



T.C.  
NECMETTİN ERBAKAN  
ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ



ADAÇAYI (*Salvia officinalis* L.)  
GENOMUNUN DE NOVO SEKANSLANMASI  
VE FONKSİYONEL ANOTASYONU

Fatma CEYLAN

DOKTORA TEZİ

Biyoteknoloji Anabilim Dalı

Kasım-2023  
KONYA  
Her Hakkı Saklıdır

## **TEZ BİLDİRİMİ**

Bu tezdeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edildiğini ve tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

## **DECLARATION PAGE**

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Fatıma CEYLAN

Tarih:

## ÖZET

### DOKTORA TEZİ

#### ADAÇAYI (*Salvia officinalis* L.) GENOMUNUN DE NOVO SEKANSLANMASI VE FONKSİYONEL ANOTASYONU

Fatıma CEYLAN

Necmettin Erbakan Üniversitesi Fen Bilimleri Enstitüsü  
Biyoteknoloji Anabilim Dalı

Danışman: Doç. Dr. Ali Tefvik UNCU

2023, 134 Sayfa

#### Jüri

Doç. Dr. Ali Tefvik UNCU  
Prof. Dr. Önder TÜRKMEN  
Prof. Dr. Kubilay Kurtuluş BAŞTAŞ  
Prof. Dr. Erdal KOCABAŞ  
Doç. Dr. Fatih ERCİ

*Salvia* cinsine ait bir tür olan adaçayı (*Salvia officinalis* L.), tarih boyunca ve günümüzde tıbbi aromatik bir bitki olarak gıda, alternatif tıp, kozmetik ve hijyen gibi alanlarda pek çok amaç için kullanılmıştır. Adaçayı yağı terpenoidlerinin tıbbi kullanımını, bu sekonder metabolitleri önemli bir araştırma konusu haline getirmektedir. Tez kapsamında yürütülen çalışmada, adaçayı genomu *de novo* sekanslanmış, assembly edilmiş, protein kodlayan gen içeriği anotasyon edilmiş ve mevcut referans genom kullanılarak varyasyon tespiti analizi gerçekleştirilmiştir. 75 terpen sentaz ve 67 terpenoid biyosentez yolak geni tahmin edilmiş ve assembly scaffoldlarına konumlandırılmıştır. Varyant analizi sonucu adaçayının kodlayan sekans bölgelerinde tahmin edilen terpenoid biyosentezi genlerine ait 188 değişken tek nükleotid lokusu tespit edilmiştir. Toplamda ise adaçayının tüm exomunda 24,570 tek nükleotid polimorfizmi tespit edilmiştir.

**Anahtar Kelimeler:** Adaçayı genomu, *de novo* sekanslama, genom sekanslama, genomiks, varyasyon analizi

## ABSTRACT

### Ph.D THESIS

#### DE NOVO SEQUENCİNG AND FUNCTIONAL ANNOTATION OF THE SAGE (*Salvia officinalis* L.) GENOME

Fatma CEYLAN

#### THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE OF NECMETTİN ERBAKAN UNIVERSITY THE DEGREE OF DOCTOR OF PHILOSOPHY IN BIOTECHNOLOGY

Advisor: Assoc. Prof. Dr. Ali Tevfik UNCU

2023, 134 Pages

#### Jury

Assoc. Prof. Dr. Ali Tevfik UNCU

Prof. Dr. Önder TÜRKMEN

Prof. Dr. Kubilay Kurtuluş BAŞTAŞ

Prof. Dr. Erdal KOCABAŞ

Assoc. Prof. Dr. Fatih ERCİ

Common sage (*Salvia officinalis* L.), a species within the *Salvia* genus, has been used throughout history and in modern times for various purposes, such as food, alternative medicine, cosmetics and hygiene, as a medicinal and aromatic plant. The medical use of the terpenoids in common sage essential oil has made these secondary metabolites a significant research subject. This thesis-based research, the common sage genome was *de novo* sequenced, assembled, its protein-coding gene content was annotated, and a variation detection analysis was performed using the existing reference genome. 75 terpene synthase genes and 67 terpenoid backbone biosynthesis pathway genes were predicted and mapped to assembly scaffolds. As a result of the variant analysis, 188 variable single nucleotide loci related to the predicted terpenoid biosynthesis genes were identified in the coding sequence regions of sage. In total, 24,570 single nucleotide polymorphisms were detected in the entire exome of common sage.

**Keywords:** Common sage genome, *de novo* sequencing, genome sequencing, genomics, variation analysis

# İÇİNDEKİLER

ÖZET .....	v
ABSTRACT .....	vi
İÇİNDEKİLER.....	vii
SİMGELER VE KISALTMALAR .....	ix
<b>1. GİRİŞ .....</b>	<b>1</b>
<b>2. KAYNAK ARAŞTIRMASI.....</b>	<b>2</b>
2.1. <i>Salvia officinalis</i> L. (Adaçayı) .....	2
2.1.1. Biyoaktif bileşenler .....	3
2.1.2. Farmakolojik aktiviteler .....	5
2.1.3. Bitkilerde terpenoid biyosentezi .....	8
2.2. Sekanslama Platformları.....	8
2.2.1. İkinci Nesil (Yeni Nesil) Sekanslama yöntemleri .....	9
2.2.2. Üçüncü Nesil (Büyük Kısım Tek Molekül (LFSM)) Sekanslama yöntemleri .....	13
2.3. Genom Assembly (Hizalama) .....	20
2.3.1. Genom hizalama işleminde de Bruijn grafik yönteminin kullanılması .....	22
2.3.2. SOAPdenovo assembler.....	23
2.3.3. Minia assembler.....	24
2.3.4. Kısa okuma <i>de novo</i> assembly analizlerinde karşılaşılan zorluklar .....	25
2.4. Genom Maskeleye .....	30
2.4.1. RepeatMasker .....	31
2.4.2. Red.....	32
2.5. Anotasyon.....	32
2.5.1. Yapısal anotasyon.....	33
2.5.2. Fonksiyonel anotasyon .....	37
2.5.3. Anotasyon yöntemleri .....	39
<b>3. MATERYAL VE YÖNTEM .....</b>	<b>41</b>
3.1. <i>Salvia officinalis</i> Bitkisinin Genomik DNA İzolasyonu .....	41
3.2. <i>Salvia officinalis</i> Genomunun Sekanslanması ve Scaffold Oluşumu .....	42
3.3. Gen İçeriğinin Değerlendirilmesi .....	42
3.4. Yapısal ve Fonksiyonel Gen Anotasyonu .....	43
3.5. Protein Kodlayan Genlerden SSR Markör Geliştirme.....	43
3.5. <i>S. officinalis</i> Exomunda Exonic Tek Nükleotid Polimorfizmlerinin (SNPs) ve Terpenoid Biyosentezinde Görev Alan Genlerin Tespiti.....	44
<b>4. ARAŞTIRMA SONUÇLARI VE TARTIŞMA.....</b>	<b>45</b>
4.1. Sekanslama Sonucu ve Assembly.....	45
vii	
4.2. Protein Kodlayan Genlerin Tespit Edilmesi ve Gen İçeriğinin Değerlendirilmesi .....	47

