS-Blast (Ters PSI-Blast) (Marchler-Bauer ve ark., 2002) kullanılarak bir tarama yapılabilir. High-throughput EST analizi binlerce Blast sonuçlarının üretilmesini, çevirilmesini ve açıklanmasını gerektirdiğinden elde edilecek BLAST değerlendirmeleri bu EST'lerin fonksiyonel analizleri bakımından oldukça önemlidir (Nagaraj ve ark., 2006).

Ayrıca EST'ler transkriptom analizi için, organizmanın kendi genom dizileri (varsa) ile de eşleştiğinden veya yakın ilişkide olduğundan genomik haritalamayı ve gen keşfini kolaylaştırmak için özel eşleşme programlarını kullanırlar. Bu programlara nasıl ulaşılabileceği Çizelge 4.4.'de görülmektedir (Çizelge 4.4.). BLAT, GMAP ve MGALIGN uygulamaları da bu süreç için güvenilir yöntemler olarak kabul edilmektedir (Wu ve Watanabe, 2005).

Çizelge 4.4.: EST / cDNA dizilerinin genom DNA dizileri ile eşleştirilmesi için programlar.

Kaynak Adı	Web sitesi
BLAT	http://genome.ucsc.edu/cgi-bin/hgBlat
Est2genome	http://bioweb.pasteur.fr/seqanal/interfaces/est2genome.html
GMAP	http://www.gene.com/share/gmap/
MGAlign	http://origin.bic.nus.edu.sg/mgalign
SSAHA	http://www.sanger.ac.uk/Software/analysis/SSAHA/
Sim4	http://globin.cse.psu.edu/html/docs/sim4.html
Splign	http://www.ncbi.nlm.nih.gov/sutils/splign/splign.cgi

Tüm programlar akademik kullanım için ücretsizdir.

4.4.1 Fonksiyon Analizlerinin Sonuçlarının İncelenmesi

4.4.1.1 Blast Sonuçlarının İncelenmesi

Yaprak EST'leri, meyve EST'leri ve bütün EST'lerin tamamı için hem Phrap hem de CAP3 programı ayrı ayrı çalıştırılmıştır. Bu programlarla elde edilen her üç EST grubuna ait contig dizileri için BLASTN analizleri yapılmıştır. BLASTN ile elde edilen sonuçlar içinde, elimizdeki contiglerden baz çifti uzunluğu ve benzerliğinin örtüşmesi bakımından GenBank veritabanındaki nükleotid dizileri ile %50'den fazla benzerlik gösteren diziler belirlenmiştir. Ayrıca daha önceden veri tabanına kaydedilmiş zeytin bitkisine ait nükleotid dizileri ile örtüşen contigler de belirlenmiştir (Çizelge 3.4., Çizelge 3.5., Çizelge 3.6.). Bu belirlemeler 6 ayrı veri grubu için yapılmıştır. Veri grupları öncelikle yaprak EST'leri, meyve EST'leri ve her iki EST grubunun toplamı olan toplam EST'ler olarak üçe ayrılmıştır. Bu üç EST veri grubu önce Phrap programı ile birleştirilmiş ve üç ayrı contig dizi grubu elde edilmiştir. Daha sonra EST birleştirme işlemi aynı üç EST veri grubu için CAP3 programı ile yapılmıştır. CAP3 programı ile de üç ayrı contig dizi grubu elde edilmiştir. Böylece toplamda 6 ayrı contig dizi grubu oluşturulmuştur.

Bu veri grupları BLASTN sonuçlarına göre veritabanında ikiden fazla kez eşleştiği gen fonksiyonları bakımından birbirleriyle karşılaştırıldığında, hem CAP3 ve Phrap programlarının değerlendirilmesine hem de yaprak ve meyve dokularının bir arada ve ayrı ayrı incelenmesine farklı bir açıdan yaklaşım olmuştur. Blast sonuçları Çizelge 4.5.'de sadece CAP3 ve Phrap programından elde edilmiş ortak fonksiyonları, Çizelge 4.6.'da iki program arasındaki farklı Blast sonuçlarını ve son olarak Çizelge 4.7.'de ise, yaprak ve meyve dokuları arasındaki benzerlik ve farklılıkları gösterecek şekilde özetlenmiştir (Çizelge 4.5.-4.6.-4.7.).

Çizelge 4.5.: Phrap ve CAP3 programlarıyla elde edilen contiglerin karşılaştırılmasından örnekler. İki program için ortak olan fonksiyonlar.

	CAP3 ve Phrap İçin Yaprak EST'lerinden oluşturulan ortak contigler	CAP3 ve Phrap İçin Meyve EST'lerinden oluşturulan contigler	CAP3 ve Phrap İçin Tüm EST'lerinden oluşturulan contigler
Ortak BLASTN sonuçları	<u>CAP3 yaprak contigleri ile Phrap yaprak contigleri için bulunan ortak BLASTN sonuçları:</u> pyruvate	<u>CAP3 meyve contigleri ile Phrap meyve contigleri için bulunan ortak BLASTN sonuçları:</u> ribosomal protein, reductase, UDP-glucuronate decarboxylase 3, tubulin, ribosomal protein, aspartic protease,	<u>CAP3 toplam contigleri ile Phrap toplam contigleri için bulunan ortak BLASTN sonuçları:</u> HTC in leaf, ribulose-

	dehydrogenase E1 beta subunit, chlorophyll a/b-binding protein, ubiquitin, Reductase, ribosomal protein, ATPase, Histone, Dehydrogenase, Acyltransferase, Protease, Hydroxylase, chloroplast product, aldolase-like protein, heat shock protein, Oxygenase, Desaturase, elongation factor,	cytochrome P450 mRNA, glutinosa ascorbate peroxidase, beta-glucosidase, polyubiquitin, myb-related transcription factor, succinate dehydrogenase 2-1, ubiquitin, chlorophyll a/b binding protein,___amino alcohol phospho transferase, ADP-glucose pyrophosphorylase, ribulose-1,5-bisphosphate, diphosphate synthase, Glycosyltransferase, ATP binding / aminoacyl-tRNA ligase, AML1, ACC oxidase homologue protein, formamide amidohydrolase, calreticulin	1,5-bisphosphate, chlorophyll a/b binding protein, cysteine protease, HTC in fruit, chloroplast protein, ribosomal protein, Polyubiquitin, malate dehydrogenase, elongation factor, phosphate synthase, glyceraldehyde-3-phosphate dehydrogenase, Proteasome, Acyltransferase, histone, chloroplast product
--	--	--	---

Çizelge 4.6.: Phrap ve CAP3 programlarıyla elde edilen contiglerin karşılaştırılmasından örnekler. İki program arasındaki farklı fonksiyonlar.

	Yaprak EST'lerinden oluşturulan contigler (CAP3 ile)	Meyve EST'lerinden oluşturulan contigler (CAP3 ile)	Tüm EST'lerinden oluşturulan contigler (CAP3 ile)	Yaprak EST'lerinden oluşturulan contigler (Phrap ile)	Meyve EST'lerinden oluşturulan contigler (Phrap ile)	Tüm EST'lerinden oluşturulan contigler (Phrap ile)
Farklı BLASTN sonuçları	<u>CAP3 yaprak contiglerinin Phrap yaprak contiglerinden farklı BLASTN sonuçları:</u> serologically defined breast cancer antigen NY-BR-84, mannitol dehydrogenase 1, acidic ribosomal protein P1a, peroxisomal membrane protein-related (AT1G52870)	<u>CAP3 meyve contiglerinin Phrap meyve contiglerinden farklı BLASTN sonuçları:</u> Ca2+ antiporter/cation exchanger, GDP-mannose-3',5'-epimerase (GME), S-RNase-binding protein (SBP1), chloroplast oxygen-evolving protein, elongation factor 1 alpha subunit, Tic20-like protein, polygalacturonase precursor (PG7), metallophosphatase/diphosphonucle	<u>CAP3 toplam contiglerinin Phrap toplam contiglerinden farklı BLASTN sonuçları:</u> profilin, dehalogenase-like hydrolase family protein, molecular chaperone, (CCD1),	<u>Phrap yaprak contiglerinin CAP3 yaprak contiglerinden farklı BLASTN sonuçları:</u> HTC in fruit, metallothionein-like protein, phosphate synthase, GTP-binding protein, putative DNA-binding protein, Olea europaeaf fruit stearyl-ACP desaturase	<u>Phrap meyve contiglerinin CAP3 meyve contiglerinden farklı BLASTN sonuçları:</u> CAX, RFLP probe ZVG28, catalase, glyceraldehyde-3-phosphate dehydrogenase	<u>Phrap toplam contiglerinin CAP3 toplam contiglerinden farklı BLASTN sonuçları:</u> cytosolic factor, vernalization independence 3-like protein, RFLP probe, CAX, Ser/Thr protein kinase, AT-hook protein, beta-tubulin

		otide phosphatase 1, Pectinesterase, aquaporin-like protein, anion exchanger family protein, thiosulfate sulfurtransferase (TST), serine/threonine protein kinase SAPK8-like protein, DNAJ protein, non-phosphorylating glyceraldehyde dehydrogenase				
--	--	--	--	--	--	--

Kullanıcı bu çalışmada olduğu gibi hem CAP3 hem de PHRAP'ı kullanabilir ve ortak dizilerini karşılaştırabilir. Phrap programı içerdiği cross-match uygulaması ile vektör dizilerden kurtulmasını sağlamıştır. Ancak birleştirme aşaması bittikten sonra Consed programı ile bu birleştirmeler izlendiğinde yer yer tek bir EST'nin sanki bir contig dizisi gibi ifade edildiği anlaşılmıştır. Örneğin 550 meyve contig dizisinden 400 tanesi birleşmemiş EST ve 150 tanesi ise birleşmiş EST. Yani output dosyalarında 550 contig dizisi olduğu gözüküyorken consed incelemesi sonucu gerçekte 150 contig olduğu görülmüştür. Phrap programının her farklı EST grubunda, benzer hataları, her defasında verdiği belirlenmiştir. Sonuç olarak Phrap programı, 2228 adet yaprak EST dizisi için 335 contig üretirken, CAP3 programı aynı yaprak EST'leri ile 205 contig üretmiştir. 1506 meyve EST'sinden Phrap ile 550 contig, CAP3 ile 69 contig oluşturulmuştur. Tüm EST'ler bir araya getirilince (3734 EST) ise, 924 contig Phrap programı ile, 299 contig CAP3 programı ile üretmiştir. CAP3 sonuçlarından da emin olabilmek için yine Consed programı ile çalışılmıştır. Phrap programında gözlemlenen hatalara rastlanmamıştır. Bu yüzden fonksiyonel analiz çalışmalarına CAP3 sonuçları ile devam edilmesi tercih edilmiştir. Fonksiyonel analizler BLAST analizlerinin yanı sıra CAP3 sonuçları için BLAST2GO ile yapılan Gen Ontoloji çalışması ile yapılmıştır.

Ancak Çizelge 4.5., 4.6. ve 4.7.'de görüldüğü gibi BLAST2GO programından bağımsız olarak yapılan BLASTN analizleri hem Phrap hem de CAP3 sonuçları için gerçekleştirilmiştir. Her iki programdan elde edilen contig dizilerinin BLASTN sonuçları genel olarak ortaktır. Ancak görülen farklılıklar önemli olmuştur. İki programdan elde edilen contig gruplarına ait BLASTN analizi sonuçları arasındaki en dikkat çekici farklar; Çizelge

4.6.'da CAP3 programı ile oluşturulmuş yaprak contiglerinde görülen fakat Phrap programı ile oluşturulan yaprak contiglerinde görülmeyen “serologically defined breast cancer antigen NY-BR-84” proteini ve yine Çizelge 4.6.'da görülen “Olea europaeafuit stearoyl-ACP desaturase” adındaki gen dizisinin sadece Phrap programından elde edilen yaprak contiglerinde görülmesidir.

Yapılan çalışmalar zeytinde bulunan oleik asit ve zeytin bitkisi yaprak dokusunda bulunan oleuropein maddelerinin başta meme kanseri olmak üzere pek çok kanser türüne karşı direnç sağladığını göstermektedir (Menendez ve ark., 2007; Han ve ark., 2009). Bu çalışmada “serologically defined breast cancer antigen 84” dizisiyle benzerlik gösteren bir contig dizisinin varlığı da bu çalışmalar açısından önemlidir. Ayrıca Stearoyl-ACP desaturase (SAD), bir plastid enzimidir ve 9. ve 10. karbon arasında bir bağ oluşturarak stearoyl-ACP nin oleoyl-ACP ye doymamasını katalize eden ve sağlıklı bir beslenme diyetinde yer alan zeytinyağ kalitesini etkileyen önemli bir enzimidir. Cap3 de görülmemiş olsada bu iki önemli gen tekrar BLAST youluyla kontrol edilerek kütüphanede yer aldığı belirlenmiştir.

Çizelge 4.7.: Phrap ve CAP3 programlarıyla elde edilen contiglerin farklı dokular bakımından karşılaştırılmasından örnekler.

	Yaprak EST'lerinden oluşturulan contigler (CAP3 ile)	Meyve EST'lerinden oluşturulan contigler (CAP3 ile)	Tüm EST'lerinden oluşturulan contigler (CAP3 ile)	Yaprak EST'lerinden oluşturulan contigler (Phrap ile)	Meyve EST'lerinden oluşturulan contigler (Phrap ile)	Tüm EST'lerinden oluşturulan contigler (Phrap ile)
Ortak BLASTN sonuçları	<u>CAP3 yaprak contigleri ile CAP3 meyve ve CAP3 toplam contigleri için bulunan ortak BLASTN sonuçları:</u> reductase, cytochrome P450, GDP-mannose-3',5'-epimerase (GME), tubulin, cytochrome P450, ascorbate peroxidase, beta-glucosidase, polyubiquitin, aldolase-like	<u>CAP3 meyve contigleri ile CAP3 yaprak ve CAP3 toplam contigleri için bulunan ortak BLASTN sonuçları:</u> reductase, cytochrome P450, GDP-mannose-3',5'-epimerase (GME), tubulin, cytochrome P450, ascorbate peroxidase, beta-glucosidase, polyubiquitin, aldolase-like	<u>CAP3 toplam contigleri ile CAP3 meyve ve CAP3 yaprak contigleri için bulunan ortak BLASTN sonuçları:</u> reductase, cytochrome P450, GDP-mannose-3',5'-epimerase (GME), tubulin, cytochrome P450, ascorbate peroxidase, beta-	<u>Phrap yaprak contigleri ile Phrap meyve ve Phrap toplam contigleri için bulunan ortak BLASTN sonuçları:</u> HTC in fruit, UBIQUITIN, Reductase, ribosomal protein, ATPase, Histone, Dehydrogenase, Acyltransferase, Protease, phosphate synthase, Oxygenase,	<u>Phrap meyve contigleri ile Phrap yaprak ve Phrap toplam contigleri için bulunan ortak BLASTN sonuçları:</u> beta-tubulin, HTC in fruit, ribosomal protein	<u>Phrap toplam contigleri ile Phrap meyve ve Phrap yaprak contigleri için bulunan ortak BLASTN sonuçları:</u> beta-tubulin, HTC in fruit, ribosomal protein, elongation factor, phosphate synthase, Acyltransferase, histone protein

	protein, ubiquitin, chlorophyll a/b binding protein	protein, ubiquitin, chlorophyll a/b binding protein	glucosidase, polyubiquitin, aldolase-like protein, ubiquitin, chlorophyll a/b binding protein	elongation factor		
Farklı BLASTN sonuçları	<u>CAP3 yaprak contiglerinin diğer CAP3 contiglerinden farklı BLASTN sonuçları:</u> 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	<u>CAP3 meyve contiglerinin diğer CAP3 contiglerinden farklı BLASTN sonuçları:</u> SDH2-1 GDP-mannose-3',5'-epimerase (GME), 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, Ca ²⁺ antiporter/cation exchanger	<u>CAP3 toplam contiglerinin diğer CAP3 contiglerinden farklı BLASTN sonuçları:</u> white-brown-complex ABC transporter family, pentose-5-phosphate-3-epimerase, Cu/Zn superoxide dismutase, protoporphyrin IX, GDSL-motif lipase-like (LIP)	<u>Phrap yaprak contiglerinin diğer Phrap contiglerinden farklı BLASTN sonuçları:</u> putative DNA-binding protein, chlorophyll a/b-binding protein, Hydroxylase, GTP-binding protein, chloroplast product, aldolase-like protein, heat shock protein, metallothionein-like protein	<u>Phrap meyve contiglerinin diğer Phrap contiglerinden farklı BLASTN sonuçları:</u> CAX	<u>Phrap toplam contiglerinin diğer Phrap contiglerinden farklı BLASTN sonuçları:</u> cytosolic factor, chloroplast protein, malate dehydrogenase, glyceraldehyde-3-phosphate dehydrogenase, Ca ²⁺ /H ⁺ exchanger

Çizelge 4.7.'de bir contig veri grubunun kendini üreten aynı programla üretilmiş diğer contig veri gruplarıyla ortak ve farklı yanları ele alınmıştır. Ortak fonksiyonlar belirlenirken bir contig dizi grubunda en fazla görülen protein türüne diğer iki grupta rastlanıp rastlanmadığına bakılmıştır. Eğer rastlanan fonksiyon varsa sadece incelenen contig grubunun Çizelge 4.7.'deki ortak kısmına bu fonksiyon yazılmıştır. Farklılıklar ise bu yaklaşımın tam tersi ile

belirlenmiştir. Bu çizelgeden çıkarılan sonuca göre, dokular arasında ortak fonksiyonlara az rastlanmaktadır. Meyve contiglerini yaprak ve toplam contiglerden ayıran farklı fonksiyonların CAP3 sonuçlarında oldukça fazla olması ancak Phrap sonuçlarında neredeyse hiç görülmemesi dikkat çekicidir.