4.3. Fonksiyonel Anotasyon.....	50
4.4. Terpenoid Biyosentezinde Rol Alan Genlerin Tespit Edilmesi.....	55
4.5. <i>S. officinalis</i> Exomunda SNP Varyantları ve Terpenoid Biyosentez Genleri .....	63
4.6. <i>S. officinalis</i> 'in Protein Kodlayan Gen Bölgelerinde SSR Tespiti.....	67
4.7. <i>S. officinalis</i> Exomunda Terpenoid Biyosentezinde Rol Aldığı Tespit Edilen Genlerde SSR Tespiti.....	69
<b>5. SONUÇLAR VE ÖNERİLER .....</b>	<b>73</b>
<b>EKLER .....</b>	<b>74</b>
<b>6. KAYNAKLAR.....</b>	<b>119</b>
<b>ÖZGEÇMİŞ .....</b>	<b>135</b>

## SİMGELER VE KISALTMALAR

### Kısaltmalar

SBS sentezleme ile sekanslama  
PCR polimeraz zincir reaksiyon  
CRT cyclic reversible termination  
WGS whole genome sequencing  
HMM Hidden Markov Model  
EST Expressed Sequence Tag  
GMOD Genel model organizme veri tabanı  
CNVs copy number variations  
GO gene ontology  
KEGG Kyoto encyclopedia of gene and genomes

## 1. GİRİŞ

Adaçayı (*Salvia officinalis* L.) bitkisi Lamiaceae ailesine, *Salvia* cinsine ait çok yıllık yuvarlak bir çalıdır. Tarih boyunca halk tarafından nöbet, ülser, gut gibi hastalıkların tedavisinde kullanılmıştır. Son yıllarda yapılan çalışmalar adaçayının antitümör, antibakteriyel, antidepresan gibi farmakolojik aktivitelerini ortaya çıkarmıştır.

Sekanslama teknolojilerinin gelişmesi ile birlikte pek çok bitki türünün genom sekans verisi elde edilebilmektedir. Illumina gibi yeni nesil sekanslama teknolojileri ile doğruluk oranı yüksek ve yüksek verimli sekanslama sonuçları elde edilmektedir.

Yapılan tez çalışması kapsamında adaçayı genomu Illumina NovaSeq 6000 platformu kullanılarak sekanslanmıştır. De novo sekanslama sonucu adaçayı genomu contigler halinde assembly edilmiştir. Adaçayı kromozomlarının referans sekansları kullanılarak contiglerden scaffold elde edilmiştir. Yapılan anotasyon analizi sonucunda adaçayında 21,443 protein kodladığı varsayılan gen bölgesi tahmin edilmiştir.

Protein kodlayan gen bölgelerinden terpenoid biyosentezinde görev aldığı varsayılan genler tespit edilmiştir. Sonuç olarak 142 terpenoid biyosentezinde görev aldığı varsayılan gen tespit edilmiştir.

Tez çalışması kapsamında adaçayı genomunun sekanslanması, genomik nükleotid varyasyonların tanımlanmasında karşılaştırmalı sekans analiz uygulamasına olanak sağlamıştır. Sonuç olarak referans sekansa ait, terpenoid biyosentezinde görev aldığı varsayılan genlerin exon bölgesinde 447 yer değişikliği, 82 tanesi benzersiz terpenoid biyosentez genlerinde olmak üzere toplamda 188 tek nükleotid polimorfizmi tespit edilmiştir. *S. officinalis* exomunda 24,570 tek nükleotid polimorfizmi belirlenmiştir.

*Salvia officinalis* genomunun protein kodlayan gen bölgelerinde ve terpenoid biyosentezinde rol aldığı tespit edilen genlerde GMATA programı kullanılarak SSR analizi gerçekleştirilmiştir. Sonuç olarak *S. officinalis* exomunda 2452 SSR lokusu, 2125 benzersiz markör tespit edilmiştir. Terpenoid biyosentezinde rol aldığı tespit edilen gen bölgelerinde ise 93 SSR lokusu ve 85 benzersiz markör tespit edilmiştir.

## 2. KAYNAK ARAŞTIRMASI

### 2.1. *Salvia officinalis* L. (Adaçayı)

*Salvia officinalis* L. (Adaçayı) bitkisi Labiatae/Lamiaceae ailesine ait çok yıllık yuvarlak bir çalıdır.



Şekil 1. *Salvia officinalis* L (Ghorbani ve Esmailizadeh, 2017).