4.4.1.2 Gen Ontoloji Sonuçlarının İncelenmesi

EST verisi doğru ve sağlam polipeptidlerin çevrimi tarafından protein merkezli açıklamalar ile ilişkilendirilebilir. Çünkü polipeptidler protein çalışmalarında lokalizasyonun ve gen ontolojilerinin belirlenmesi için domainleri ve motifleri tanımlamada daha uygun kaynaklardır (GO) (Nagaraj ve ark., 2006).

Gen Ontoloji birliği 1998 yılında, ontolojiyi tanımlamak için aynı dili paylaşma ile elde edilen verileri birleştirmeye karar veren bir grup bilim adamı tarafından kurulmuştur (Lewis, 2004)..

Model organizma veritabanları için çalışan biyologların en temel soruları genler etrafında dönmektedir. Genlerin ne olduğu, mRNA'ları ve peptid bağlarının ne olduğu, genomların neresinde yer aldıkları, ne zaman açıklandıkları ve nasıl aktivitelerinin kontrol edildiği, hangi doku, organ ve hücre kısmında açıklandıkları, organizmanın biyolojisinde hangi rolü oynadığı gibi sorular sorulmaktadır. Bu soruların cevaplanması tüm genom çalışmalarının ortak gereksinimidir (Lewis, 2004).

İlk fonksiyonel sınıflandırma sistemlerinden biri 1993 yılında Monica Riley tarafından *Escherichia coli* için yaratılmıştır (Riley, 1993). Bu sistemler üstüne öncelikli olarak inşa etmeye bakıldığında, Michael Ashburner, Gen Ontolojisinin öncüsü haline gelen birleşmeyi başlatmış, ama aslında Sirke sineği'nin (FlyBase) ihtiyaçları için hizmet etmiştir. Benzer olarak, TIGR bu zamanlarda kendi fonksiyonel sınıflandırma sistemini oluşturmuştur. Bu erken çabalar sistematikti, tanımlamalar için iyi belirlenmiş bir dizi konsept kullanıyorlardı, ancak aynı organizmalar arasında paylaşılmadıkları için kısıtlı kalmışlardır. SGD (www.yeastgenome.org), FlyBase (www.flybase.org), TIGR (www.tigr.org), Mouse Genome Informatics (MGI) (www.informatics.jax.org) ve diğerleri bağımsızca farkına varmışlardır ki eğer bir ters ürün fonksiyonel sınıflandırma sistemi yaratılmışsa, veri entegrasyonunun önemli bir kısmı çözülebilir (Lewis, 2004).

1998'de, topluluk model organizma veritabanı için, belli bir sayıda tamamen dizilmiş genomlar ve geniş ölçekli fonksiyonel deneyler temel zorunluluk haline geldi. O yılın Temmuz ayında, Micheal Ashburner basitçe hiyerarşik olarak kontrol edilmiş "Moleküler

Biyoloji için Akıllı Sistemler (Intelligent Systems for Molecular Biology – ISMB)” adında bir sözlüğün planını Montreal uluslararası konferansında biyo-ontoloji seminerinde sunmuştur; ancak amacı başlangıçta pek ilgi görmemiştir. Ama daha sonraları, FlyBase, SGD ve MGI temsilcileri planı benimsemiştir ve kendi veritabanlarında her gen için moleküler fonksiyonlar ve biyolojik rolleri tanımlamak amacıyla aynı sözlüğe başvurmaya karar vermişlerdir. Böylece Gen Ontoloji (GO) Birliği kurulmuştur (Lewis, 2004)..

GO şimdi tamamen tanımlanmakta ve biyolojik sistemleri tanımlamak için yapısal, standartlaştırılmış sözlükleri oluşturmada başarı amaçlayan sayısız diğer biyolojik ontoloji projeleri için bir model haline gelmiştir. GO tanımlanmış terimleri içeren yapısal bir ağ ve gen ürünlerinin üç özelliği olan Moleküler Fonksiyonu, Biyolojik Süreci, Hücre Bileşenini ve aralarındaki ilişkiyi tanımlayan bir yapıdır (www.geneontology.org/GO.doc.html). Büyük kurumsal veritabanlarının arasında, devamlılığını sağlayan peptid dizilerini notlandırmak için Swiss-Prot, GO’yu kullanmaktadır (Lewis, 2004).

GO'daki her giriş GO:nnnnnnn formunda numaralı birer tanımlama ile, ve örneğin ; hücre (cell), fibroblast büyüme faktörü reseptörünü bağlayıcı (fibroblast growth factor receptor binding)veya sinyal iletimi (transduction) gibi bir term adı alır. Her term aynı zamanda üç ontoloji bakımından moleküler fonksiyonu, hücresel bileşen ve biyolojik süreci şeklinde değerlendirilir. Terimlerin çoğu bir metin tanımı ve tanımın kaynağını belirten referanslara sahiptir. Eğer terim kullanımı ile ilgili bilgi veya herhangi bir açıklama gerekli ise, bunlar ayrı bir alanda yorumlanmaktadır [12].

Diğer veritabanları ile GO veritabanı örtüşmektedir. Bir GO terminin başka bir veritabanı parçası ile anlamsal olarak eşleştiği durumlarda, karşı veritabanındaki referans terime eklenir. Bu karşıdan eklenen referanslar GO sayfası için haritalamadan indirilebilir. GO terimleri için eş anlamlı GO terimleri ve obsolete (kullanılmayan) GO terimleri gibi ifadelerle karşılaştırılabilir. Eş anlamlı GO terimi, tamamen eş anlamlı olan veya birbirlerine benzer bilgiler içeren GO terimlerini ifade eder. Obsolete (kullanılmayan) GO terimleri ise GO kapsamı dışında olan fakat yanlışlıkla adlandırılmış, bununla birlikte kendisi ve tanımı hala GO veritabanında bulunan, ancak kullanılmayan (obsolete) şeklinde işaretlenmiş terimlerdir. Genellikle, bu durumun sebebi belirtilerek yedek bir terim de önerilmektedir[12].

GO 'nun üç temel düzenleme ilkesi bulunmaktadır. Bunlar; hücresel bileşen, biyolojik süreç ve moleküler fonksiyondur. Herhangi bir gen ürünü işlevi bakımından bu üç temel grubun her hangi birine girebileceği gibi, birden fazla gruba da dahil olabilir. GO gruplarından hücresel bileşen adından da anlaşılacağı gibi hücrenin büyük ölçekli parçalarını ifade eder. Bu büyük

parça bir anatomik yapı (ör: granüllü endoplazmik retikulum veya nükleus) ya da bir gen ürün grubu olabilir. Biyolojik süreç ise, bir veya daha fazla moleküler işlevin birleşmesi ile düzenlenmesi sonucu tamamlanmış seri olaylardır. Geniş biyolojik süreç terimlerine örnek olarak hücresel fizyolojik işlevler veya sinyal iletimi verilebilir. Daha spesifik örnekler olarak ise; primidin metabolik süreç veya alfa-glukozid transferini verilebilir. Son GO terimi Moleküler fonksiyon ise, moleküler aşamada ortaya çıkan katalitik veya bağlayıcı aktiviteler gibi, aktiviteleri tanımlar. Bir biyolojik süreç ve bir moleküler fonksiyon arasındaki ayrımı yapmak zor olabilir, bu durumda genel kural, bir sürecin birden fazla farklı adımları olması gerekliliğini bilmektir. Bir biyolojik süreç bir sentez yoluna (pathway) eşdeğer değildir.

Bu çalışmada olduğu gibi BLAST2GO programı ile Gen ontoloji analizlerinin yapılması kütüphanenin incelenmesini daha da kolaylaştırmıştır. EST projelerinin çeşitli gruplandırmalar ve grafiklerle daha kolay ve anlaşılabilir ifade edilmesi araştırmacı açısından son derece önemlidir. BLAST2GO programının BLASTN ve BLASTX analizlerini daha kısa sürede, daha az insan gücüyle yapabilmesi çok büyük bir kolaylıktır. Kütüphane çalışmalarında BLAST analizleri yapılacak çok fazla dizi bulunması ve her bir dizi için tek tek çalışılması çok uzun zaman aldığı gibi bu sonuçlardan oluşturulan çizelgeler okunup değerlendirilemeyecek kadar uzun metinler oluşturmaktadır. BLAST2GO programı Blast, Mapping, Annotation ve Analysis aşamaları ile fonksiyonel analizin tamamen otomatik olmasını sağlamaktadır. Böylece dizilerle ilgili daha güvenilir, daha az yorucu ve daha hızlı şekilde değerlendirme yapılması sağlanır (Şekil 3.8., Şekil 3.9., Şekil3.10., Şekil 3.14., Şekil 3.15., Şekil 3.16., Şekil 3.19., Şekil 3.20., Şekil 3.21.). Sonuçlar bölümünde görülen Gen Ontoloji sonuçlarına ait grafiklerin kısaca özeti Çizelge 4.8.'de görülmektedir.

Çizelge 4.8.: Gen Ontoloji sonuçları.

Doku Tipi	GO Terimi	Contig Sayısı
Yaprak 205 contig	Moleküler Fonksiyon = 20 tip	162 contig
	Hücresel Bileşen = 11 tip	98 contig
	Biyolojik Proses = 20 tip	194 contig
Meyve 69 contig	Moleküler Fonksiyon = 8 tip	55 contig
	Hücresel Bileşen = 26 tip	51 contig
	Biyolojik Proses = 11 tip	89 contig
Toplam 299 contig	Moleküler Fonksiyon = 25 tip	207 contig
	Hücresel Bileşen = 14 tip	151 contig
	Biyolojik Proses = 30 tip	240 contig

4.4.1.3 Housekeeping İncelemesi

Elde edilen Blast sonuçlarına bakarak farklı anlatım yapan genleri bulmak için fonksiyonel analiz sonuçlarının değerlendirilmesi yapılmalıdır. Var olan bilgisayar programlarıyla bu konu hakkında kesin bilgiler sunulamamaktadır. Bir organizmanın housekeeping genlerini ve farklı anlatım yapan genlerini belirlemek için çok farklı koşullar ve çok farklı dokular göz önüne alınarak kütüphaneler kurulmalı ve elde edilen dizilerin fonksiyonlarından emin olunmalıdır.

Bir kompleks olarak transkriptom hiçbir zaman in-vivo ortamda tamamlanmış olarak görülmemiştir. Bunun sebebi, aynı dokuda, aynı zamanda tüm genlerin anlatımının yapılmamasıdır. Hücreler anlatımı her zaman yapılan, temel fonksiyonları olan “housekeeping” genlerin transkripsiyonunu yapar, fakat diğer “luxury” genlerin ekspresyonu belirli koşullara göre olur, örneğin; gelişim aşamasının belli bir parçasında veya dışardan gelen uyarılara göre düzenlenir. (Primrose ve Twyman, 2006)

Tipik bir insan hücresinde yaklaşık 15.000 – 20.000 arasında farklı mRNA sentezlenir. Bunların bazıları housekeeping fonksiyonlarındadır, bir kısmı ise daha özeldir. Aynı primer transkriptin farklı şekilde kesilmesi ile bu oran belirlenir. Bazı mRNA’lar çok fazla olabileceği gibi, bazıları orta miktarda ve diğerleri ise çok az olacaktır. (Primrose ve Twyman, 2006)

Hücrede RNA sentezine gerçek bir küresel bakış açısı için, tüm bu transkriptler aynı anda sayılabilir olmalıdır. Bu da hem duyarlı hem de seçici olan yüksek bir paralel analiz biçimi gerektirir. (Primrose ve Twyman, 2006)

İlk cDNA kütüphanesi çalışmaları Okubo ve arkadaşlarına aitti. (Okubo ve ark., 1992) Normalize edilmemiş bir cDNA kütüphanesi kullanılan organizmaya ait mRNA’ları sunmaktadır. Bazı mRNA’lar (ve onları karşılayan cDNA’lar) çok fazla bulunurken, bazıları çok az görülmektedir. Eğer kütüphaneden rasgele 5000 klon seçilir ve kısmi diziler belirlenirse, fazla bulunan transkriptler az bulunan transkriptlerden daha sıklıkla gözlemlenecektir. Bu sonuçların istatistiksel analizi birbiriyle ilişkili ekspresyon aşamalarının belirlenmesinde, farklı kaynaklardan oluşturulmuş kütüphanelerin karşılaştırılmasında kullanılarak farklı anlatım yapan genlerin tanımlanabilmesine izin verir. Bu yöntemin sınırları; cDNA kütüphaneleri üreten ve istatistiksel olarak önemli veriler için gerekli büyük ölçekli dizileme projeleri yürütülmesine kadar gider (Audic ve Claverie, 1997). Bununla

birlikte, EST veri miktarı şimdi bu verilerin ne sıklıkta gözleendiğini sorgulamak için uygun bir yaklaşım olan gen anlatımının dijital sunumu uygulanır. Bu yaklaşım çok fazla duyarlı olmamakla ve EST verilerinin ulaşılabilirliğine bağılı olmakla birlikte, silikodaki dizilerin doğrudan örneklerinin temeli veriler elde etmek amacıyla kullanılabilir. Konuyla ilgili olarak; EST dizi örneklerinin kullanılarak farklı anlatım yapan genlerin tanımlanması ile ilgili birkaç rapor (yayın) bulunmaktadır. (Claudio ve ark., 1998, Vasmatzis ve ark., 1998) (Primrose ve Twyman, 2006).

Biz bu çalışmada iki ayrı kütüphane oluşturduk. Zeytine ait iki farklı dokudan yaprak ve zeytin meyvesinden elde edilen iki kütüphanenin sonuçları birlikte analiz edildi. Fonskiyonel analizler için çizelgelerden de gözlenebileceği gibi yaprak ve meyve dokusu için farklı genleri yakalama şansı bulunmuş oldu. Zeytin meyvesinde gözlenmeyen “asparagine synthetase”, “desacetoxyvind oline”, “glutamate synthase1”, “chloroplast ferredoxin I”, “glyceraldehyde-3-phosphate dehydrogenase”, “chlorophyll a/b-binding protein”, “malate dehydrogenase”, “alcohol dehydrogenase”, “mannitol dehydrogenase1” ve “serologically defined breast cancer antigen NY-BR-84” gibi fonksiyonlar yaprakta gözleendi. Analiz edilen klon sayısının farklı olması yanı sıra iki farklı dokudan kütüphane oluşturulmasında bize bu sonucu gösterdi.

Zeytin için muhtemel housekeeping gen dizileri belirlenmediği için biz sonuçlarımızı Gemlik zeytininin Blast sonuçlarına göre en fazla örtüştüğü organizma olan Populus (kavak) bitkisi için belirlenen housekeeping genler ile karşılaştırdık (Sterky ve ark., 2004) ve bu genlerden 6 tanesi yapraktaki 6 contigde yer alırken 2 gende meyvedeki iki farklı contigde gözleendi. Çizelge 4.9. Populus housekeeping genleriyle örtüşen yaprak, meyve ve toplam contiglerini göstermektedir.

Çizelge 4.9.: Populus housekeeping genleri ile ortak fonksiyona sahip contig dizileri

POPULUS HOUSEKEEPİNG GENLERİ	AİT	TOPLAM	YAPRAK	MEYVE
Polyubiquitin		Contig 157	Contig 148	Contig 20
Hydroxymethyltransferase		Contig 127	Contig 120	
Elongation factor 1		Contig 62, Contig 234	Contig 110	Contig 21
60S ribosomal protein		Contig 102	Contig 93	
Asparagine synthetase		Contig 107	Contig 98	
Histone H3		Contig 279	Contig 94	

Populus housekeeping genleri içinden meyve yaprakta ortak gözlenen sadece “Polyubiquitin” ve “Elongation factor 1” için housekeeping gen olarak yaklaşıması muhtemeldir. Ancak daha önce de değinildiği gibi elimizdeki EST sayısının az ve sadece iki dokuya özgü kütüphaneden elde edilen veriler ile bir organizmanın housekeeping genleri hakkında kesin sonuca varmak mümkün olmaz. Ancak daha sonra yapılacak çalışmalar için bir kaynak oluşturacaktır bu sonuçlar.

Gemlik zeytin bitkisi için oluşturduğumuz bu çalışma bugüne kadar zeytin genomuyla ilgili hazırlanmış en kapsamlı EST çalışmasıdır. Zeytin genomu hakkında elde edilen bu bilgiler birçok araştırmacının kullanımına sunulmak üzere NCBI GenBankasına yüklenmiş olması nedeniyle özellikle zeytin genomunu çalışan ve diğer genom çalışması yapan birçok araştırmacıya karşılaştırmalı genomik çalışmaları için kaynak oluşturacaktır. Aynı zamanda EST analizleri yapmak isteyen araştırmacılar için kullanılan yöntemler ve elde edilen sonuçlar bakımından benzer çalışmalara ışık tutabilecek niteliktedir. Özellikle genom çalışmalarında EST dizilerinin analizleri için kullanılacak bilgisayar programlarının seçimi konusunda Phrap ve CAP3 programları karşılaştırılarak kullanıcılar için avantaj ve dezavantajları bu çalışma ile Zeytin EST leri kullanılarak ortaya çıkarılmıştır. Elde edilen bu EST koleksiyonunda belirlenen önemli genlerin fonksiyon çalışmalarının yapılması ve yeni genlerin varlığının tespit edilmesi için bu sonuçların devamında deneysel ve analiz çalışmaları yapılmalıdır.

KAYNAKLAR

- Aaronson JS, Eckman B, Blevins RA, ve ark. (1996) "Toward the development of a gene index to the human genome: an assessment of the nature of high-throughput EST sequence data." *GenomeRes* 6:829–45.
- Adams, M.D. ,Kelley, J.M. ve Gocayne, J.D. , (1991), "Complementary DNA sequencing: expressed sequence tags and human genome Project", *Science*, 252:1651-1656.
- Adams, M.D., Dubinick, M. Ve Kerlavage, A.R., (1992), "Sequence identification of 2375 human brain genes", *Nature*, 355: 632 – 634.
- Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., Kerlavage, A.R., McCombie, W.R. ve Venter, J.C., (1991), "Complementary DNA sequencing: expressed sequence tags and human genome Project", *Science*, 252(5013):1651-1656.
- Altschul SF, Madden TL, Schaffer AA, ve ark. (1997) "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." *Nucleic Acids Res*;25:3389–402.
- Angiolillo, A., Mencuccini, M., Baldoni, L., (1999), "Olive genetic diversity assessed using amplified fragment length polymorphisms.", *Theor. Appl. Genet.* 98 411-421.
- Audic, S. ve Claverie J. (1997) "The significance of digital gene expression profiles." *Genome Res.* 7,986-995.
- Baldoni, L., Tosti, N., Ricciolini, C., Belaj, A., Arcioni, S., Pannelli, G., ve ark., (2006), "Genetic structure of wild and cultivated olives in the Central Mediterranean.", *Basin. Ann. Bot.*, 98: 935–942.
- Banfi, S. ,Borsani, G. ,Rossi, E. ,Bernard, L. ve Guffanti, A. , (1996), "Identification and mapping of human cDNAs homologous to *Drosophila* mutant genes through EST database searching", *Nature Genetics* 13:167 – 174.
- Bedell JA, Korf I, Gish W. (2000) "MaskerAid: a performance enhancement to RepeatMasker." *Bioinformatics*;16:1040–1.
- Belaj, A., Trujillo, I., Rossa, R., Rallo, L., Gimenez, M.J., (2001), "Polymorphism and discrimination capacity of randomly amplified polymorphic markers in an olive germplasm bank." *J. Am. Soc. Hortic. Sci.* 126 64-71.
- Bennetzen, J.L., (2002), "Mechanisms and rates of genome expansion and contraction in flowering plants.", *Genetica*, 115: 29-36.
- Benson, D.A. ,Karsch-Mizrachi, I. ,Lipman, D.J. ve Ostell, J., (2001), "GenBank", *Nucleic Acids Research*, 30:17-20
- Benson, D.A. ,Karsch-Mizrachi, I. ve Lipman, D.J. ,(2004), "GenBank: update." , *Nucleic Acids Research*, 1:32

Besnard, G., Berville, A., (2000), "Multiple origins of the Mediterranean olive deduced from mitochondrial DNA polymorphisms.", C. R. Acad. Sci. 323 173-181.

Besnard, G., Baradat, P., Breton, C., Khadari, B. ve Berville, A., (2001), "Olive domestication from structure of oleasters and cultivars using RAPDs and mitochondrial RFLP.", Genetics Selection Evolution, 33 (1): 251-268.

Besnard, G., Gren, P.S., Berville, A., (2002a), "The genus *Olea*: molecular approaches of its structure and relationships to other *Oleaceae*.", Acta Bot Gallica, 149: (in press).

Besnard, G., Khadari, B., Baradat, P., Berville, A., (2002b), "*Olea europaea* (Oleaceae) phylogeography based on chloroplast DNA polymorphism.", Theor Appl Genet, DOI 10.1007/s00122-001- 832-x.

Besnard, G., Khadari, B., Baradat, P., Berville, A., (2002c), "Combination of chloroplast and mitochondrial DNA polymorphisms to study cytoplasm genetic differentiation in the olive (*Olea europaea* L.) complex.", Theor Appl Genet, DOI 10.1007/s00122-002-868-6.

Boguski MS, Schuler GD. (1995) "ESTablishing a human transcript map." Nat Genet;10:369–71.

Boguski, M.S. , (1995), "The turning point in genome research", Trends in Biochemical Sciences, 20:295-332

Brett D, Hanke J, Lehmann G, ve ark. (2000) "EST comparison indicates 38% of human mRNAs contain possible alternative splice forms." FEBS Lett 474:83–6.

Buetow KH, Edmonson MN, Cassidy AB. (1999) "Reliable identification of large numbers of candidate SNPs from public EST data." Nat Genet 21:323–5.

Burke J, Davison D, Hide W. (1999) "d2_cluster: a validated method for clustering EST and full-length cDNA sequences." Genome Res 9:1135–42.

Caminici, P. ve Kvam, C., (1996), "High efficiency full length cDNA cloning by biotinylated CAP-trapper", Genomics, 37, 327-36

Champoux, J.J. , (1993), "Roles of ribonuclease H in reverse transcription Reverse Transcriptase", Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NewYork

Chen, Y. , Lin, C.C. , Wang, C.D. , Wu, H.B. ve Hwang, P. , (2007), "An optimized procedure greatly improves EST vector contamination removal", BMC Genomics, 8:416

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

Claudio, J.O. ve arkadaşları (1998) "Identification of sequence tagged transcripts differentially expressed within the human hematopoietic hierarchy." Genomics 50, 44-52

Coleclough, C. ve Erlitz, F.L. , (1985), “Use of primer-restriction-end adapters in a novel cDNA cloning strategy”, *Gene*, 34(2-3):305-314

Conesa A, Gotz S, Garcia-Gomez JM, ve ark. (2005) “Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.” *Bioinformatics* 21:3674–6.

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M. ve Robles, M., (2005), “Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.”, *Bioinformatics Applications note*, 21, 18:3674–3676

Cooke, N.E. , Coit, D. , Weiner, R.I. ,Baxter, J.D. ve Martial, J.A. , (1980) “Structure of Cloned DNA Complementary to Rat Prolactin Messenger RNA”, *Journal of Biological Chemistry*, 255:6502-6510

Dayhoff, M.O. ,Eck, R.V. ,Chang, M.A. ve Sochard, M.R. , (1965), “Atlas of Protein Sequence and Structure”, National Biomedical Research Foundation, Silver Spring MD. 1

De Caraffa, V.B., Maury, J., Gambotti, C., Breton, C., Berville, A., ve Giannettini, J., (2002), “Mitochondrial DNA variation and RAPD mark oleasters, olive and feral olive from Western and Eastern Mediterranean.”, *Theor. Appl. Genet.*, 104: 1209–1216.

Dieffenbach, C.W. ve Dvesler, G.S. , (1995), “PCR Primer: A Laboratory Manual”, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NewYork

Dong, Q., Kroiss, L., Oakley, F.D., Wang, B.B., Brendel, V., (2005), “Comparative EST analyses in plant systems.”, *Methods Enzymol*, 395: 400-418.

Durbin, R. ve Dear, S. , (1998), “Base Qualities Help Sequencing Software” *Genome Research*, 8: 161-162

Edery,I. , Chu, L.L. , Sonenberg, N. ve Pelletier, J. , (1995), “An Efficient Strategy To Isolate Full-Length cDNAs Based on an mRNA Cap Retention Procedure (CAPture)” *Molecular and Cellular Biology*, American Society for Microbiology, 15:3363-3371

Efstratiadis, A., Kafatos, F. C., Maxam, A. M. ve Maniatis, T., (1976), “ Enzymatic in vitro synthesis of globin genes.”, *Cell*, 7:279-288

Elbaum, R., Melamed-Bessudo, C., Boaretto, E., Galili, E., Lev-Yadun, S., Levy, A.A. ve Weiner, S., (2006), “Ancient olive DNA in pits: preservation, amplification and sequence analysis.”, *Journal of Archaeological Science*, 33, 77–88.

Ergun, M.E. (2003), “Sofralık Zeytin İşletmelerinin Yapısı Ve Sorunları” Gemlik Zeytin Paneli, 13 Eylül 2003, Gemlik, Bursa, Türkiye

Ewing B, Green P. (1998) “Base-calling of automated sequencer traces using phred. II. Error probabilities.” *GenomeRes*;8:186–94.

Ewing RM, Ben Kahla A, Poirot O, ve ark. (1999) “Large-scale statistical analyses of rice

ESTs reveal correlated patterns of gene expression.” *Genome Res* 9:950–9.

Ewing, B. ve Green, P. , (1988), “Base-Calling of Automated Sequencer Traces Using Phred.II. Error Probabilities”, *Genome Research*, 8: 186-194

Ewing, B., Hillier, L. ,Wendl, M.ve Green.,P. , (1998),”Base-calling of automated sequencer traces using Phred. I. Accuracy assessment”, *Genome Research*, 8: 175-185

Feliciello, I. ve Chinali, G., (1993), “A modified alkaline lysis method for the preparation of highly purified plasmid DNA from *Escherichia coli*.”, *Anal. Biochem.*, 212: 394–401.

Gerard, G.F. ve D’ Allesio, J.M. , (1993), “Reverse Transcriptase (EC2.7.7.49): The Use of Cloned Moloney Murine leukemia virus reverse transcriptase to synthesize DNA from RNA.”, *Methods Mol. Bio.* 16: 73-94.

Gerhold, D. ve Caskey, C.T. , (1996), “It's the genes! EST access to human genome content”, *Bioessays*, 18:973-81

Gingeras, T.R., Milazzo, J.P., Sciaky, D. ve Roberts, R.J., (1979), "Computer programs for the assembly of DNA sequences", *Nuc. Acids Res.* 7: 529-545.

Gordon, D., Abajian, C. ve Green, P., (1998), “Consed: A Graphical Tool for Sequence Finishing”, *Genome Res.*, 8: 195-202

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

Gubler, U. ve Hoffman, B.J. , (1983), “A simple and very efficient method for generating cDNA libraries.” *Gene*, 26, 263-269

Han, J., Talorete, T.P.N., Yamada, P. ve Isoda, H., (2009), “Anti-proliferative and apoptotic effects of oleuropein and hydroxytyrosol on human breast cancer MCF-7 cells.”, *Cytotechnology*, 59:45–53.

Hawkins V, Doll D, Bumgarner R, ve ark., (1999) “PEDB: the Prostate Expression Database.” *Nucleic Acids Res* 27:204–8.

Heyer LJ, Kruglyak S, Yooseph S. (1999) “Exploring expression data: identification and analysis of coexpressed genes.” *Genome Res* 9:1106–15.

<http://www.phrap.org/>. Phred, Phrap and Consed. Laboratory of Phil Green, Department of Genome Sciences, University of Washington (17 February 2006 date last accessed).

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

Hulbert, S.H., ve Michelmore, R.W., (1988), “DNA restriction fragment length polymorphism and somatic variation in the lettuce downy mildew fungus, *Bremia lactucae*.” *Mol. Plant–*

Microbe Interact., 1: 17–24.

Jiang J, Jacob HJ. (1998) “EbEST: an automated tool using expressed sequence tags to delineate gene structure.” *Genome Res* 8:268–75.

Jongeneel, C.V., (2000), “Searching the expressed sequence tag (EST) databases: panning for genes.”, *Brief Bioinform*, 1: 76-92.

Kan Z, Castle J, Johnson JM, Tsinoremas NF. (2004) “Detection of novel splice forms in human and mouse using cross-species approach.” *Pac Symp Biocomput* 42–53.

Kan Z, Rouchka EC, Gish WR, States DJ. (2001) “Gene structure prediction and alternative splicing analysis using genomically aligned ESTs.” *GenomeRes* 11:889–900.

Kawasaki, E.S. ,Innis, M.A. ,Gelfand, D.H. ve Sninski, J.J. , (1990), “Amplification of RNA in PCR Protocols”, *A Guide to Methods and Applications*, Academic Press, NewYork.

Kececioglu, J.D. ,(1991), "Exact and approximate algorithms for DNA sequence reconstruction", Ph.D. Thesis. Technical Report 91-26, Dept. of Computer Science, U. of Arizona, Tucson, AZ 85721.

Kececioglu, J.D. ve Myers, E.W., (1992), "Combinatorial algorithms for DNA sequence assembly", Accepted for publication in *Algorithmica*

Kim SI, Kim JY, Kim EA, ve ark. (2003) “Proteome analysis of hairy root from *Panax ginseng* C.A. Meyer using peptide fingerprinting, internal sequencing and expressed sequence tag data.” *Proteomics* 3:2379–92.

Kockar, F. ve Ilikci, R., (2003), “Investigations of genetic variation between olive (*Olea europaea* L.) cultivars using arbitrarily primed polymerase chain reaction (AP-PCR).”, *Journal of biosciences*, 58(11-12):837-42.

Kotewicz, M.L. , Sampson, C.M. , D’Alessio, J.M. ve Gerard, G.F. , (1988), “Isolation of cloned Moloney murine leukemia virus reverse transcriptase lacking ribonuclease H activity.”, *Nucleic Acids Research* , 16: 265-277

Land H., Grey M., Hanser H., Lindenmaier W. ve Schutz G., (1981), “5’ Terminal Sequences of eucaryotic mRNA can be cloned with a high efficiency.”, *Nucleic Acids Research*, 9: 2251-66

Lapeyre, B. Ve Amalric, F., (1985), “A powerful method for the preparation of cDNA libraries: isolation of cDNA encoding a 100-kDal nucleolar protein”, *Gene(Amsterdam)*, 37:1-31-3, 215-220

Lee Y, Tsai J, Sunkara S, ve ark. (2005) “The TIGR Gene Indices: clustering and assembling EST and known genes and integration with eukaryotic genomes.” *Nucleic Acids Res*;33:D71–4.

Lewis, S.E., (2004), “Gene Ontology: looking backwards and forwards.” *Genome Biology*, 6:103

Liang F, Holt I, Pertea G, ve ark. (2000) "An optimized protocol for analysis of EST sequences." *Nucleic Acids Res* 28:3657–65.

Liang, P. ve Pardee, A.B., (1992), "Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction", *Science*, 257:967-971

Lisacek FC, Traini MD, Sexton D, etal. (2001) "Strategy for protein isoform identification from expressed sequence tags and its application to peptide mass fingerprinting." *Proteomics*;1:186–93.

Madison, J.T. ,Everett, G.A. ve Kung, H. , (1966), "Nucleotide Sequence of a Yeast Tyrosine Transfer RNA" , *Science*, 153: 531 – 534

Maniatis, T., Sim, G. K., Efstratiadis, A. ve Kafatos, F. C., (1976), "Amplification and characterization of a beta-globin gene synthesized in vitro.", *Cell*, 8:163-182.

Marchler-Bauer A, Panchenko AR, Shoemaker BA, ve ark. (2002) "CDD: a database of conserved domain alignments with links to domain three-dimensional structure." *Nucleic Acids Res* 30:281–3.

Martienssen, R.A., (2000), "Weeding out the genes: The *Arabidopsis* genome project.", *Funct. Integr. Genomics*, 1:2–11.

Maruyama, K. , Sugano, S. , Suzuki, Y. ve Yoshitomo, K. , (1994), "Oligo-capping: a simple method to replace the cap structure of eukaryotic mRNAs with oligoribonucleotides." *Gene*, 138:171-174

Maxam, A.M. ve Gilbert, W., (1977), "A new method for sequencing DNA". *Proc. NatL. Acad. Sci. USA* 74:560-564.

Mekuria, G. T., Collins, G.G., Sedgley, M., (1999), "Genetic variability between different accessions of some common commercial olive cultivars.", *J. Hort. Sci. Biotech.* 74 309-314.

Menendez, J.A., Vazquez-Martin, A., Colomer, R., Brunet, J., Carrasco-Pancorbo, A., Garcia-Villalba, R., Fernandez-Gutierrez, A. ve Segura-Carretero, A., (2007), "Olive oil's bitter principle reverses acquired autoresistance to trastuzumab (Herceptin™) in HER2-overexpressing breast cancer cells.", *BMC Cancer*, 7:80

Miller RT, Christoffels AG, Gopalakrishnan C, ve ark. (1999) "A comprehensive approach to clustering of expressed human gene sequence: the sequence tag alignment and consensus knowledge base." *Genome Res*;9:1143–55.

Mironov AA, Fickett JW, Gelfand MS. (1999) "Frequent alternative splicing of human genes." *Genome Res* 9:1288–93.

Modrek B, Resch A, Grasso C, Lee C. (2001) "Genome-wide detection of alternative splicing in expressed sequences of human genes." *Nucleic Acids Res* 29:2850–9.