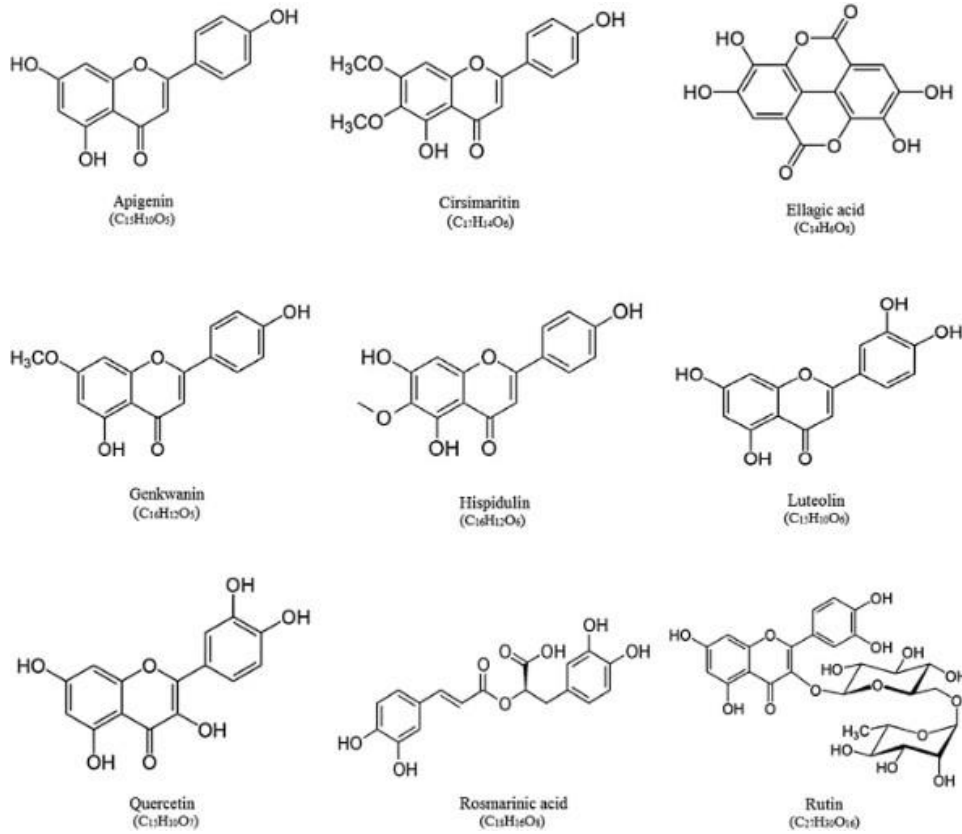
*Salvia* cinsi 900'e yakın tür içerir ve ailenin en büyük cinsidir. *S. officinalis* türünün anavatanı Orta Doğu ve Akdeniz bölgeleri iken bu cinsin bitkileri tüm dünyada yetişir. *S. officinalis* türünün toprak üstü kısımları, aşçılık ve geleneksel tıpta uzun bir kullanım geçmişine sahiptir. Aynı zamanda tatlandırıcı ve çeşni verici özelliklerinden dolayı birçok gıdanın hazırlanmasında da yaygın olarak kullanılmaktadır.

Nöbet, ülser, gut, romatizma, iltihaplanma, baş dönmesi, titreme, felç, ishal ve hiperglisemi gibi farklı türdeki rahatsızlıkların tedavisinde alternatif tıp tedavisi olarak kullanılmıştır (Garcia vd., 2016). Ayrıca bazı bölgelerde hafif hazımsızlık (mide ekşimesi ve şişkinlik gibi), aşırı terleme, yaşa bağlı bilişsel bozukluklar ve boğaz ve derideki iltihapları tedavi etmek için de kullanılmıştır (Perry vd., 2010).

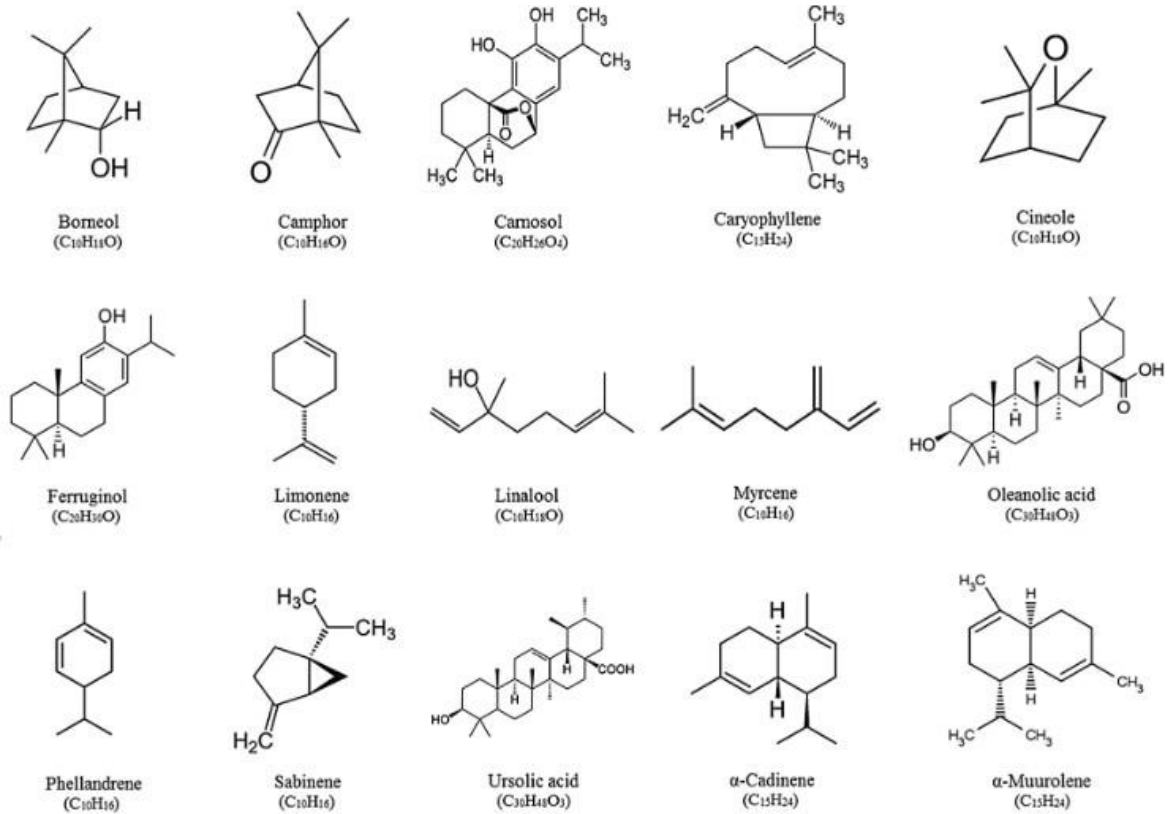
Son yıllarda, *S. officinalis*'in geleneksel kullanımını belgelemek ve bu bitki için yeni biyolojik etkiler bulmak için birçok araştırma yapılmıştır. Bu çalışmalar, antikanser, antiinflamatuvar, antioksidan, antimikrobiyal, antimutajenik, antidemans, hipoglisemik ve hipolipidemik etkiler dahil olmak üzere çok çeşitli farmakolojik aktiviteleri ortaya çıkarmıştır.

### 2.1.1. Biyoaktif bileşenler

*S. officinalis*'in çiçeklerindeki, yapraklarındaki ve gövdesindeki başlıca fitokimyasallar iyi bir şekilde karakterize edilmiştir. Çok çeşitli bileşenleri arasında alkaloidler, karbonhidrat, yağ asitleri, glikozidik türevler (örn. kardiyak glikozitler, flavonoid glikozitler, saponinler), fenolik bileşikler (örn. monoterpeneoidler, diterpeneoidler, triterpeneoidler) bulunur (Badiee vd., 2012; M. Wang vd., 1999). *S. officinalis*'ten izole edilen ana flavonoidlerin ve terpenlerin/terpeneoidlerin yapısı Şekil 2 ve 3'te gösterilmektedir.