Mooney BP, Krishnan HB, Thelen JJ. (2004) "High-throughput peptide mass fingerprinting

of soybean seed proteins: automated workflow and utility of UniGene expressed sequence tag databases for protein identification.” *Phytochemistry* 65:1733–44.

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

Moyle, R., Fairbairn, D.J., Ripi, J., Crowe, M., Botella, J.R., (2005), “Developing pineapple fruit has a small transcriptome dominated by metallothionine.”, *J. Exp. Bot.*, 56: 101–112.

Myers, E.W. (1994). “Advances in sequence assembly.” Adams, M.D., Fields, C. and Venter, J.C., eds., *Automated DNA Sequencing and Analysis*, 231-238. San Diego, CA: Academic Press.

Nagaraj, S.H., Gasser, R.B., ve Ranganathan, S., (2006), “A hitchhiker’s guide to expressed sequence tag (EST) analysis.”, *Briefings In Bioinformatics*, 8:1:6-21

Nelson RT, Grant D, Shoemaker RC. (2005) “ESTminer: a suite of programs for gene and allele identification.” *Bioinformatics* 21:691–3.

Newcomb, R.D., Crowhurst, R.N., Gleave, A.P., Rikkerink, E.H.A., Allan, A.C., Beuning, L.L., Bowen, J.H., Gera, E., Jamieson, K.R., Janssen, B.J., Laing, W.A., McCartney, S., Nain, B., Ross, G.S., Snowden, K.C., Souleyre, E.J.F., Walton, E.F., ve Yauk Y.K., (2006), “Analyses of Expressed Sequence Tags from Apple.”, *Plant Physiol.*, 141 (1): 147.

Okayama, H. ve Berg, P. , (1982), “High-efficiency cloning of full-length cDNA.”, *Molecular and Cellular Biology*, 2(2): 161-170

Okubo, K. ,Hori, N. ,Matoba, R. ,Niiyama, T. ve Fukushima, A. , (1992), “Large scale cDNA sequencing for analysis of quantitative and qualitative aspects of gene expression”, *Nature Genetics* 2: 173 – 179

Okubo, K. ve Matsubara, K. , (1997), “Complementary DNA sequence (EST) collections and the expression information of the human genome”, *FEBS Letters*, 403: 225-229

Peltola, H., So"derlund, H. ve Ukkonen, E., (1984), "SEQAID: A DNA sequence assembly program based on a mathematical model", *Nucleic Acids Research* 12: 307-321.

Pertea G, Huang X, Liang F, ve ark. (2003) “TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets.” *Bioinformatics*;19:651–2.

Picoult-Newberg L, Ideker TE, Pohl MG, ve ark. (1999) “Mining SNPs from EST databases.” *Genome Res* 9:167–74.

Primrose, S.B. ve Twyman, R.M., (2006), “Principles of Gene Manipulation and Genomics”, Blackwell Publishing, Oxford

Ptitsyn A, Hide W. (2005) “CLU: a new algorithm for EST clustering.” *BMC Bioinformatics* 6 [Suppl 2]:S3.

- Qiu F, Guo L, Wen TJ, ve ark. (2003) "DNA sequence-based "bar codes" for tracking the origins of expressed sequence tags from a maize cDNA library constructed using multiple mRNA sources." *Plant Physiol* 133:475–81.
- Quackenbush J, Liang F, Holt I, ve ark. (2000) "The TIGR gene indices: reconstruction and representation of expressed gene sequences." *Nucleic Acids Res*;28:141–5.
- Rallo, P., Dorado, G., Martin A., (2000), "Development of simple sequence repeats (SSRs) in olive tree (*Olea europaea* L.).", *Theor. Apopl. Genet.* 101 984-989.
- Riley M. (1993) "Functions of the gene products of *Escherichia coli*." *Microbiol Rev*, 57:862-952.
- Rudd, S., (2003), "Expressed sequence tags: alternative or complement to whole genome sequences?", *Trends Plant Sci*, 8: 321-329.
- Ryle, A.P., Sanger, F., Smith, L.F. ve Kitai, R., (1955), "The disulphide bonds of insulin", *Biochemical Journal*, 60:541-556.
- Sambrook, J., Fritsch, E.F., ve Maniatis, T. (1989) "Molecular Cloning: A Laboratory Manual", 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Sanger, F., Coulson, A.R., Hong, G.F., Hill, D.F. ve Petersen, G.B., (1982), "Nucleotide inhibitors". *Proc. NatL. Acad. Sci. USA* 74:5463-5467.
- Sanger, F., Nicklen, S. ve Coulson, A.R., (1977), "DNA sequencing with chain-terminating sequence of bacteriophage λ DNA". *Journal of Molecular Biology* 162:729-773.
- Schena, M., Shalon, D., Davis, R.W., ve Brown, P.O., (1995), "Quantitative Monitoring of Gene Expression Patterns with a Complementary DNA Microarray", *Science*, 270: 467 - 470
- Schmitt AO, Specht T, Beckmann G, ve ark. (1999) "Exhaustive mining of EST libraries for genes differentially expressed in normal and tumour tissues." *Nucleic Acids Res* 27:4251–60.
- Sefc, K.M., Lopes, S., Mendonca, D., Dos Santos, M.R., Machado, M.L.D., ve Machado, A.D., (2000), "Identification of microsatellite loci in olive (*Olea europaea*) and their characterization in Italian and Iberian olive trees.", *Mol. Ecol.*, 9: 1171–1173.
- Smith T.F. ve Waterman M.S.,(1981), "Identification of common molecular Subsequences", *Journal of Molecular Biology*, 147(1):195-197.
- Staden, R. (1979), "A strategy of DNA sequencing employing computer programs", *Nuc. Acids Research*, 7: 2601-2610.
- Sterky, F., Bhalerao, R.R., Unneberg, P., Segerman, B., Nilsson, P., Brunner, A.M., Charbonnel-Campaa, L., Lindvall, J.J., Tandre, K., Strauss, S.H., Sundberg, B., Gustafsson, P., Uhlen, M., Bhalerao, R.P., Nilsson, O., Sandberg, G., Karlsson, J., Lundeberg, J. ve Stefan, J., (2004), "A *Populus* EST resource for plant functional genomics." *Proceedings of the National Academy of Sciences of the United States of America*, 101 no. 38 13951-13956

Suzuki, Y. ,Ishihara, D. ve Sasaki, M. , (2000), “Statistical Analysis of the 5' Untranslated Region of Human mRNA Using “Oligo-Capped” cDNA Libraries” . Genomics, 64: 286-97

Suzuki, Y. ,Yoshitomo-Nakagawa, K. ve Maruyama, K. , (1997), “Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library “. Gene, 200: 149-56

Tatusov RL, Fedorova ND, Jackson JD, ve ark. (2003) “The COG database: an updated version includes eukaryotes.” BMC Bioinformatics;4:41.

Technical Report WUCS-86-16, Dept. of Computer Science, Washington University, St.

Tunalıođlu, R. (2003), “İç ve Dış Pazar Boyutu ile Türkiye Sofralık Zeytin Pazarlaması; Son Durum, Gelişmeler, Beklentiler.” Gemlik Zeytin Paneli, 13 Eylül 2003, Gemlik, Bursa, Türkiye

Turner, J., (1986), "Approximation algorithms for the shortest common superstring problem", Louis, MO 63130.

Useche FJ, Gao G, Harafey M, Rafalski A. (2001) “Highthroughput identification, database storage and analysis of SNPs in EST sequences.” Genome InformSerWorkshop Genome Inform 12:194–203.

Van der Hoeven, R., Ronning, C., Giovannoni, J., Tanksley, M.G., (2002), “Deductions about the number, organization, and evolution of genes in the tomato genome based on analysis of a larger expressed sequence tag collection and selective genomic sequencing.”, Plant Cell, 14: 1441–1456.

Vasmatzis G. ve arkadaşları (1998) “Discovery of three genes specifically expressed in human prostate by expressed sequence tag database analysis.” Proc. Nat. Acad. Sci. USA 95, 300-304.

Velculescu, V. E. ve Kinzler, K. W., (1995), “Serial Analysis of Gene Expression”, Science, 270:484-487

Wan H, Li L, Federhen S, Wootton JC. (2003) “Discovering simple regions in biological sequences associated with scoring schemes.” JComput Biol 10:171–85.

Wang JP, Lindsay BG, Leebens-Mack J, etal. (2004) “EST clustering error evaluation and correction.” Bioinformatics 20:2973–84.

Wiesman, Z., Avidan, N., Lavee, S., Quebedeaux, B., (1998), “Molecular characterization of common olive varieties in Israel and the West Bank using randomly amplified polymorphic DNA (RAPD) markers.” J. Am. Soc. Hort. Sci. 123 837-841.

Wilcox, A.S. ,Khan, A.S. ,Hopkins, J.A. ve Sikela, J.M. , (1991), “Use of 3' untranslated sequences of human cDNAs for rapid chromosome assignment and conversion to STSs: implications for an expression map of the genome” , Nucleic Acids Research, 19:1837-1843

Wu TD, Watanabe CK. (2005) “GMAP: a genomic mapping and alignment program for

mRNA and EST sequences.” *Bioinformatics* 21:1859–75.

Yamamoto, K., ve Sasaki, T., (1997), “Large-scale EST sequencing in rice.”, *Plant Mol. Biol.*, 35:135–144.

Yu, J., ve ark., (2002), “A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*).”, *Science*, 296:79–92.

İNTERNET KAYNAKLARI

- [1] <http://www.ddbj.nig.ac.jp/>
- [2] <http://www.embl.de/>
- [3] <http://www.ncbi.nlm.nih.gov/About/primer/bioinformatics.html>
- [4] <http://www.ncbi.nlm.nih.gov/About/primer/est.html>
- [5] <http://www.ncbi.nlm.nih.gov/VecScreen/VecScreen.html>
- [6] <http://www.phrap.org/>
- [7] <http://www.zae.gov.tr/ekonomi/2.asp>
- [8] <http://www.yeastgenome.org/>
- [9] <http://www.flybase.org>
- [10] <http://www.tigr.org/>
- [11] <http://www.informatics.jax.org/>
- [12] <http://www.geneontology.org/GO.doc.html>
- [13] <http://www.ncbi.nlm.nih.gov/Sequin/>
- [14] <http://www.ensembl.org/index.html>
- [15] <http://genome.ucsc.edu/>
- [16] <http://www.genome.jp/kegg/>

EKLER

Çizelge Ek 1: İşlenmiş “*Olea europaea L. cv. Gemlik*” EST dizilerinin GenBank’a kaydedilmesinden sonra GenBank veritabanından alınan “accession numaraları”.

GenBank_Accn		
GO242703	OEAA-070810_Plate10a01.b1	64487766
GO242704	OEAA-070810_Plate10a02.b1	64487767
GO242705	OEAA-070810_Plate10a03.b1	64487768
GO242706	OEAA-070810_Plate10a04.b1	64487769
GO242707	OEAA-070810_Plate10a05.b1	64487770
GO242708	OEAA-070810_Plate10a06.b1	64487771
GO242709	OEAA-070810_Plate10a07.b1	64487772
GO242710	OEAA-070810_Plate10a08.b1	64487773
GO242711	OEAA-070810_Plate10a09.b1	64487774
GO242712	OEAA-070810_Plate10a10.b1	64487775
GO242713	OEAA-070810_Plate10a11.b1	64487776
GO242714	OEAA-070810_Plate10a12.b1	64487777
GO242715	OEAA-070810_Plate10a13.b1	64487778
GO242716	OEAA-070810_Plate10a14.b1	64487779
GO242717	OEAA-070810_Plate10a15.b1	64487780
GO242718	OEAA-070810_Plate10a16.b1	64487781
GO242719	OEAA-070810_Plate10a17.b1	64487782
GO242720	OEAA-070810_Plate10a18.b1	64487783
GO242721	OEAA-070810_Plate10a19.b1	64487784
GO242722	OEAA-070810_Plate10a20.b1	64487785
GO242723	OEAA-070810_Plate10a21.b1	64487786
GO242724	OEAA-070810_Plate10a22.b1	64487787
GO242725	OEAA-070810_Plate10a23.b1	64487788
GO242726	OEAA-070810_Plate10a24.b1	64487789
GO242727	OEAA-070810_Plate10b01.b1	64487790
GO242728	OEAA-070810_Plate10b02.b1	64487791
GO242729	OEAA-070810_Plate10b03.b1	64487792
GO242730	OEAA-070810_Plate10b04.b1	64487793
GO242731	OEAA-070810_Plate10b05.b1	64487794
GO242732	OEAA-070810_Plate10b06.b1	64487795
GO242733	OEAA-070810_Plate10b07.b1	64487796
GO242734	OEAA-070810_Plate10b08.b1	64487797
GO242735	OEAA-070810_Plate10b09.b1	64487798
GO242736	OEAA-070810_Plate10b10.b1	64487799
GO242737	OEAA-070810_Plate10b11.b1	64487800
GO242738	OEAA-070810_Plate10b12.b1	64487801
GO242739	OEAA-070810_Plate10b13.b1	64487802
GO242740	OEAA-070810_Plate10b14.b1	64487803
GO242741	OEAA-070810_Plate10b15.b1	64487804
GO242742	OEAA-070810_Plate10b16.b1	64487805
GO242743	OEAA-070810_Plate10b17.b1	64487806
GO242744	OEAA-070810_Plate10b18.b1	64487807
GO242745	OEAA-070810_Plate10b19.b1	64487808
GO242746	OEAA-070810_Plate10b20.b1	64487809

GO242703 ile başlayan ve GO246436 ile son bulan ve toplam 3734 adet işlenmiş “*Olea europaea L. cv. Gemlik*” EST dizisine ait olan accession numaraları çizelgesinin bir kısmı görülmektedir (Çizelge Ek 1).

Çizelge Ek 2: Yaprak dokusuna ait contig dizilerinden her birinin içerdiği yaprak EST'lerinin adı ve accession numaraları.

CONTİG no	EST no	EST adı		Accession no
1	2	OEAA-070810_Plate1b06.b1 070810_Plate6c10.b1	+	OEAA- GO243111 GO244996
2	2	OEAA-070810_Plate1b10.b1 070810_Plate5m13.b1	+	OEAA- GO243115 GO244860
3	2	OEAA-070810_Plate1b13.b1 070810_Plate4i19.b1	+	OEAA- GO243118 GO244408
4	2	OEAA-070810_Plate1c03.b1 070810_Plate6c14.b1	+	OEAA- GO243132 GO245000
5	2	OEAA-070810_Plate1c16.b1 070810_Plate1n20.b1	+	OEAA- GO243145 GO243402
6	3	OEAA-070810_Plate1d09.b1 070810_Plate1a03.b1 070810_Plate1e02.b1	+ +	OEAA- GO243086 OEAA- GO243175 GO243160
7	4	OEAA-070810_Plate1d17.b1 070810_Plate2c05.b1 070810_Plate3d20.b1 070810_Plate1o20.b1	+ + +	OEAA- GO243166 OEAA- GO243504 OEAA- GO243909 GO243426
8	2	OEAA-070810_Plate1d21.b1 070810_Plate1i03.b1	+	OEAA- GO243170 GO243270
9	2	OEAA-070810_Plate1d23.b1 070810_Plate5a24.b1	+	OEAA- GO243172 GO244596
10	2	OEAA-070810_Plate1e09.b1 070810_Plate1n05.b1	+	OEAA- GO243182 GO243387

11	3	OEAA-070810_Plate1e16.b1 + OEAA-070810_Plate5j08.b1 + OEAA-070810_Plate4p11.b1		OEAA-	GO243189 GO244785 GO244562
12	2	OEAA-070810_Plate1f04.b1 + OEAA-070810_Plate4p01.b1		OEAA-	GO243200 GO244552
13	2	OEAA-070810_Plate1f13.b1 + OEAA-070810_Plate4f21.b1		OEAA-	GO243209 GO244315
14	2	OEAA-070810_Plate1f14.b1 + OEAA-070810_Plate4m18.b1		OEAA-	GO243210 GO244499
15	3	OEAA-070810_Plate1f15.b1 + OEAA-070810_Plate2p19.b1 + OEAA-070810_Plate6m01.b1		OEAA- OEAA-	GO243211 GO243814 GO245221
16	7	OEAA-070810_Plate1f24.b1 + OEAA-070810_Plate6a03.b1 + OEAA-070810_Plate6l24.b1 + OEAA-070810_Plate1d02.b1 + OEAA-070810_Plate1a16.b1 + OEAA-070810_Plate2m15.b1 + OEAA-070810_Plate6l05.b1		OEAA- OEAA- OEAA- OEAA-	GO243220 GO244943 GO245220 GO243155 GO243098 GO243739 GO245202
17	3	OEAA-070810_Plate1h03.b1 + OEAA-070810_Plate5m21.b1 + OEAA-070810_Plate4a11.b1		OEAA- OEAA-	GO243247 GO244868 GO244210
18	4	OEAA-070810_Plate1h18.b1 + OEAA-070810_Plate1e24.b1 + OEAA-070810_Plate4i04.b1 + OEAA-070810_Plate3o17.b1		OEAA-	GO243262 GO243196 GO244393 GO244169
19	2	OEAA-070810_Plate1i18.b1 + OEAA-070810_Plate2b01.b1		OEAA-	GO243284 GO243478
20	2	OEAA-070810_Plate1i23.b1		OEAA-	GO243289