Şekil 2. *Salvia officinalis*'ten izole edilen ana flavonoidlerin yapısı (Ghorbani ve Esmailzadeh, 2017).



**Şekil 3.** *Salvia officinalis*'ten izole edilen ana terpenlerin ve terpenoidlerin yapısı (Ghorbani ve Esmailizadeh, 2017).

*S. officinalis*'ten bildirilen fitokimyasalların çoğu, uçucu yağdan, alkollü ekstraktından, sulu ekstraktından, bütanol fraksiyonundan ve infüzyon preparasyonundan izole edilmiştir. *S. officinalis*'in toprak üstü kısımlarından hazırlanan uçucu yağda 120'den fazla bileşen karakterize edilmiştir. Alkollü ve sulu ekstraktları, flavonoidler ve özellikle rosmarinik asit ve luteolin-7-glukozit açısından zengindir. Ayrıca, metanolik ekstraktında kafeik asit ve 3-Kafeoilkinik asit gibi fenolik asitler bulunmuştur (Lima vd., 2007). Rosmarinik asit ve ellagik asit, füzyon ekstraktında en bol bulunan flavonoidlerdir, bunu klorojenik asit ve kuersetin takip eder (Hernández-Saavedra vd., 2016). Bu bitkide en çok bulunan karbonhidratlar arabinoz, galaktoz, glikoz, mannoz, ksiloz, üronik asitler ve rhamnozdur (Capek ve Hřibalová, 2004).

*S. officinalis*'in çiçek, yaprak ve gövdesindeki fitokimyasallar karşılaştırılacak olursa; linalool gövdede en çok bulunan fitokimyasaldır; çiçekler en yüksek düzeyde α-pinene ve cineole sahiptir; ve bornil asetat, kamfen, kafur, humulen, limonen ve tujon yapraklarda en çok bulunan fitokimyasallardır (Veličković vd., 2003). Bununla birlikte, diğer bitkiler gibi *S. officinalis*'in de kimyasal bileşiminin iklim, su mevcudiyeti ve

rakım gibi çevresel koşullara bağlı olarak değişebileceği göz önünde bulundurulmalıdır (A. Russo vd., 2013).

## **2.1.2. Farmakolojik aktiviteler**

### **2.1.2.1. Antikanser ve antimutajenik etkileri**

*S. officinalis*'in potansiyel antitümör aktivitesi, çeşitli kanser hücre hatları ve hayvan kanser modellerinde incelenmiştir. Ada çayı içmenin kolon karsinogenezisinin başlangıç evrelerini engellediği bildirilmiştir (Pedro vd., 2016). *S. officinalis*'in mutajenez inhibitörü olarak davrandığına dair artan kanıtlar bulunmaktadır. Uçucu yağının, *Escherichia coli* ve *Saccharomyces cerevisiae*'de UV kaynaklı mutasyonları azalttığı gösterilmiştir. Antimutajenik etkisi esas olarak tujone, kafur, limonen ve 1,8-sineol gibi monoterpen bileşiklerine atfedilir (Simić vd., 1998; Vujošević ve Blagojević, 2004; Vuković-Gaćić vd., 2006).

### **2.1.2.2. Antioksidan etkisi**

Oksidatif stres, kanser, kardiyovasküler bozukluklar, diyabet ve nörolojik hastalıklar gibi birçok hastalığın başlamasında ve ilerlemesinde önemli bir rol oynar (Neves Carvalho vd., 2017; Toyokuni, 2016). İçme suları *S. officinalis* ekstraktı ile zenginleştirilen farelerde, hepatositlerinin oksidatif strese karşı dirençlerinin arttığı gösterilmiştir (Horváthová vd., 2016). *S. officinalis*'in en etkili antioksidan bileşenleri karnosol, rosmarinik asit ve karnosik asittir, bunu kafeik asit, rosmanol, rosmadial, genkwanin ve cirsimaritin takip eder (Cuvelier vd., 1996). Bunlara ek olarak, *S. officinalis*'in diğer flavonoidleri, özellikle quercetin ve rutin, güçlü antioksidan aktivitelere sahiptir (Azevedo vd., 2013).

### 2.1.2.3. Antiseptik etkisi

*S. officinalis*'in uçucu yağı ve etanolik ekstraktı, hem Gram-pozitif hem de Gram-negatif bakterilere karşı güçlü bakterisidal ve bakteriyostatik etkiler gösterir. Gram-pozitif patojenlerden *Bacillus cereus*, *Bacillus megaterium*, *Bacillus subtilis*, *Enterococcus faecalis*, *Listeria monocytogenes* ve *Staphylococcus epidermidis*; *S. officinalis*'e karşı yüksek hassasiyet göstermektedir (Bozin vd., 2007; Hayouni vd., 2008; Longaray Delamare vd., 2007; Mitic-Culafic vd., 2005; Veličković vd., 2003).

*S. officinalis*'in Gram-negatif bakteriler üzerindeki etkileri ise kullanılan ekstraktın türüne bağlıdır. *S. officinalis*'in esansiyel yağı *Aeromonashydrophila*, *Aeromonas sobria*, *E. coli*, *Klebsiella oxytoca*, *Klebsiella pneumonia*, *Pseudomonas morgani*, *Salmonella anatum*, *Salmonella enteritidis*, *Salmonella typhi* ve *Shigellasonnei*'nin büyümesi üzerinde önemli inhibitör etkiye sahipken, etanolik ekstraktının *E. coli*, *Pseudomonas aeruginosa* ve *S. enteritidis* üzerindeki etkisi zayıftır (Bozin vd., 2007; Hayouni vd., 2008; Longaray Delamare vd., 2007; Mitic-Culafic vd., 2005; Veličković vd., 2003).