					GO244867
					GO244058
					GO245302
					GO243355
					GO245282
					GO243885
					GO243489
					GO244617
29	3	OEAA-070810_Plate1o05.b1 070810_Plate1o08.b1 070810_Plate5h01.b1	+ +	OEAA- OEAA-	GO243411 GO243414 GO244732
30	2	OEAA-070810_Plate1o07.b1 070810_Plate5o04.b1	+	OEAA-	GO243413 GO244898
31	2	OEAA-070810_Plate1o14.b1 070810_Plate2o24.b1	+	OEAA-	GO243420 GO243795
32	2	OEAA-070810_Plate1p02.b1 070810_Plate1n02.b1	+	OEAA-	GO243432 GO243384
33	2	OEAA-070810_Plate1p20.b1 070810_Plate3f23.b1	+	OEAA-	GO243449 GO243960
34	2	OEAA-070810_Plate1p24.b1 070810_Plate5p10.b1	+	OEAA-	GO243453 GO244927
35	3	OEAA-070810_Plate2a18.b1 070810_Plate6c06.b1 + OEAA-070810_Plate2f07.b1	+	OEAA-	GO243471 GO244992 GO243575
36	2	OEAA-070810_Plate2a19.b1 070810_Plate1n10.b1	+	OEAA-	GO243472 GO243392
37	2	OEAA-070810_Plate2b05.b1 070810_Plate4g20.b1	+	OEAA-	GO243482 GO244362
38	2	OEAA-070810_Plate2b20.b1	+	OEAA-	GO243497

		070810_Plate2e20.b1			GO243564
47	2	OEAA-070810_Plate2f23.b1 070810_Plate4i10.b1	+	OEAA-	GO243591 GO244399
48	2	OEAA-070810_Plate2g06.b1 070810_Plate5g20.b1	+	OEAA-	GO243598 GO244727
49	2	OEAA-070810_Plate2g11.b1 070810_Plate1i21.b1	+	OEAA-	GO243603 GO243287
50	2	OEAA-070810_Plate2g17.b1 070810_Plate3a24.b1	+	OEAA-	GO243608 GO243841
51	2	OEAA-070810_Plate2h07.b1 070810_Plate2g16.b1	+	OEAA-	GO243621 GO243607
52	4	OEAA-070810_Plate2h10.b1 070810_Plate6e03.b1 070810_Plate3n05.b1 070810_Plate3b19.b1	+ + +	OEAA- OEAA- OEAA-	GO243624 GO245036 GO244133 GO243860
53	3	OEAA-070810_Plate2h21.b1 070810_Plate2e22.b1 070810_Plate3n14.b1	+ +	OEAA- OEAA-	GO243635 GO243566 GO244142
54	2	OEAA-070810_Plate2h22.b1 070810_Plate4i01.b1	+	OEAA-	GO243636 GO244459
55	2	OEAA-070810_Plate2i06.b1 070810_Plate5j11.b1	+	OEAA-	GO243644 GO244788
56	2	OEAA-070810_Plate2i11.b1 070810_Plate2e12.b1	+	OEAA-	GO243649 GO243558
57	3	OEAA-070810_Plate2j09.b1 070810_Plate2d19.b1 070810_Plate1m11.b1	+ +	OEAA- OEAA-	GO243670 GO243541 GO243369

58	2	OEAA-070810_Plate2j18.b1 070810_Plate3m05.b1	+	OEAA-	GO243679 GO244109
59	2	OEAA-070810_Plate2k02.b1 070810_Plate2p15.b1	+	OEAA-	GO243687 GO243810
60	3	OEAA-070810_Plate2k03.b1 070810_Plate4b23.b1 + OEAA-070810_Plate3i13.b1	+	OEAA-	GO243688 GO244245 GO244021
61	2	OEAA-070810_Plate2k11.b1 070810_Plate6d08.b1	+	OEAA-	GO243695 GO244649
62	9	OEAA-070810_Plate2k21.b1 070810_Plate2m23.b1 070810_Plate1e22.b1 070810_Plate6a09.b1 070810_Plate3h13.b1 070810_Plate1b22.b1 070810_Plate5a18.b1 070810_Plate1g11.b1 070810_Plate5n16.b1	+ + + + + + + +	OEAA- OEAA- OEAA- OEAA- OEAA- OEAA- OEAA- OEAA-	GO243702 GO243747 GO243194 GO244948 GO243997 GO243127 GO244591 GO243231 GO244886
63	2	OEAA-070810_Plate2k22.b1 070810_Plate2o22.b1	+	OEAA-	GO243703 GO243793
64	3	OEAA-070810_Plate2l01.b1 070810_Plate4k04.b1 070810_Plate2d13.b1	+ +	OEAA- OEAA-	GO243706 GO244439 GO243535
65	5	OEAA-070810_Plate2l06.b1 070810_Plate1m03.b1 070810_Plate6m09.b1 070810_Plate1k10.b1 070810_Plate6d17.b1	+ + + +	OEAA- OEAA- OEAA- OEAA-	GO243711 GO243361 GO245229 GO243321 GO245026
66	2	OEAA-070810_Plate2l08.b1	+	OEAA-	GO243713

		070810_Plate2h06.b1			GO243620
67	2	OEAA-070810_Plate2l20.b1 070810_Plate5i22.b1	+	OEAA-	GO243723 GO244775
68	3	OEAA-070810_Plate2n22.b1 070810_Plate2f03.b1 070810_Plate3h11.b1	+ +	OEAA- OEAA-	GO243769 GO243571 GO243995
69	2	OEAA-070810_Plate2o11.b1 070810_Plate1b16.b1	+	OEAA-	GO243782 GO243121
70	3	OEAA-070810_Plate2o18.b1 070810_Plate5a04.b1 070810_Plate6d07.b1	+ +	OEAA- OEAA-	GO243789 GO244578 GO245016
71	3	OEAA-070810_Plate2p13.b1 070810_Plate4m21.b1 070810_Plate6a17.b1	+ +	OEAA- OEAA-	GO243808 GO244502 GO244956
72	2	OEAA-070810_Plate3a03.b1 070810_Plate1l19.b1	+	OEAA-	GO243821 GO243353
73	2	OEAA-070810_Plate3a05.b1 070810_Plate1k02.b1	+	OEAA-	GO243823 GO243313
74	2	OEAA-070810_Plate3a15.b1 070810_Plate3b15.b1	+	OEAA-	GO243833 GO243856
75	7	OEAA-070810_Plate3a21.b1 070810_Plate3a20.b1 070810_Plate3b04.b1 070810_Plate3c18.b1 070810_Plate4a10.b1 070810_Plate4a08.b1 070810_Plate1m23.b1	+ + + + +	OEAA- OEAA- OEAA- OEAA- OEAA- OEAA-	GO243839 GO243838 GO243845 GO243883 GO244209 GO244207 GO243381
76	2	OEAA-070810_Plate3b01.b1	+	OEAA-	GO243842

		070810_Plate1124.b1			GO243358
77	3	OEAA-070810_Plate3i07.b1 070810_Plate3a10.b1 070810_Plate3b06.b1	+ +	OEAA- OEAA-	GO244015 GO243828 GO243847
78	8	OEAA-070810_Plate6e13.b1 070810_Plate2e13.b1 070810_Plate3b08.b1 + OEAA-070810_Plate1114.b1 + OEAA-070810_Plate6a23.b1 070810_Plate2o10.b1 070810_Plate2e15.b1 070810_Plate5b11.b1	+ + + + +	OEAA- OEAA- OEAA- OEAA- OEAA- OEAA-	GO245046 GO243559 GO243849 GO243349 GO244962 GO243781 GO243561 GO244607
79	5	OEAA-070810_Plate3b14.b1 070810_Plate4o24.b1 070810_Plate5a03.b1 070810_Plate2h05.b1 070810_Plate5b23.b1	+ + + +	OEAA- OEAA- OEAA- OEAA-	GO243855 GO244551 GO244577 GO243619 GO244619
80	2	OEAA-070810_Plate3b23.b1 070810_Plate5p21.b1	+	OEAA-	GO243864 GO244938
81	2	OEAA-070810_Plate3c02.b1 070810_Plate2e05.b1	+	OEAA-	GO243867 GO243551
82	2	OEAA-070810_Plate3c10.b1 070810_Plate1g16.b1	+	OEAA-	GO243875 GO243236
83	2	OEAA-070810_Plate3c12.b1 070810_Plate3h03.b1	+	OEAA-	GO243877 GO243987
84	2	OEAA-070810_Plate3c14.b1 070810_Plate5o03.b1	+	OEAA-	GO243879 GO244897
85	2	OEAA-070810_Plate3d15.b1 070810_Plate3d16.b1	+	OEAA-	GO243904 GO243903

86	2	OEAA-070810_Plate3e20.b1 070810_Plate6g21.b1	+	OEAA-	GO243933 GO245101
87	3	OEAA-070810_Plate3e21.b1 070810_Plate6k15.b1 070810_Plate6e18.b1	+ +	OEAA- OEAA-	GO243958 GO245189 GO245051
88	2	OEAA-070810_Plate3f16.b1 070810_Plate5j16.b1	+	OEAA-	GO243953 GO244793
89	3	OEAA-070810_Plate3g19.b1 070810_Plate3b13.b1 070810_Plate4k14.b1	+ +	OEAA- OEAA-	GO243979 GO243854 GO244449
90	2	OEAA-070810_Plate3g22.b1 070810_Plate4n13.b1	+	OEAA-	GO243982 GO244517
91	3	OEAA-070810_Plate3h15.b1 070810_Plate5j20.b1 + OEAA-070810_Plate3o24.b1	+ +	OEAA- OEAA-	GO243999 GO244797 GO244176
92	3	OEAA-070810_Plate3h17.b1 070810_Plate2g22.b1 070810_Plate4e04.b1	+ +	OEAA- OEAA-	GO244001 GO243612 GO244298
93	2	OEAA-070810_Plate3h21.b1 070810_Plate2i04.b1	+	OEAA-	GO244005 GO243642
94	2	OEAA-070810_Plate3i22.b1 070810_Plate3h23.b1	+	OEAA-	GO244030 GO244007
95	2	OEAA-070810_Plate3i15.b1 070810_Plate4p09.b1	+	OEAA-	GO244023 GO244560
96	2	OEAA-070810_Plate3j05.b1 070810_Plate2g02.b1	+	OEAA-	GO244037 GO243594

104	4	OEAA-070810_Plate3m16.b1 + OEAA-070810_Plate2i05.b1 + OEAA-070810_Plate5k11.b1 + OEAA-070810_Plate3d02.b1	+	OEAA-	GO244120 GO243643 GO244812 GO243891
105	2	OEAA-070810_Plate3m23.b1 + OEAA-070810_Plate5o11.b1	+	OEAA-	GO244127 GO244905
106	3	OEAA-070810_Plate3n10.b1 + OEAA-070810_Plate2l23.b1 + OEAA-070810_Plate1f05.b1	+	OEAA-	GO244138 GO243726 GO243201
107	2	OEAA-070810_Plate3n11.b1 + OEAA-070810_Plate5c21.b1	+	OEAA-	GO244139 GO244638
108	2	OEAA-070810_Plate3n18.b1 + OEAA-070810_Plate2c11.b1	+	OEAA-	GO244146 GO243509
109	2	OEAA-070810_Plate3n24.b1 + OEAA-070810_Plate2b18.b1	+	OEAA-	GO244152 GO243495
110	2	OEAA-070810_Plate3o12.b1 + OEAA-070810_Plate5h13.b1	+	OEAA-	GO244164 GO244744
111	33	OEAA-070810_Plate3p16.b1 + OEAA-070810_Plate4l08.b1 + OEAA-070810_Plate2k06.b1 + OEAA-070810_Plate5a06.b1 + OEAA-070810_Plate1b14.b1 + OEAA-070810_Plate2e23.b1 + OEAA-070810_Plate4e17.b1 + OEAA-070810_Plate6i12.b1 + OEAA-070810_Plate1c24.b1 + OEAA-070810_Plate2p10.b1 + OEAA-070810_Plate3h01.b1 + OEAA-070810_Plate4e01.b1 + OEAA-070810_Plate5l21.b1 + OEAA-070810_Plate3i01.b1 + OEAA-070810_Plate5k21.b1 + OEAA-070810_Plate2e21.b1 + OEAA-070810_Plate3j03.b1 + OEAA-070810_Plate3b13.b1 + OEAA-070810_Plate1c05.b1 + OEAA-070810_Plate1i07.b1 + OEAA-070810_Plate3f24.b1 + OEAA-070810_Plate2f18.b1 + OEAA-070810_Plate1c13.b1 + OEAA-070810_Plate1n04.b1 + OEAA-070810_Plate3m08.b1 + OEAA-070810_Plate5h10.b1 + OEAA-070810_Plate4k22.b1 + OEAA-070810_Plate1g12.b1	+	OEAA-	GO244191 GO244466 GO243691 GO244580 GO243119 GO243567 GO244311 GO245139 GO243153 GO243805 GO243985 GO244295 GO244844 GO244009

		070810_Plate3p23.b1	+	OEAA-	GO244821
		070810_Plate4h07.b1	+	OEAA-	GO243565
		070810_Plate2a17.b1	+	OEAA-	
		070810_Plate3a09.b1 + OEAA-070810_Plate4j17.b1			GO244035
					GO243854
					GO243134
					GO243274
					GO243961
					GO243586
					GO243142
					GO243386
112	3	OEAA-070810_Plate3p20.b1	+	OEAA-	GO244195
		070810_Plate3e15.b1 + OEAA-070810_Plate3f07.b1			GO243928
					GO243944
113	2	OEAA-070810_Plate4a16.b1	+	OEAA-	GO244215
		070810_Plate4b16.b1			GO244238
114	2	OEAA-070810_Plate4a17.b1	+	OEAA-	GO244216
		070810_Plate3a12.b1			GO243830
115	2	OEAA-070810_Plate4a21.b1	+	OEAA-	GO244243
		070810_Plate3k04.b1			GO244060
116	6	OEAA-070810_Plate4b02.b1	+	OEAA-	GO244225
		070810_Plate2a24.b1	+	OEAA-	GO243477
		070810_Plate6p10.b1	+	OEAA-	GO245299
		070810_Plate5p24.b1	+	OEAA-	GO244940
		070810_Plate2e01.b1	+	OEAA-	
		070810_Plate6k21.b1			GO243547
					GO245194
117	2	OEAA-070810_Plate4b03.b1	+	OEAA-	GO244226
		070810_Plate4f01.b1			GO244319
118	2	OEAA-070810_Plate4b21.b1	+	OEAA-	GO244243
		070810_Plate4c21.b1			GO244267
119	2	OEAA-070810_Plate4c08.b1	+	OEAA-	GO244254

		070810_Plate4c07.b1			GO244253
120	3	OEAA-070810_Plate4c11.b1 070810_Plate4c24.b1 070810_Plate4k09.b1	+ +	OEAA- OEAA-	GO244257 GO244270 GO244444
121	3	OEAA-070810_Plate4c13.b1 070810_Plate4o16.b1 + OEAA-070810_Plate2j13.b1	+ +	OEAA- OEAA-	GO244259 GO244543 GO243674
122	2	OEAA-070810_Plate4d17.b1 070810_Plate5o02.b1	+ +	OEAA- OEAA-	GO244287 GO244896
123	3	OEAA-070810_Plate4e03.b1 070810_Plate1e18.b1 070810_Plate4g05.b1	+ +	OEAA- OEAA-	GO244297 GO243191 GO244347
124	5	OEAA-070810_Plate4e12.b1 070810_Plate3d22.b1 + OEAA-070810_Plate5j22.b1 + OEAA-070810_Plate1g03.b1 + OEAA- 070810_Plate2k23.b1	+ + +	OEAA- OEAA- OEAA-	GO244306 GO243911 GO244799 GO243223 GO243704
125	3	OEAA-070810_Plate4e21.b1 070810_Plate4c02.b1 070810_Plate2k09.b1	+ +	OEAA- OEAA-	GO244315 GO244248 GO243693
126	2	OEAA-070810_Plate4f03.b1 070810_Plate1m14.b1	+ +	OEAA- OEAA-	GO244321 GO243372
127	4	OEAA-070810_Plate2k07.b1 070810_Plate3j23.b1 + OEAA-070810_Plate2h20.b1 + OEAA-070810_Plate4f09.b1	+ + +	OEAA- OEAA- OEAA-	GO243692 GO244055 GO243634 GO244327
128	2	OEAA-070810_Plate4g04.b1 070810_Plate1i24.b1	+ +	OEAA- OEAA-	GO244346 GO243290

129	2	OEAA-070810_Plate4g17.b1 070810_Plate1e21.b1	+	OEAA-	GO244359 GO243193
130	6	OEAA-070810_Plate4g21.b1 070810_Plate3f18.b1 070810_Plate3d07.b1 070810_Plate3e11.b1 + OEAA-070810_Plate5f10.b1 + OEAA-070810_Plate4p21.b1	+ + + + +	OEAA- OEAA- OEAA- OEAA- OEAA-	GO244363 GO243955 GO243896 GO243924 GO244697 GO244571
131	2	OEAA-070810_Plate4h03.b1 070810_Plate2e08.b1	+	OEAA-	GO244369 GO243554
132	2	OEAA-070810_Plate4h24.b1 070810_Plate2i02.b1	+	OEAA-	GO244389 GO243640
133	5	OEAA-070810_Plate4i05.b1 070810_Plate6d16.b1 070810_Plate6f22.b1 070810_Plate6d22.b1 070810_Plate5p22.b1	+ + + + +	OEAA- OEAA- OEAA- OEAA- OEAA-	GO244394 GO245025 GO245055 GO245031 GO244939
134	3	OEAA-070810_Plate4i20.b1 070810_Plate5a04.b1 070810_Plate6b02.b1	+ + +	OEAA- OEAA- OEAA-	GO244409 GO244578 GO244598
135	3	OEAA-070810_Plate4j07.b1 070810_Plate1f18.b1 070810_Plate4g07.b1	+ + +	OEAA- OEAA- OEAA-	GO244420 GO243191 GO244349
136	2	OEAA-070810_Plate4j23.b1 070810_Plate5b06.b1	+	OEAA-	GO244434 GO244602
137	2	OEAA-070810_Plate4k13.b1 070810_Plate3j18.b1	+	OEAA-	GO244448 GO244050
138	2	OEAA-070810_Plate4l16.b1 070810_Plate5c07.b1	+	OEAA-	GO244474 GO244625

139	8	OEAA-070810_Plate4m05.b1	+	OEAA-	GO244486
		070810_Plate1g21.b1	+	OEAA-	GO243241
		070810_Plate1g06.b1	+	OEAA-	GO243226
		070810_Plate6n01.b1 + OEAA-070810_Plate2i12.b1	+	OEAA-	GO245244
		+ OEAA-070810_Plate2c13.b1	+	OEAA-	GO243717
		070810_Plate5b10.b1	+	OEAA-	GO243511
		070810_Plate5e22.b1			GO244606
					GO244686
140	2	OEAA-070810_Plate4m23.b1	+	OEAA-	GO244504
		070810_Plate1k20.b1			GO243331
141	2	OEAA-070810_Plate4n01.b1	+	OEAA-	GO244506
		070810_Plate1n15.b1			GO243397
142	2	OEAA-070810_Plate4n14.b1	+	OEAA-	GO244518
		070810_Plate2f01.b1			GO243569
143	2	OEAA-070810_Plate4o04.b1	+	OEAA-	GO244531
		070810_Plate6o07.b1			GO245273
144	2	OEAA-070810_Plate4o17.b1	+	OEAA-	GO244544
		070810_Plate2j05.b1			GO243666
145	2	OEAA-070810_Plate5a13.b1	+	OEAA-	GO244586
		070810_Plate1j12.b1			GO243299
146	2	OEAA-070810_Plate5b03.b1	+	OEAA-	GO244599
		070810_Plate5m07.b1			GO244854
147	2	OEAA-070810_Plate5b13.b1	+	OEAA-	GO244609
		070810_Plate1p15.b1			GO243445
148	4	OEAA-070810_Plate4o21.b1	+	OEAA-	GO244548
		070810_Plate5b20.b1	+	OEAA-	GO244616
		070810_Plate3b21.b1	+	OEAA-	GO243862
		070810_Plate5d18.b1			GO244659

149	2	OEAA-070810_Plate5b22.b1 070810_Plate4k07.b1	+	OEAA-	GO244618 GO244442
150	2	OEAA-070810_Plate5f05.b1 070810_Plate1c12.b1	+	OEAA-	GO244693 GO243141
151	8	OEAA-070810_Plate5f13.b1 070810_Plate6h23.b1 + OEAA-070810_Plate3j11.b1 + OEAA-070810_Plate6n19.b1 + OEAA- 070810_Plate4p22.b1 + OEAA-070810_Plate4l13.b1 + OEAA-070810_Plate4i13.b1 + OEAA- 070810_Plate3h19.b1	+	OEAA-	GO244700 GO245126 GO244043 GO245261 GO244549 GO244471 GO244402 GO244003
152	2	OEAA-070810_Plate5f22.b1 070810_Plate2c21.b1	+	OEAA-	GO244709 GO243519
153	2	OEAA-070810_Plate5g15.b1 070810_Plate6f14.b1	+	OEAA-	GO244724 GO245071
154	2	OEAA-070810_Plate5g23.b1 070810_Plate5g24.b1	+	OEAA-	GO244730 GO244731
155	3	OEAA-070810_Plate5h02.b1 070810_Plate5b02.b1 + 070810_Plate1b23.b1	+	OEAA- OEAA-	GO244733 GO244598 GO243128
156	2	OEAA-070810_Plate5h06.b1 070810_Plate5e13.b1 +	+	OEAA-	GO244737 GO244677
157	5	OEAA-070810_Plate5h09.b1 070810_Plate4h05.b1 + 070810_Plate4m06.b1 + 070810_Plate6c05.b1 + 070810_Plate1h01.b1	+	OEAA- OEAA- OEAA- OEAA-	GO244740 GO244371 GO244487 GO244991 GO243245

158	3	OEAA-070810_Plate5j02.b1 070810_Plate1g14.b1 070810_Plate4b11.b1	+ +	OEAA- OEAA-	GO244779 GO243234 GO244233
159	2	OEAA-070810_Plate5j17.b1 070810_Plate6b08.b1	+	OEAA-	GO244794 GO244970
160	2	OEAA-070810_Plate5j18.b1 070810_Plate2i13.b1	+	OEAA-	GO244795 GO243651
161	2	OEAA-070810_Plate5j24.b1 070810_Plate6j03.b1	+	OEAA-	GO244801 GO245154
162	5	OEAA-070810_Plate5l16.b1 070810_Plate5c04.b1 070810_Plate6e10.b1 070810_Plate6e02.b1 070810_Plate4b05.b1	+ + + +	OEAA- OEAA- OEAA- OEAA-	GO244840 GO244622 GO245043 GO245035 GO244228
163	3	OEAA-070810_Plate5l17.b1 070810_Plate4o05.b1 070810_Plate3d23.b1 +	+ +	OEAA- OEAA-	GO244841 GO244532 GO243912
164	2	OEAA-070810_Plate5m23.b1 070810_Plate1g24.b1	+	OEAA-	GO244870 GO243244
165	4	OEAA-070810_Plate5n04.b1 070810_Plate1p23.b1 070810_Plate5h23.b1 070810_Plate4c03.b1	+ + +	OEAA- OEAA- OEAA-	GO244874 GO243452 GO244754 GO244249
166	3	OEAA-070810_Plate5n07.b1 070810_Plate5e16.b1 + OEAA-070810_Plate5j04.b1	+	OEAA-	GO244877 GO244680 GO244781
167	2	OEAA-070810_Plate5n09.b1 070810_Plate1g10.b1	+	OEAA-	GO244879 GO243230

168	2	OEAA-070810_Plate5n12.b1 070810_Plate6g07.b1	+	OEAA-	GO244882 GO245088
169	5	OEAA-070810_Plate5n13.b1 070810_Plate6d24.b1 070810_Plate1m05.b1 070810_Plate3i14.b1 + OEAA-070810_Plate3h07.b1	+ + + +	OEAA- OEAA- OEAA-	GO244883 GO245033 GO243363 GO244094 GO243991
170	3	OEAA-070810_Plate5n15.b1 070810_Plate5h22.b1 070810_Plate1h19.b1	+ +	OEAA- OEAA-	GO244885 GO244753 GO243263
171	2	OEAA-070810_Plate5o01.b1 070810_Plate2f15.b1	+	OEAA-	GO244895 GO243583
172	2	OEAA-070810_Plate5p11.b1 070810_Plate5p11.b1	+	OEAA-	GO244928 ?
173	4	OEAA-070810_Plate6b10.b1 070810_Plate6m24.b1 070810_Plate5f16.b1 070810_Plate2d04.b1	+ + +	OEAA- OEAA- OEAA-	GO244972 GO245243 GO244703 GO243526
174	2	OEAA-070810_Plate6b22.b1 070810_Plate3i12.b1	+	OEAA-	GO244984 GO244020
175	2	OEAA-070810_Plate6c17.b1 070810_Plate6n21.b1	+	OEAA-	GO245003 GO245263
176	5	OEAA-070810_Plate6e20.b1 070810_Plate1m22.b1 070810_Plate6a10.b1 070810_Plate6a11.b1 070810_Plate3p15.b1	+ + + +	OEAA- OEAA- OEAA- OEAA-	GO245053 GO243380 GO244949 GO244950 GO244190
177	3	OEAA-070810_Plate6f02.b1 070810_Plate6b05.b1 070810_Plate4p10.b1	+ +	OEAA- OEAA-	GO245059 GO244967 GO244561

178	10	OEAA-070810_Plate6f03.b1	+	OEAA-	GO245060
		070810_Plate4o10.b1	+	OEAA-	GO244537
		070810_Plate6k01.b1	+	OEAA-	GO245176
		070810_Plate6d13.b1	+	OEAA-	GO245022
		070810_Plate4h09.b1	+	OEAA-	GO245022
		070810_Plate2b08.b1	+	OEAA-	GO244375
		070810_Plate6a08.b1	+	OEAA-	GO244375
		070810_Plate2e19.b1	+	OEAA-	GO243485
		070810_Plate3o23.b1	+	OEAA-	GO244947
		070810_Plate2h17.b1			GO243563
					GO244175
					GO243631
179	2	OEAA-070810_Plate6f19.b1	+	OEAA-	GO245076
		070810_Plate6e19.b1			GO245052
180	2	OEAA-070810_Plate6g02.b1	+	OEAA-	GO245083
		070810_Plate2g04.b1			GO243596
181	4	OEAA-070810_Plate6d16.b1	+	OEAA-	GO245025
		070810_Plate3e14.b1	+	OEAA-	GO243927
		070810_Plate3k15.b1	+	OEAA-	GO244071
		070810_Plate3p13.b1			GO244188
182	2	OEAA-070810_Plate6g18.b1	+	OEAA-	GO245098
		070810_Plate6l16.b1			GO245213
183	2	OEAA-070810_Plate6g23.b1	+	OEAA-	GO245103
		070810_Plate6l22.b1			GO245218
184	2	OEAA-070810_Plate6g24.b1	+	OEAA-	GO245104
		070810_Plate6m20.b1			GO245239
185	2	OEAA-070810_Plate6h12.b1	+	OEAA-	GO245116
		070810_Plate3m03.b1			GO244107
186	2	OEAA-070810_Plate6h17.b1	+	OEAA-	GO245121
		070810_Plate6l15.b1			GO245212
187	6	OEAA-070810_Plate6i04.b1	+	OEAA-	GO245131

		070810_Plate3e18.b1	+	OEAA-	GO243931
		070810_Plate3m17.b1	+	OEAA-	GO244121
		070810_Plate2h16.b1	+	OEAA-	GO243630
		070810_Plate6h03.b1	+	OEAA-	GO245107
		070810_Plate2h19.b1			GO243633
188	3	OEAA-070810_Plate6i16.b1	+	OEAA-	GO245143
		070810_Plate6i15.b1 + OEAA-070810_Plate6j16.b1			GO245142
					GO245167
189	2	OEAA-070810_Plate6i18.b1	+	OEAA-	GO245145
		070810_Plate5g22.b1			GO244729
190	2	OEAA-070810_Plate6i22.b1	+	OEAA-	GO245149
		070810_Plate6l21.b1			GO245217
191	2	OEAA-070810_Plate6j04.b1	+	OEAA-	GO245155
		070810_Plate2k15.b1			GO243698
192	2	OEAA-070810_Plate6j08.b1	+	OEAA-	GO243529
		070810_Plate2d07.b1			GO245159
193	3	OEAA-070810_Plate6j19.b1	+	OEAA-	GO245170
		070810_Plate1m21.b1	+	OEAA-	GO243379
		070810_Plate5o23.b1			GO244917
194	2	OEAA-070810_Plate6j23.b1	+	OEAA-	GO245174
		070810_Plate2e17.b1			GO243562
195	2	OEAA-070810_Plate6k04.b1	+	OEAA-	GO245178
		070810_Plate3k03.b1			GO244059
196	5	OEAA-070810_Plate6k10.b1	+	OEAA-	GO245184
		070810_Plate5n24.b1 + OEAA-070810_Plate2i17.b1			GO244894
		+ OEAA-070810_Plate4i09.b1	+	OEAA-	GO243655
		070810_Plate2n24.b1			GO244398
					GO243771
197	3	OEAA-070810_Plate6k11.b1	+	OEAA-	GO245185

		070810_Plate6l08.b1 + OEAA-070810_Plate6c02.b1		GO245205
				GO244988
198	3	OEAA-070810_Plate6m07.b1 + OEAA-070810_Plate6n06.b1 + 070810_Plate4b14.b1		GO245227 GO245249 GO244236
199	8	OEAA-070810_Plate6m11.b1 + OEAA-070810_Plate4l15.b1 + OEAA-070810_Plate5b09.b1 + OEAA-070810_Plate3c24.b1 + OEAA-070810_Plate1j03.b1 + OEAA-070810_Plate5n01.b1 + OEAA-070810_Plate2m22.b1 + OEAA-070810_Plate2e02.b1		GO245231 GO244473 GO244605 GO243889 GO243293 GO244872 GO243746 GO243548
200	3	OEAA-070810_Plate6n05.b1 + OEAA-070810_Plate2l11.b1 + OEAA-070810_Plate2f14.b1		GO245248 GO243716 GO243582
201	2	OEAA-070810_Plate6n13.b1 + 070810_Plate6j02.b1		GO245256 GO245153
202	2	OEAA-070810_Plate6n14.b1 + 070810_Plate6m14.b1		GO245257 GO245234
203	2	OEAA-070810_Plate6n24.b1 + 070810_Plate6i21.b1		GO245266 GO245148
204	3	OEAA-070810_Plate6o21.b1 + OEAA-070810_Plate6j22.b1 + OEAA-070810_Plate3f22.b1		GO245287 GO245173 GO243959
205	2	OEAA-070810_Plate6o24.b1 + 070810_Plate5n19.b1		GO245290 GO244889

Çizelge Ek 3: Meyve dokusuna ait contig dizilerinden her birinin içerdiği meyve EST'lerinin adı ve accession numaraları.

CONTİG no		EST adı	EST no
1	2	OEAA-070810_Plate10a15.b1+OEAA-070810_Plate10a21.b1	GO242717 + GO242723
2	69	OEAA-070810_Plate10e13.b1+OEAA-070810_Plate10k17.b1+OEAA-070810_Plate10o05.b1+OEAA-070810_Plate10k01.b1+OEAA-070810_Plate8f22.b1+OEAA-070810_Plate8m05.b1+OEAA-070810_Plate9m19.b1+OEAA-070810_Plate9n02.b1+OEAA-070810_Plate10c21.b1+OEAA-070810_Plate10e01.b1+OEAA-070810_Plate10e03.b1+OEAA-070810_Plate10g03.b1+OEAA-070810_Plate10g11.b1+OEAA-070810_Plate10h04.b1+OEAA-	GO242763+GO242958+GO243041+GO242942 + GO245827+GO245974+GO246360+GO246367 + GO242771+GO242799+GO242801+GO242848 + GO242856+GO242873+GO242912+GO242993 + GO246381+GO242753+GO242761+GO242815 + GO242864+GO242944+GO242947+GO242948 + GO242950+GO242954+GO242964+GO242991 + GO243005+GO243037+GO243039+GO243049

	070810_Plate10i19.b1+OEAA- 070810_Plate10m05.b1+OEAA- 070810_Plate9n16.b1+OEAA- 070810_Plate10c03.b1+OEAA- 070810_Plate10c11.b1+OEAA- 070810_Plate10e17.b1+OEAA- 070810_Plate10g19.b1+OEAA- 070810_Plate10k03.b1+OEAA- 070810_Plate10k06.b1+OEAA- 070810_Plate10k07.b1+OEAA- 070810_Plate10k09.b1+OEAA- 070810_Plate10k13.b1+OEAA- 070810_Plate10k23.b1+OEAA- 070810_Plate10m03.b1+OEAA- 070810_Plate10m17.b1+OEAA- 070810_Plate10o01.b1+OEAA- 070810_Plate10o03.b1+OEAA- 070810_Plate10o13.b1+OEAA- 070810_Plate10o19.b1+OEAA- 070810_Plate8b09.b1+OEAA- 070810_Plate8d10.b1+OEAA- 070810_Plate9n20.b1+OEAA- 070810_Plate10c09.b1+OEAA- 070810_Plate10e19.b1+OEAA- 070810_Plate10b02.b1+OEAA- 070810_Plate10b22.b1+OEAA- 070810_Plate10c05.b1+OEAA- 070810_Plate10c13.b1+OEAA- 070810_Plate10c17.b1+OEAA- 070810_Plate10c19.b1+OEAA- 070810_Plate10d22.b1+OEAA- 070810_Plate10e23.b1+OEAA- 070810_Plate10g09.b1+OEAA- 070810_Plate10h03.b1+OEAA- 070810_Plate10k21.b1+OEAA- 070810_Plate10m22.b1+OEAA- 070810_Plate10n19.b1+OEAA- 070810_Plate10o07.b1+OEAA- 070810_Plate10o21.b1+OEAA- 070810_Plate10o23.b1+OEAA-	+ GO243055+GO245718+GO245767+GO246385 + GO242759+GO242817+GO242728+GO242748 + GO242755+GO242763+GO242767+GO242769 + GO242796+GO242821+GO242854+GO242872 + GO242962+GO243010+GO243031+GO243043 + GO243057+GO243059+GO243079+GO245398 + GO245654+GO246125+GO246252+GO246347 + GO246432+GO246435+GO242885+GO242987 + GO246069+GO246037+GO242896+GO242837 + GO245771+
--	--	---