Antibakteriyel etkisine ilave olarak, *S. officinalis*'in antifungal, antiviral ve sıtmaya karşı etkileri de bildirilmiştir (Akkawi vd., 2012; Badiee vd., 2012; Carta vd., 2011). *Botrytis cinerea*, *Candida glabrata*, *Candida albicans*, *Candida krusei* ve *Candida parapsilosis*'e karşı antifungal aktivite bildirilmiştir (Badiee vd., 2012; Carta vd., 2011). *S. officinalis*'in antimikrobiyal etkileri, bu bitkide bulunan terpen ve terpenoid bileşiklere atfedilir. Kafur, tujon ve 1,8-sineolün *Aeromonas hydrophila*, *Aeromonas sobria*, *B. mega-therium*, *B. subtilis*, *B. cereus* ve *Klebsiella oxytoca*'ya karşı antibakteriyel etkileri olduğu gösterilmiştir (Longaray Delamare vd., 2007). Ayrıca, *S. officinalis*'e ait iki triterpenoid olan oleanolik asit ve ursolik asit, vankomisine dirençli enterokoklar, penisiline dirençli *Streptococcus pneumonia* ve metisiline dirençli *Staphylococcus aureus* gibi çoklu ilaca dirençli bakterilerin büyümesi üzerinde inhibitör etkiye sahiptir. Ursolik asidin *Enterococcus faecium* ve çoklu ilaca dirençli bakteriler üzerindeki etkisi, ampisilinin etkisinden daha güçlüdür (Horiuchi, vd., 2007). Bir diterpenoid olan karnosol ve bununla ilgili bir bileşen olan karnosik asit, *S. officinalis*'ten elde edilen diğer iki antibakteriyel bileşiktir. Bu bileşenler aminoglikozitlerin, metisiline dirençli olan *S. aureus* üzerindeki etkilerini güçlendirir (Horiuchi, vd., 2007).

#### 2.1.2.4. Hafıza geliştirici etkileri

Hayvan çalışmalarında, *S. officinalis*'in etanolik ekstraktının farelerde pasif kaçınma öğreniminin hafızada tutulmasını arttırdığı gösterilmiştir (M. Eidi vd., 2006). Ayrıca, *S. officinalis* hidroalkolik ekstraktının, morfinin neden olduğu hafıza bozukluğunu hafiflettiği belirtilmiştir (Gomar vd., 2014). Klinik araştırmalar sonucu *S. officinalis*'in hem sağlıklı bireylerde hem de bilişsel bozukluğu veya demansı olan hastalarda bilişsel performansı artırdığı gösterilmiştir (Miroddi vd., 2014). Aynı zamanda bazı bilim insanları, *S. officinalis* esansiyel yağının aromasının sağlıklı yetişkinlerde geleceğe yönelik hafıza performansını artırabileceğini bildirmektedir (L. Moss vd., 2010; M. Moss vd., 2014). Ayrıca, bu bitkinin etanolik ekstraktının sağlıklı yaşlı deneklerde hafızayı ve dikkati geliştirdiği bildirilmektedir (Scholey vd., 2008). Yapılan bir başka çalışmada ise *S. officinalis*'in hidroalkolik ekstraktı ile 4 aylık bir tedavinin, hafif ve orta şiddette Alzheimer hastalığı olan hastalarda bilişsel işlevleri iyileştirdiğini gösterilmiştir (Akhondzadeh vd., 2003).

*S. officinalis*'in asetilkolinesteraz aktivitesini inhibe ettiği bildirilmiştir. Asetilkolinesteraz inhibitörleri, Alzheimer hastalığının önde gelen terapötikleridir ve *S. officinalis*, bu hastalık için terapötik ajanlar geliştirmek için umut verici bir kaynak olabilir (Kennedy vd., 2005; P. Russo vd., 2013).

#### 2.1.2.5. Metabolik etkisi

Farmakolojik araştırmalar, *S. officinalis*'in toprak üstü kısımlarının farklı ekstraktlarının normal ve diyabetik koşullarda kan şekeri azaltabildiğini göstermiştir (Behradmanesh vd., 2013). Yapılan çalışmalar ayrıca *S. officinalis*'in farklı ekstraktlarının serum lipitlerini azalttığını ortaya çıkarmıştır. *S. officinalis*'in lipid profili üzerindeki yararlı etkileri diyabetik hayvanlarda da gösterilmiştir. Streptozotosin ile indüklenen diyabetik farelerde trigliserit, kolesterol, üre, ürik asit, kreatinin, aspartat aminotransferaz (AST) ve alanin aminotransferaz (ALT) düzeylerini düşürebildiği görülmüştür (A. Eidi ve Eidi, 2009; Khattab ve Mohamed, 2012).

### 2.1.2.6. Yan etkileri

Bir dizi klinik çalışma, *S. officinalis* tüketiminin ciddi yan etkilere neden olmadığını bildirmektedir (Akhondzadeh vd., 2003; Kianbakht vd., 2011; Sá vd., 2009). Ancak *S. officinalis*'in (yapraklarının 15 g'dan fazlasına tekabül eden) etanolik ekstrakt ve uçucu yağının aşırı dozda veya uzun süreli kullanımı durumunda kusma, taşikardi, baş dönmesi, sıcak basması, alerjik reaksiyonlar gibi bazı istenmeyen etkiler, dilde yutkunma, morarma ve hatta kasılma meydana gelebilir (Mills ve Bone, 2005). Kafur, tujon ve terpen ketonları, *S. officinalis*'teki en toksik bileşikler olarak kabul edilir. Bu bileşikler fetüs ve yenidoğan üzerinde toksik etkilere neden olabilir. Bu nedenle hamilelik ve emzirme döneminde *S. officinalis* tüketimi önerilmemektedir (Halicioğlu vd., 2011; Mills ve Bone, 2005).

### 2.1.3. Bitkilerde terpenoid biyosentezi

Terpenoidler pek çok önemli biyolojik ve fizyolojik fonksiyonlara sahiptir. Bazıları farmasötik önemleri ile bilinmektedir (Dubey vd., 2003). Terpenoidler 33,000'den fazla üye ile en geniş doğal bitki ürünü ailesidir (Sacchetti ve Poulter, 1997). Terpenoidler yapılarında bulunan izopren birimindeki homolog karbon sayılarına göre gruplandırılmaktadır: hemiterpenler  $C_5$  (1 izopren birimi), monoterpenler  $C_{10}$  (2 izopren birimi), sesquiterpenler  $C_{15}$  (3 izopren birimi), diterpenler  $C_{20}$  (4 izopren birimi), triterpenler  $C_{30}$  (6 izopren birimi), tetraterpenler  $C_{40}$  (8 izopren birimi), politerpenler  $(C_5)_n$  (Mcgarvey ve Croteau', 1995).