		070810_Plate10p19.b1+OEAA- 070810_Plate7d15.b1+OEAA- 070810_Plate7o14.b1+OEAA- 070810_Plate9c15.b1+OEAA- 070810_Plate9h23.b1+OEAA- 070810_Plate9m05.b1+OEAA- 070810_Plate9p20.b1+OEAA- 070810_Plate9p23.b1+OEAA- 070810_Plate10h16.b1+OEAA- 070810_Plate10l23.b1+OEAA- 070810_Plate9a07.b1+OEAA- 070810_Plate8o22.b1+OEAA- 070810_Plate10i03.b1+OEAA- 070810_Plate10f15.b1+OEAA- 070810_Plate8d14.b1	
3	2	OEAA- 070810_Plate10k24.b1+OEAA- 070810_Plate8e13.b1	GO242965 GO245794
4	5	OEAA- 070810_Plate10m24.b1+OEAA- 070810_Plate10l24.b1+OEAA- 070810_Plate7n10.b1+OEAA- 070810_Plate9f17.b1+OEAA- 070810_Plate8p19.b1	GO243012 GO242988 GO245626 GO246199 GO246058
5	4	OEAA- 070810_Plate10n20.b1+OEAA- 070810_Plate9k20.b1+OEAA- 070810_Plate10h18.b1+OEAA- 070810_Plate10m18.b1	GO243032 GO246317 GO242887 GO243006
6	2	OEAA-070810_Plate7c08.b1+OEAA- 070810_Plate7k20.b1	GO245367 GO245565
7	2	OEAA-070810_Plate7c12.b1+OEAA- 070810_Plate7d07.b1	GO245371 GO245390

8	2	OEAA-070810_Plate7c18.b1+OEAA-070810_Plate7h02.b1	GO245377 GO245477
9	9	OEAA-070810_Plate7c19.b1+OEAA-070810_Plate7j05.b1+OEAA-070810_Plate7l21.b1+OEAA-070810_Plate8o21.b1+OEAA-070810_Plate7m11.b1+OEAA-070810_Plate9f01.b1+OEAA-070810_Plate7f06.b1+OEAA-070810_Plate7h13.b1+OEAA-070810_Plate9p09.b1	GO245378 GO245527 GO245589 GO246036 GO245603 GO246183 GO245436 GO245488 GO246421
10	2	OEAA-070810_Plate7c24.b1+OEAA-070810_Plate7b24.b1	GO245383 GO245359
11	2	OEAA-070810_Plate7e08.b1+OEAA-070810_Plate7e09.b1	GO245416 GO245415
12	2	OEAA-070810_Plate7e15.b1+OEAA-070810_Plate7e14.b1	GO245422 GO245421
13	2	OEAA-070810_Plate7e20.b1+OEAA-070810_Plate9p03.b1	GO245426 GO246415
14	3	OEAA-070810_Plate7e22.b1+OEAA-070810_Plate7f22.b1+OEAA-070810_Plate7g22.b1	GO245428 GO245452 GO245473

15	3	OEAA-070810_Plate7f03.b1+OEAA-070810_Plate9d07.b1+OEAA-070810_Plate9i03.b1	GO245433 GO246141 GO246324
16	2	OEAA-070810_Plate7f09.b1+OEAA-070810_Plate7f11.b1	GO245439 GO245441
17	2	OEAA-070810_Plate7f10.b1+OEAA-070810_Plate7a15.b1	GO245440 GO245326
18	3	OEAA-070810_Plate7g10.b1+OEAA-070810_Plate7n17.b1+OEAA-070810_Plate8k13.b1	GO245462 GO245633 GO245936
19	6	OEAA-070810_Plate7g23.b1+OEAA-070810_Plate7p02.b1+OEAA-070810_Plate10a10.b1+OEAA-070810_Plate8c01.b1+OEAA-070810_Plate9j16.b1+OEAA-070810_Plate7i23.b1	GO245474 GO245666 GO242712 GO245734 GO246291 GO245521
20	3	OEAA-070810_Plate7i02.b1+OEAA-070810_Plate7e13.b1+OEAA-070810_Plate8a01.b1	GO245500 GO245420 GO245689
21	2	OEAA-070810_Plate7i09.b1+OEAA-070810_Plate7c23.b1	GO245507 GO245382
22	2	OEAA-070810_Plate7i18.b1+OEAA-070810_Plate8d01.b1	GO245516 GO245758

23	2	OEAA-070810_Plate7i20.b1+OEAA-070810_Plate7i19.b1	GO245518 GO245517
24	2	OEAA-070810_Plate7i21.b1+OEAA-070810_Plate7i08.b1	GO245519 GO245506
25	2	OEAA-070810_Plate7j15.b1+OEAA-070810_Plate7l05.b1	GO245537 GO245573
26	2	OEAA-070810_Plate7j23.b1+OEAA-070810_Plate7a18.b1	GO245545 GO245329
27	2	OEAA-070810_Plate7k13.b1+OEAA-070810_Plate7l06.b1	GO245559 GO245574
28	2	OEAA-070810_Plate7k22.b1+OEAA-070810_Plate7l23.b1	GO245566 GO245591
29	2	OEAA-070810_Plate7l08.b1+OEAA-070810_Plate7l09.b1	GO245576 GO245577
30	2	OEAA-070810_Plate7l13.b1+OEAA-070810_Plate7l14.b1	GO245581 GO245582
31	2	OEAA-070810_Plate7m20.b1+OEAA-070810_Plate7n19.b1	GO245612 GO245635
32	2	OEAA-070810_Plate7n03.b1+OEAA-070810_Plate8b11.b1	GO245619 GO245720

33	3	OEAA-070810_Plate7n11.b1+OEAA-070810_Plate7m22.b1+OEAA-070810_Plate7n21.b1	GO245627 GO245614 GO245637
34	385	OEAA-070810_Plate9p07.b1+OEAA-070810_Plate8c03.b1+OEAA-070810_Plate9e17.b1+OEAA-070810_Plate7o16.b1+OEAA-070810_Plate9b15.b1+OEAA-070810_Plate9g07.b1+OEAA-070810_Plate9o06.b1+OEAA-070810_Plate10h01.b1+OEAA-070810_Plate7m01.b1+OEAA-070810_Plate8o24.b1+OEAA-070810_Plate8p15.b1+OEAA-070810_Plate10c15.b1+OEAA-070810_Plate10d09.b1+OEAA-070810_Plate10g02.b1+OEAA-070810_Plate10k19.b1+OEAA-070810_Plate10p08.b1+OEAA-070810_Plate8c10.b1+OEAA-070810_Plate9i21.b1+OEAA-070810_Plate9o05.b1+OEAA-070810_Plate9o17.b1+OEAA-070810_Plate10i09.b1+OEAA-070810_Plate10i17.b1+OEAA-070810_Plate10n04.b1+OEAA-070810_Plate7c14.b1+OEAA-070810_Plate8j02.b1+OEAA-070810_Plate9b23.b1+OEAA-070810_Plate9f02.b1+OEAA-070810_Plate10e09.b1+OEAA-070810_Plate10i07.b1+OEAA-070810_Plate10n23.b1+OEAA-070810_Plate7l24.b1+OEAA-070810_Plate8d09.b1+OEAA-070810_Plate8m21.b1+OEAA-070810_Plate8o15.b1+OEAA-070810_Plate9b13.b1+OEAA-	GO246419+GO245736+GO246175+GO245656 + GO246101+GO246213+GO246395+GO242870 + GO245593+GO246039+GO246054+GO242765 + GO242783+GO242847+GO242960+GO243068 + GO245743+GO246274+GO246394+GO246406 + GO242902+GO242910+GO243016+GO245373 + GO245901+GO246109+GO246184+GO242807 + GO242900+GO243035+GO245592+GO245766 + GO245990+GO246031+GO246099+GO242731 + GO242741+GO242746+GO242776+GO242777 + GO242819+GO242824+GO242899+GO242959 + GO242967+GO242985+GO243044+GO243056 + GO243066+GO243071+GO243077+GO243080 + GO245353+GO245431+GO245472+GO245649 + GO245725+GO245770+GO245772+GO245912 + GO245945+GO245950+GO245967+GO246023 + GO246048+GO246093+GO246110+GO246131 + GO246158+GO246177+GO246190+GO246234

	070810_Plate10b05.b1+OEAA-	+
	070810_Plate10b15.b1+OEAA-	GO246298+GO246314+GO246335+GO246338
	070810_Plate10b20.b1+OEAA-	+
	070810_Plate10d02.b1+OEAA-	GO246379+GO246380+GO246434+GO242708
	070810_Plate10d03.b1+OEAA-	+
	070810_Plate10e21.b1+OEAA-	GO242729+GO242779+GO242805+GO242829
	070810_Plate10f02.b1+OEAA-	+
	070810_Plate10i06.b1+OEAA-	GO242850+GO242852+GO242889+GO242914
	070810_Plate10k18.b1+OEAA-	+
	070810_Plate10l02.b1+OEAA-	GO242920+GO242955+GO242971+GO242982
	070810_Plate10l21.b1+OEAA-	+
	070810_Plate10o08.b1+OEAA-	GO242992+GO242995+GO242998+GO243004
	070810_Plate10o20.b1+OEAA-	+
	070810_Plate10p06.b1+OEAA-	GO243017+GO243033+GO243038+GO243040
	070810_Plate10p11.b1+OEAA-	+
	070810_Plate10p17.b1+OEAA-	GO243042+GO243050+GO243051+GO243053
	070810_Plate10p20.b1+OEAA-	+
	070810_Plate7b18.b1+OEAA-	GO243061+GO243063+GO243065+GO243082
	070810_Plate7f01.b1+OEAA-	+
	070810_Plate7g21.b1+OEAA-	GO245320+GO245336+GO245350+GO245355
	070810_Plate7o09.b1+OEAA-	+
	070810_Plate8b16.b1+OEAA-	GO245476+GO245534+GO245645+GO245791
	070810_Plate8d13.b1+OEAA-	+
	070810_Plate8d15.b1+OEAA-	GO245818+GO245865+GO245886+GO245902
	070810_Plate8j13.b1+OEAA-	+
	070810_Plate8k23.b1+OEAA-	GO245915+GO245951+GO245958+GO245992
	070810_Plate8l04.b1+OEAA-	+
	070810_Plate8l22.b1+OEAA-	GO246016+GO246090+GO246095+GO246100
	070810_Plate8o07.b1+OEAA-	+
	070810_Plate8p09.b1+OEAA-	GO246103+GO246104+GO246105+GO246111
	070810_Plate9b07.b1+OEAA-	+
	070810_Plate9b21.b1+OEAA-	GO246119+GO246127+GO246136+GO246138
	070810_Plate9c21.b1+OEAA-	+
	070810_Plate9d24.b1+OEAA-	GO246167+GO246206+GO246297+GO246318
	070810_Plate9e19.b1+OEAA-	+
	070810_Plate9f08.b1+OEAA-	GO246340+GO246353+GO246392+GO246430
	070810_Plate9h04.b1+OEAA-	+
	070810_Plate9j23.b1+OEAA-	GO242727+GO242816+GO242888+GO242890
	070810_Plate9k16.b1+OEAA-	+
	070810_Plate9l14.b1+OEAA-	GO242919+GO242932+GO243048+GO243076

	070810_Plate9l19.b1+OEAA-	+
	070810_Plate9n14.b1+OEAA-	GO243084+GO245360+GO245385+GO245610
	070810_Plate9n15.b1+OEAA-	+
	070810_Plate9p22.b1+OEAA-	GO245640+GO245684+GO245712+GO245773
	070810_Plate10a06.b1+OEAA-	+
	070810_Plate10b03.b1+OEAA-	GO245922+GO246025+GO246071+GO246076
	070810_Plate10d05.b1+OEAA-	+
	070810_Plate10e07.b1+OEAA-	GO246151+GO246215+GO246289+GO246294
	070810_Plate10f07.b1+OEAA-	+
	070810_Plate10g05.b1+OEAA-	GO242730+GO242732+GO242743+GO242744
	070810_Plate10g07.b1+OEAA-	+
	070810_Plate10h20.b1+OEAA-	GO242745+GO242752+GO242756+GO242772
	070810_Plate10i21.b1+OEAA-	+
	070810_Plate10j03.b1+OEAA-	GO242773+GO242780+GO242785+GO242787
	070810_Plate10k14.b1+OEAA-	+
	070810_Plate10l06.b1+OEAA-	GO242790+GO242793+GO242795+GO242803
	070810_Plate10l17.b1+OEAA-	+
	070810_Plate10m04.b1+OEAA-	GO242809+GO242825+GO242827+GO242834
	070810_Plate10m07.b1+OEAA-	+
	070810_Plate10m10.b1+OEAA-	GO242841+GO242858+GO242860+GO242861
	070810_Plate10m16.b1+OEAA-	+
	070810_Plate10n05.b1+OEAA-	GO242863+GO242866+GO242867+GO242868
	070810_Plate10n21.b1+OEAA-	+
	070810_Plate10o02.b1+OEAA-	GO242871+GO242874+GO242876+GO242886
	070810_Plate10o04.b1+OEAA-	+
	070810_Plate10o06.b1+OEAA-	GO242895+GO242897+GO242906+GO242908
	070810_Plate10o14.b1+OEAA-	+
	070810_Plate10o15.b1+OEAA-	GO242922+GO242923+GO242925+GO242934
	070810_Plate10o17.b1+OEAA-	+
	070810_Plate10p01.b1+OEAA-	GO242943+GO242961+GO242966+GO242968
	070810_Plate10p03.b1+OEAA-	+
	070810_Plate10p05.b1+OEAA-	GO242970+GO242990+GO242999+GO243011
	070810_Plate10p22.b1+OEAA-	+
	070810_Plate7a09.b1+OEAA-	GO243022+GO243027+GO243028+GO243034
	070810_Plate7b01.b1+OEAA-	+
	070810_Plate7b15.b1+OEAA-	GO243045+GO243046+GO243047+GO243052
	070810_Plate7b20.b1+OEAA-	+
	070810_Plate7h01.b1+OEAA-	GO243054+GO243060+GO243062+GO243064
	070810_Plate7j12.b1+OEAA-	+
	070810_Plate7o05.b1+OEAA-	GO243067+GO243069+GO243070+GO243073

	070810_Plate8e10.b1+OEAA-	+
	070810_Plate8f13.b1+OEAA-	GO243075+GO243078+GO243081+GO243083
	070810_Plate8h13.b1+OEAA-	+
	070810_Plate8i11.b1+OEAA-	GO245316+GO245334+GO245351+GO245352
	070810_Plate8j03.b1+OEAA-	+
	070810_Plate8j16.b1+OEAA-	GO245356+GO245363+GO245369+GO245393
	070810_Plate8l05.b1+OEAA-	+
	070810_Plate8l12.b1+OEAA-	GO245408+GO245409+GO245455+GO245496
	070810_Plate8m23.b1+OEAA-	+
	070810_Plate8n23.b1+OEAA-	GO245544+GO245558+GO245570+GO245639
	070810_Plate9b04.b1+OEAA-	+
	070810_Plate9b09.b1+OEAA-	GO245660+GO245661+GO245681+GO245686
	070810_Plate9b14.b1+OEAA-	+
	070810_Plate9b17.b1+OEAA-	GO245714+GO245715+GO245717+GO245722
	070810_Plate9b18.b1+OEAA-	+
	070810_Plate9b19.b1+OEAA-	GO245724+GO245737+GO245755+GO245762
	070810_Plate9c01.b1+OEAA-	+
	070810_Plate9c09.b1+OEAA-	GO245765+GO245779+GO245803+GO245816
	070810_Plate9c17.b1+OEAA-	+
	070810_Plate9d02.b1+OEAA-	GO245817+GO245861+GO245862+GO245905
	070810_Plate9d04.b1+OEAA-	+
	070810_Plate9e09.b1+OEAA-	GO245908+GO245962+GO245965+GO245985
	070810_Plate9f24.b1+OEAA-	+
	070810_Plate9j22.b1+OEAA-	GO245986+GO246005+GO246009+GO246038
	070810_Plate9k21.b1+OEAA-	+
	070810_Plate9l21.b1+OEAA-	GO246011+GO246079+GO246080+GO246081
	070810_Plate9m11.b1+OEAA-	+
	070810_Plate9o03.b1+OEAA-	GO246091+GO246096+GO246108+GO246113
	070810_Plate9p18.b1+OEAA-	+
	070810_Plate10b01.b1+OEAA-	GO246114+GO246116+GO246121+GO246126
	070810_Plate10e18.b1+OEAA-	+
	070810_Plate10h19.b1+OEAA-	GO246133+GO246134+GO246152+GO246157
	070810_Plate10h21.b1+OEAA-	+
	070810_Plate10j02.b1+OEAA-	GO246163+GO246168+GO246179+GO246180
	070810_Plate10j15.b1+OEAA-	+
	070810_Plate10o12.b1+OEAA-	GO246181+GO246186+GO246205+GO246208
	070810_Plate10p16.b1+OEAA-	+
	070810_Plate10p24.b1+OEAA-	GO246232+GO246255+GO246256+GO246270
	070810_Plate7c01.b1+OEAA-	+
	070810_Plate7d02.b1+OEAA-	GO246287+GO246307+GO246320+GO246321