## 2.2. Sekanslama Platformları

DNA'yı sekanslamak için geliştirilen ilk yöntemler Sanger dideoksi sentez yöntemi (Sanger vd., 1977; Sanger ve Coulson, 1975) ve Maxam-Gilbert kimyasal bölünme (Maxam ve Gilbert, 1980) yöntemleridir. Maxam-Gilbert yönteminin kimyası, toksik reaktiflerin ortadan kaldırılması için değiştirilmiş olsa da Sanger sekanslama (SBS) yöntemi yaygın olarak kullanılmaktadır. Kapiller elektroforez yönteminin

kullanılmasıyla gelişmeye devam eden yöntem halen yüksek verim gerektirmeyen, plazmid yapılarının ve PCR ürünlerinin tespiti gibi, uygulamalarda kullanılmaktadır. Küçük ve kompleks olmayan genomlarda kullanımı daha yaygındır (Slatko vd., 2018). Zaman ve maliyet bakımından büyük ve kompleks genomlarda yaygın olarak tercih edilmemektedir. İlk insan genom projesinde Sanger sekanslama yöntemi kullanılmış ve projenin tamamlanması 10 yıldan fazla zaman alarak 2,7 milyar dolara mal olmuştur.

Bu maliyeti 10 milyon dolara indiren gelişmeler olmasına rağmen büyük ve kompleks genomlarda bu yöntemin kullanılması kısıtlıdır (Levy vd., 2007). Sekanslama verilerinin çoğalması için yeni teknolojilerin gelişmesi gerektiği açıktır (Hu vd., 2021).

### **2.2.1. İkinci Nesil (Yeni Nesil) Sekanslama yöntemleri**

Büyük genomların daha düşük maliyetle daha yüksek verimli sekanslanmasına duyulan ihtiyaç Yeni Nesil veya İkinci Nesil sekanslama teknolojilerinin gelişimini tetikledi. Bu yeni teknolojiler kalıp (template) hazırlama, sekanslama, görüntüleme, genom birleştirme ve genom assembly aşamalarından oluşmaktadır. Yeni nesil teknolojilerin sunduğu en büyük ilerleme büyük miktarda veriyi ucuz ve kısa zamanda üretebiliyor olmasıdır. Bu gelişme sayesinde pek çok organizmanın tüm genomunun sekanslanabilir olması, karşılaştırmalı ve evrimsel biyoloji çalışmalarına olanak sağlamıştır (Metzker, 2009).

İkinci nesil sekanslama devrimi, 454 pyro-sequencing platformunun (Margulies vd., 2005) piyasaya sürülmesi ile başladı. Daha sonra ise okuma başına 600 Gb'ye kadar sekans verisi sağlayan HiSeq2500 ile Illumina tarafından yönetildi (Ferrari vd., 2013).

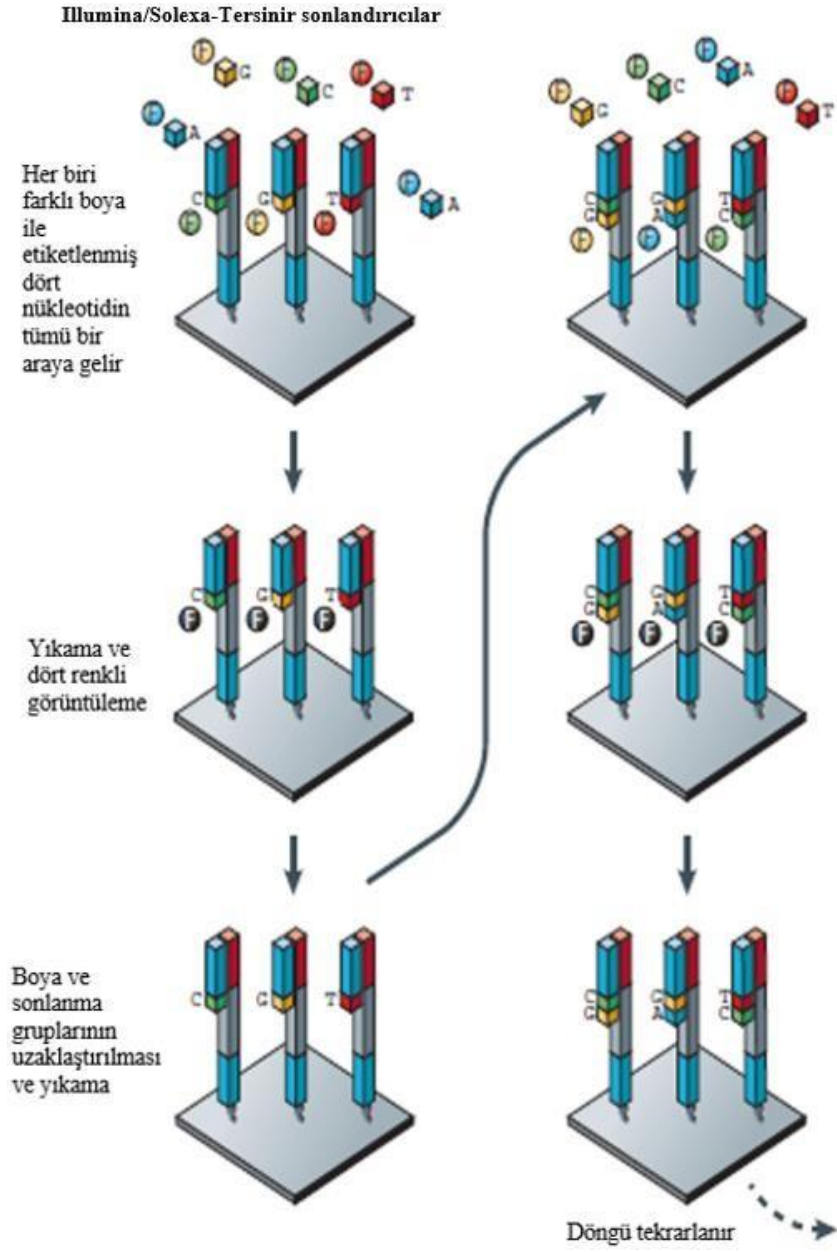
2005'te Roche/454 pyro-sequencing'in, 2007'de Illumina/Solexa sekanslamasının ve diğer yüksek verimli teknolojilerin ticarileştirilmesiyle genom sekanslamasının maliyeti hızla düştü (Mardis, 2008b). Bu durum, genomik çeşitliliği analiz etmek için geniş çaplı yeniden sekanslama çalışmalarının (Altshuler vd., 2010) yanı sıra birçok yeni genomun da sekanslanmasını (Schatz vd., 2010) sağlamıştır.

### 2.2.1.1. Illumina teknoloji (Döngüsel Tersinir Sonlandırma (CRT))

Illumina teknolojisi sentezleme ile sekanslama (SBS) yöntemine dayanmaktadır ve floresan işaretli reversibl (tersinir) bir reaksiyon gerçekleşmektedir (Goodwin vd., 2016; Mardis, 2013). Sanger sekanslama metodunda olduğu gibi reaksiyon sonlandırıcı bir molekül ile riboz 3'-OH grubu bloke edilir. Böylece DNA'nın uzaması sonlandırılmış olur (Guo vd., 2008; Ju vd., 2006).

İlk adımda, DNA polimeraz enzimi DNA kalıbına bağlanır ve floresanla modifiye edilmiş nükleotidi ekler. Floresanla modifiye edilmiş olan bu nükleotid, reaksiyonu sonlandırarak uzamayı durdurur. Tek bir nükleotidin eklenmesinden sonra DNA sentezinin sonlanması, CRT'nin önemli bir özelliğidir. Bu adımdan sonra bağlanmayan nükleotidler yıkanarak ortamdan uzaklaştırılır. Ve ortama dahil edilen nükleotidin kimliğini belirlemek için görüntüleme işlemi gerçekleştirilir. Bu işlemi, floresan boyanın ve sonlandırıcı grubun ortamdan uzaklaştırıldığı, ayırma aşaması takip eder. Sonraki birleştirme adımına başlamadan önce tekrar yıkama işlemi gerçekleştirilir (Metzker, 2009) (Şekil 4).

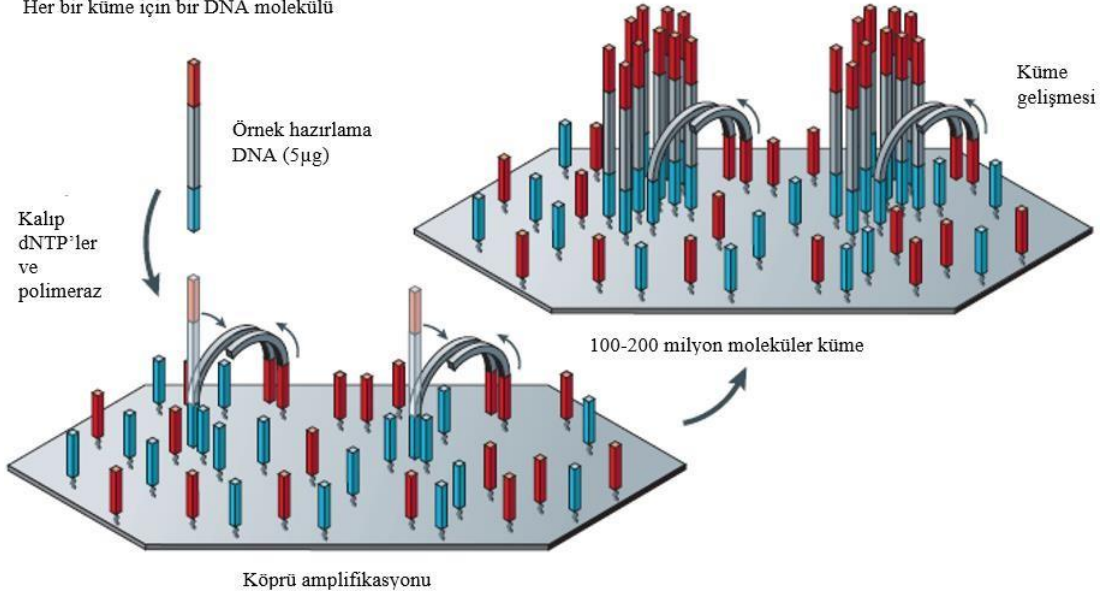
Illumina CRT sistemi (Şekil 6), diğer ikinci nesil platformlara kıyasla sekanslama araçları içinde en büyük pazar payına sahiptir (Metzker, 2009). Illumina'nın kısa okumalı sekanslama cihazları, düşük verimli tezgâh üstü ünitelerden, popülasyon düzeyinde tüm genom sekanslamaya (WGS) uygun ultra yüksek verimli araçlara kadar uzanır. Dört renkli CRT yöntemi Şekil 4 ile Şekil 5'te gösterilen klonal çoğaltılan kalıp yöntemlerinin birlikte kullanıldığı prensiple çalışır. Dört renk, iki lazer kullanılarak toplam iç yansıma floresan (TIRF) görüntüleme ile algılanır (Şekil 7).



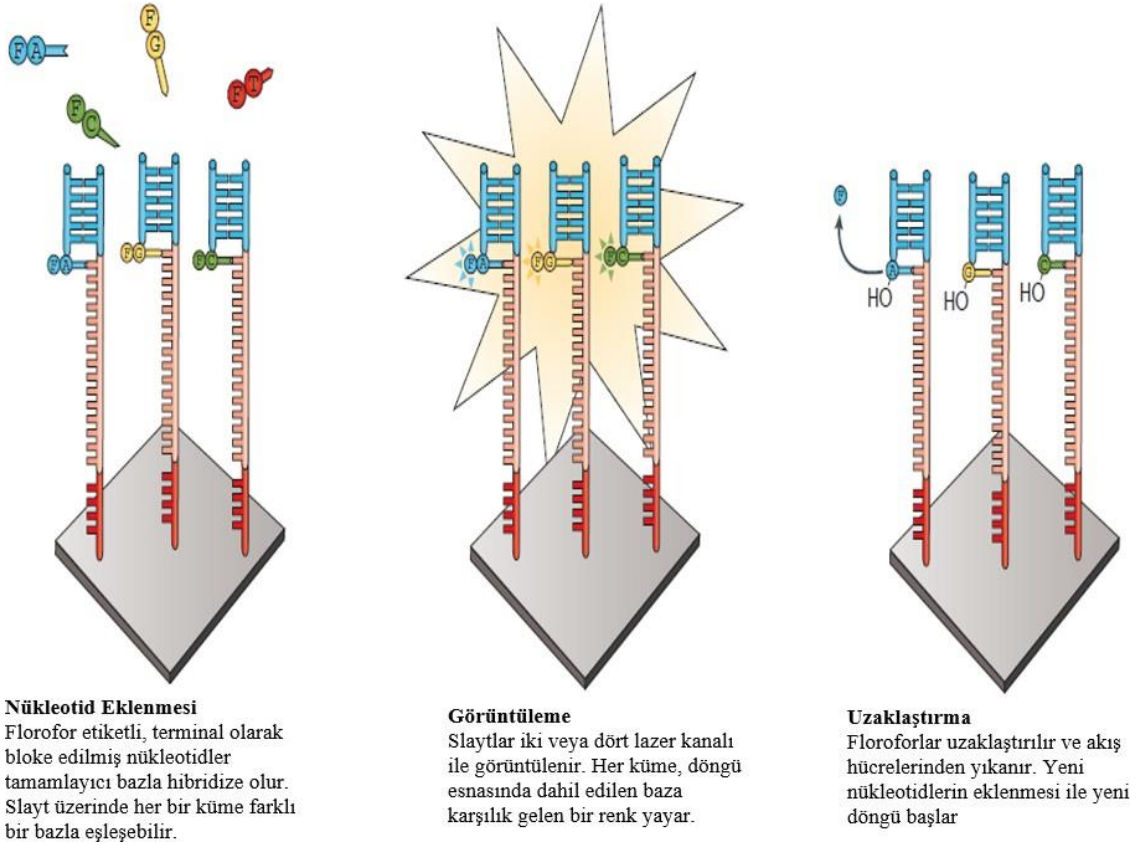
Şekil 4. Döngüsel Tersinir Sonlandırma teknolojisinin gösterimi (Metzker, 2009)