	070810_Plate7m18.b1+OEAA- 070810_Plate7n24.b1+OEAA- 070810_Plate7p20.b1+OEAA- 070810_Plate8b01.b1+OEAA- 070810_Plate8d16.b1+OEAA- 070810_Plate8j23.b1+OEAA- 070810_Plate8o09.b1+OEAA- 070810_Plate9a09.b1+OEAA- 070810_Plate9a14.b1+OEAA- 070810_Plate9d17.b1+OEAA- 070810_Plate9g09.b1+OEAA- 070810_Plate9j14.b1+OEAA- 070810_Plate9j19.b1+OEAA- 070810_Plate10b04.b1+OEAA- 070810_Plate10b06.b1+OEAA- 070810_Plate10b17.b1+OEAA- 070810_Plate10b18.b1+OEAA- 070810_Plate10b19.b1+OEAA- 070810_Plate10c02.b1+OEAA- 070810_Plate10c06.b1+OEAA- 070810_Plate10c22.b1+OEAA- 070810_Plate10c23.b1+OEAA- 070810_Plate10d06.b1+OEAA- 070810_Plate10d11.b1+OEAA- 070810_Plate10d13.b1+OEAA- 070810_Plate10d16.b1+OEAA- 070810_Plate10d19.b1+OEAA- 070810_Plate10d21.b1+OEAA- 070810_Plate10e05.b1+OEAA- 070810_Plate10e11.b1+OEAA- 070810_Plate10f03.b1+OEAA- 070810_Plate10f05.b1+OEAA- 070810_Plate10f12.b1+OEAA- 070810_Plate10f19.b1+OEAA- 070810_Plate10g13.b1+OEAA- 070810_Plate10g15.b1+OEAA- 070810_Plate10g16.b1+OEAA- 070810_Plate10g18.b1+OEAA- 070810_Plate10g21.b1+OEAA- 070810_Plate10g22.b1+OEAA-	+ GO246336+GO246337+GO246342+GO246343 + GO246345+GO246346+GO246362+GO246374 + GO246398+GO246404+GO246408+GO246410 + GO246412+GO246426+GO246428+GO245375 + GO246402+GO245321+GO246014+GO242733 + GO242820+GO242831+GO246137+GO246253 + GO246269+GO242788+GO245615+GO245802 + GO246386+GO242936+GO243058+GO246207 + GO242754+GO242818+GO245317+GO246117 + GO246209+GO242835+GO245616+GO246364 + GO242849+GO242937+GO242940+GO242973 + GO245667+GO245731+GO245991+GO246082 + GO246092+GO246159+GO246211+GO246225 + GO246248+GO246303+GO246407+GO242782 + GO245523+GO242709+GO242778+GO243007 + GO245509+GO245346+GO245870+GO245744 + GO245776+GO246106+GO245963+GO246003 + GO246238+GO245671+GO245685+GO242894 + GO245606
--	---	--

	070810_Plate10g23.b1+OEAA- 070810_Plate10h02.b1+OEAA- 070810_Plate10h05.b1+OEAA- 070810_Plate10h07.b1+OEAA- 070810_Plate10h17.b1+OEAA- 070810_Plate10i02.b1+OEAA- 070810_Plate10i04.b1+OEAA- 070810_Plate10i13.b1+OEAA- 070810_Plate10i15.b1+OEAA- 070810_Plate10j05.b1+OEAA- 070810_Plate10j06.b1+OEAA- 070810_Plate10j08.b1+OEAA- 070810_Plate10j17.b1+OEAA- 070810_Plate10k02.b1+OEAA- 070810_Plate10k20.b1+OEAA- 070810_Plate10l01.b1+OEAA- 070810_Plate10l03.b1+OEAA- 070810_Plate10l05.b1+OEAA- 070810_Plate10m02.b1+OEAA- 070810_Plate10m11.b1+OEAA- 070810_Plate10m23.b1+OEAA- 070810_Plate10n10.b1+OEAA- 070810_Plate10n15.b1+OEAA- 070810_Plate10n16.b1+OEAA- 070810_Plate10n22.b1+OEAA- 070810_Plate10o09.b1+OEAA- 070810_Plate10o10.b1+OEAA- 070810_Plate10o11.b1+OEAA- 070810_Plate10o16.b1+OEAA- 070810_Plate10o18.b1+OEAA- 070810_Plate10o24.b1+OEAA- 070810_Plate10p02.b1+OEAA- 070810_Plate10p04.b1+OEAA- 070810_Plate10p07.b1+OEAA- 070810_Plate10p09.b1+OEAA- 070810_Plate10p10.b1+OEAA- 070810_Plate10p13.b1+OEAA- 070810_Plate10p15.b1+OEAA- 070810_Plate10p18.b1+OEAA- 070810_Plate10p21.b1+OEAA-	
--	--	--

	070810_Plate10p23.b1+OEAA- 070810_Plate7a04.b1+OEAA- 070810_Plate7a23.b1+OEAA- 070810_Plate7b16.b1+OEAA- 070810_Plate7b17.b1+OEAA- 070810_Plate7b21.b1+OEAA- 070810_Plate7c04.b1+OEAA- 070810_Plate7c10.b1+OEAA- 070810_Plate7d10.b1+OEAA- 070810_Plate7e01.b1+OEAA- 070810_Plate7e02.b1+OEAA- 070810_Plate7g01.b1+OEAA- 070810_Plate7h22.b1+OEAA- 070810_Plate7j22.b1+OEAA- 070810_Plate7k12.b1+OEAA- 070810_Plate7l02.b1+OEAA- 070810_Plate7n23.b1+OEAA- 070810_Plate7o20.b1+OEAA- 070810_Plate7o21.b1+OEAA- 070810_Plate7p17.b1+OEAA- 070810_Plate7p22.b1+OEAA- 070810_Plate8b05.b1+OEAA- 070810_Plate8b06.b1+OEAA- 070810_Plate8b08.b1+OEAA- 070810_Plate8b13.b1+OEAA- 070810_Plate8b15.b1+OEAA- 070810_Plate8c04.b1+OEAA- 070810_Plate8c22.b1+OEAA- 070810_Plate8d05.b1+OEAA- 070810_Plate8d08.b1+OEAA- 070810_Plate8d22.b1+OEAA- 070810_Plate8e22.b1+OEAA- 070810_Plate8f11.b1+OEAA- 070810_Plate8f12.b1+OEAA- 070810_Plate8h09.b1+OEAA- 070810_Plate8h10.b1+OEAA- 070810_Plate8j06.b1+OEAA- 070810_Plate8j09.b1+OEAA- 070810_Plate8l16.b1+OEAA- 070810_Plate8l20.b1+OEAA-	
--	---	--

	070810_Plate8m16.b1+OEAA- 070810_Plate8m17.b1+OEAA- 070810_Plate8n12.b1+OEAA- 070810_Plate8n16.b1+OEAA- 070810_Plate8n18.b1+OEAA- 070810_Plate8o23.b1+OEAA- 070810_Plate9a17.b1+OEAA- 070810_Plate9a18.b1+OEAA- 070810_Plate9a19.b1+OEAA- 070810_Plate9b05.b1+OEAA- 070810_Plate9b10.b1+OEAA- 070810_Plate9b22.b1+OEAA- 070810_Plate9c03.b1+OEAA- 070810_Plate9c04.b1+OEAA- 070810_Plate9c06.b1+OEAA- 070810_Plate9c11.b1+OEAA- 070810_Plate9c16.b1+OEAA- 070810_Plate9c23.b1+OEAA- 070810_Plate9c24.b1+OEAA- 070810_Plate9d18.b1+OEAA- 070810_Plate9d23.b1+OEAA- 070810_Plate9e05.b1+OEAA- 070810_Plate9e10.b1+OEAA- 070810_Plate9e21.b1+OEAA- 070810_Plate9e22.b1+OEAA- 070810_Plate9e23.b1+OEAA- 070810_Plate9f04.b1+OEAA- 070810_Plate9f23.b1+OEAA- 070810_Plate9g02.b1+OEAA- 070810_Plate9h02.b1+OEAA- 070810_Plate9i02.b1+OEAA- 070810_Plate9i03.b1+OEAA- 070810_Plate9i17.b1+OEAA- 070810_Plate9j12.b1+OEAA- 070810_Plate9k09.b1+OEAA- 070810_Plate9k23.b1+OEAA- 070810_Plate9k24.b1+OEAA- 070810_Plate9l15.b1+OEAA- 070810_Plate9l16.b1+OEAA- 070810_Plate9l23.b1+OEAA-	
--	--	--

	070810_Plate9l24.b1+OEAA- 070810_Plate9m02.b1+OEAA- 070810_Plate9m03.b1+OEAA- 070810_Plate9m21.b1+OEAA- 070810_Plate9n09.b1+OEAA- 070810_Plate9o09.b1+OEAA- 070810_Plate9o15.b1+OEAA- 070810_Plate9o19.b1+OEAA- 070810_Plate9o21.b1+OEAA- 070810_Plate9o23.b1+OEAA- 070810_Plate9p14.b1+OEAA- 070810_Plate9p16.b1+OEAA- 070810_Plate7c16.b1+OEAA- 070810_Plate9o13.b1+OEAA- 070810_Plate7a10.b1+OEAA- 070810_Plate8n21.b1+OEAA- 070810_Plate10b07.b1+OEAA- 070810_Plate10e22.b1+OEAA- 070810_Plate10f09.b1+OEAA- 070810_Plate9d03.b1+OEAA- 070810_Plate9h24.b1+OEAA- 070810_Plate9i16.b1+OEAA- 070810_Plate10d14.b1+OEAA- 070810_Plate7m23.b1+OEAA- 070810_Plate8e21.b1+OEAA- 070810_Plate9n21.b1+OEAA- 070810_Plate10j19.b1+OEAA- 070810_Plate10o22.b1+OEAA- 070810_Plate9g01.b1+OEAA- 070810_Plate10c04.b1+OEAA- 070810_Plate10e20.b1+OEAA- 070810_Plate7a05.b1+OEAA- 070810_Plate9c07.b1+OEAA- 070810_Plate9g03.b1+OEAA- 070810_Plate10f13.b1+OEAA- 070810_Plate7m24.b1+OEAA- 070810_Plate9m23.b1+OEAA- 070810_Plate10g04.b1+OEAA- 070810_Plate10j20.b1+OEAA- 070810_Plate10j23.b1+OEAA-	
--	--	--

		070810_Plate10l08.b1+OEAA- 070810_Plate7p03.b1+OEAA- 070810_Plate8b22.b1+OEAA- 070810_Plate8m22.b1+OEAA- 070810_Plate9a20.b1+OEAA- 070810_Plate9b06.b1+OEAA- 070810_Plate9e01.b1+OEAA- 070810_Plate9g05.b1+OEAA- 070810_Plate9g19.b1+OEAA- 070810_Plate9h18.b1+OEAA- 070810_Plate9k05.b1+OEAA- 070810_Plate9o18.b1+OEAA- 070810_Plate10d08.b1+OEAA- 070810_Plate7j01.b1+OEAA- 070810_Plate10a07.b1+OEAA- 070810_Plate10d04.b1+OEAA- 070810_Plate10m19.b1+OEAA- 070810_Plate7i11.b1+OEAA- 070810_Plate7b11.b1+OEAA- 070810_Plate8h18.b1+OEAA- 070810_Plate8c11.b1+OEAA- 070810_Plate8d19.b1+OEAA- 070810_Plate9b20.b1+OEAA- 070810_Plate8l18.b1+OEAA- 070810_Plate8n10.b1+OEAA- 070810_Plate9h08.b1+OEAA- 070810_Plate7p07.b1+OEAA- 070810_Plate7p21.b1+OEAA- 070810_Plate10i01.b1+OEAA- 070810_Plate7m14.b1	
35	2	OEAA-070810_Plate7o19.b1+OEAA- 070810_Plate7o17.b1	GO245659 GO245657
36	2	OEAA-070810_Plate7o23.b1+OEAA- 070810_Plate7o22.b1	GO245663 GO245662
37	3	OEAA-070810_Plate7p01.b1+OEAA- 070810_Plate7o02.b1+OEAA-	GO245665

		070810_Plate7n22.b1	GO245642 GO245638
38	2	OEAA-070810_Plate7p24.b1+OEAA-070810_Plate9a10.b1	GO245688 GO246072
39	2	OEAA-070810_Plate8a15.b1+OEAA-070810_Plate9i11.b1	GO245702 GO246332
40	2	OEAA-070810_Plate8b07.b1+OEAA-070810_Plate8c08.b1	GO245716 GO245741
41	2	OEAA-070810_Plate8d07.b1+OEAA-070810_Plate8e07.b1	GO245764 GO245788
42	2	OEAA-070810_Plate8e14.b1+OEAA-070810_Plate8e15.b1	GO245795 GO245796
43	2	OEAA-070810_Plate8f10.b1+OEAA-070810_Plate8a09.b1	GO245815 GO245696
44	2	OEAA-070810_Plate8f15.b1+OEAA-070810_Plate8a02.b1	GO245820 GO245690
45	2	OEAA-070810_Plate8g16.b1+OEAA-070810_Plate8f16.b1	GO245845 GO245821
46	2	OEAA-070810_Plate8g19.b1+OEAA-070810_Plate8i19.b1	GO245847 GO245894
47	2	OEAA-070810_Plate8g21.b1+OEAA-	GO245849

		070810_Plate7c06.b1	GO245365
48	2	OEAA-070810_Plate8h07.b1+OEAA-070810_Plate8m11.b1	GO245859 GO245980
49	2	OEAA-070810_Plate8h08.b1+OEAA-070810_Plate7g14.b1	GO245860 GO245466
50	2	OEAA-070810_Plate8i03.b1+OEAA-070810_Plate8i04.b1	GO245878 GO245879
51	2	OEAA-070810_Plate8i05.b1+OEAA-070810_Plate8i06.b1	GO245880 GO245881
52	2	OEAA-070810_Plate8j18.b1+OEAA-070810_Plate8j19.b1	GO245917 GO245918
53	2	OEAA-070810_Plate8l07.b1+OEAA-070810_Plate8k06.b1	GO245953 GO245929
54	2	OEAA-070810_Plate8n04.b1+OEAA-070810_Plate8m10.b1	GO245997 GO245979
55	3	OEAA-070810_Plate8o04.b1+OEAA-070810_Plate8p02.b1+OEAA-070810_Plate8p03.b1	GO246020 GO246041 GO246042
56	2	OEAA-070810_Plate8p04.b1+OEAA-070810_Plate8p10.b1	GO246043 GO246049
57	2	OEAA-070810_Plate8p07.b1+OEAA-	GO246046

		070810_Plate9e08.b1	GO246166
58	2	OEAA-070810_Plate8p11.b1+OEAA-070810_Plate8p13.b1	GO246050 GO246052
59	2	OEAA-070810_Plate9b02.b1+OEAA-070810_Plate7l21.b1	GO246088 GO245613
60	2	OEAA-070810_Plate9c08.b1+OEAA-070810_Plate7c13.b1	GO246118 GO245372
61	2	OEAA-070810_Plate9c13.b1+OEAA-070810_Plate9c12.b1	GO246122 GO246123
62	2	OEAA-070810_Plate9d19.b1+OEAA-070810_Plate7i07.b1	GO246153 GO245505
63	3	OEAA-070810_Plate9e13.b1+OEAA-070810_Plate7i17.b1+OEAA-070810_Plate9a04.b1	GO246171 GO245515 GO246066
64	2	OEAA-070810_Plate9g12.b1+OEAA-070810_Plate9h01.b1	GO246218 GO246231
65	2	OEAA-070810_Plate9g16.b1+OEAA-070810_Plate9h19.b1	GO246222 GO246249
66	3	OEAA-070810_Plate9g24.b1+OEAA-070810_Plate9h13.b1+OEAA-070810_Plate7p11.b1	GO246230 GO246243 GO245675

67	2	OEAA-070810_Plate9l06.b1+OEAA-070810_Plate8p18.b1	GO246327 GO246057
68	8	OEAA-070810_Plate9n04.b1+OEAA-070810_Plate10m09.b1+OEAA-070810_Plate10n08.b1+OEAA-070810_Plate8n14.b1+OEAA-070810_Plate9d15.b1+OEAA-070810_Plate9n13.b1+OEAA-070810_Plate10h11.b1+OEAA-070810_Plate9h16.b1	GO246369 GO242997 GO243020 GO246007 GO246149 GO246378 GO242880 GO246246
69	2	OEAA-070810_Plate9o02.b1+OEAA-070810_Plate9o01.b1	GO246391 GO246390

ÖZGEÇMİŞ

Doğum tarihi	17.08.1984	
Doğum yeri	Malatya	
Lise	1998 - 2002	Küçükçekmece Lisesi
Lisans	2002 - 2006	İstanbul Üniversitesi Fen Fakültesi Biyoloji Bölümü
Yüksek Lisans	2007 - 2009	Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Biyoloji Anabilim Dalı, Moleküler Biyoloji ve Genetik Programı