### Illumina/Solexa Katı-faz amplifikasyonu

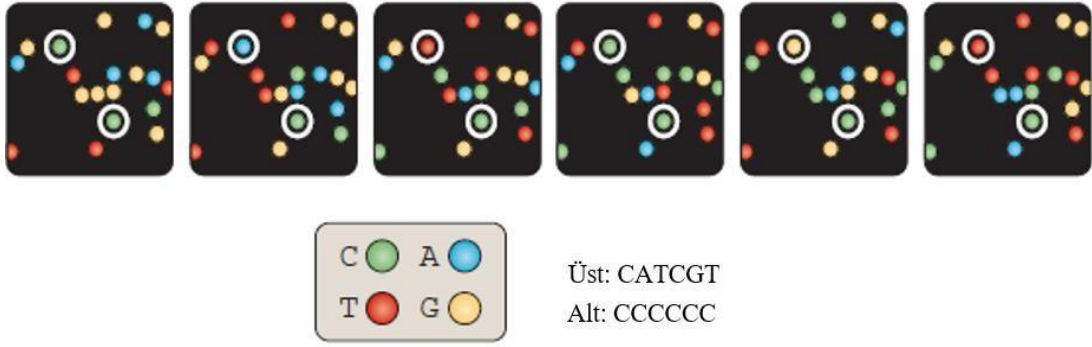
Her bir küme için bir DNA molekülü



Şekil 5. Döngüsel Tersinir Sonlandırma teknolojisinin gösterimi (Metzker, 2009)



Şekil 6. Illumina CRT sistemi (Goodwin vd., 2016)



**Şekil 7.** İç yansıma floresan görüntüleme (Metzker, 2009)

Dört renkli döngüsel tersinir sonlandırma (CRT) yöntemi, Illumina/Solexa'nın katı fazla güçlendirilmiş kalıp kümeleri kullanan 3'-O-azidometil tersinir sonlandırıcı kimyasını kullanır. Görüntülemenin ardından, bir ayırma adımı floresan boyaları uzaklaştırır ve indirgeyici ajan tris(2-karboksietil) fosfin (TCEP) kullanarak 3'-OH grubunu yeniden oluşturur.

### 2.2.2. Üçüncü Nesil (Büyük Kısım Tek Molekül (LFSM)) Sekanslama yöntemleri

İkinci nesil sekanslama yöntemlerinin aksine üçüncü nesil sekanslama yöntemleri uzun DNA ve RNA moleküllerini sekanslamayı amaçlar (Slatko vd., 2018). İkinci nesil sekanslama yöntemleri birkaç yüz baz çifti uzunluğunda kısa okumalar üretirken, üçüncü nesil tek molekül teknolojileri 10.000 baz çifti üzerinde okuma üretir veya 100.000 baz çifti üzerinde moleküllü haritalar (Lee vd., 2016).

İkinci nesil sekanslama yöntemleri ile, tek nükleotit varyasyonların ve diğer küçük varyantların popülasyon ölçeğinde analizleri mümkün kılınmış olsa da daha büyük yapısal varyasyonların analizinde teknolojinin veriminin düştüğü kanıtlanmıştır. Ayrıca, ikinci nesil teknolojiler kullanılarak yeniden bir araya getirilen yeni genomlar, daha eski, daha pahalı yöntemler kullanılarak sekanslanan genomlara kıyasla genellikle daha düşük kalitededir (Huang vd., 2010; Lander vd., 2001). Yalnızca kısa okumalar ile oluşturulmuş *de novo* genom analizlerinde genomun önemli miktarda bölümünden veri gelmeyebilir, önemli gen bölgeleri parçalanmış veya eksik olabilir ve genel kromozom yapısını incelemek için gereken yeterlilik olmayabilir (Li vd., 2009; Schatz vd., 2010). Bazı durumlarda, assembl edilmiş sekanslar, ortalama gen boyutundan önemli ölçüde daha küçük olabilmektedir ve bu da sekansı önceki referans genomlarından daha az kullanışlı hale getirmektedir (Jia vd., 2013).

10.000 baz çiftinden fazla ortalama okuma uzunlukları ve 100.000 baz çifti veya daha fazla okuma uzunlukları üretebilen yeni tek molekül sekanslama teknolojilerinin mevcudiyeti, genom yapısının büyük ölçüde geliştirilmiş analizini sağlamıştır. Daha da

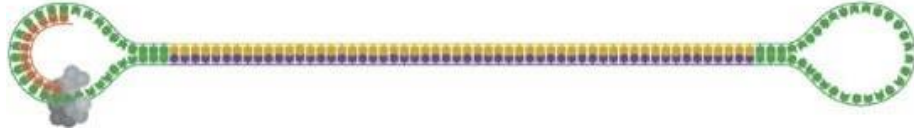
önemlisi, daha uzun okuma uzunlukları daha fazla tekrar eden bölge kapsar ve böylece daha büyük genom bütünlüğü üretir (Roberts vd., 2013). Yapısal varyasyon analizi ile ilgili olarak ise uzun okumalar, eklemeler (insertions), silmeler (deletions), yer değiştirmeler (translocations) ve diğer yapısal değişikliklerin daha kolay tespit edilebilmesi için geliştirilmiş "bölünmüş okuma" analizlerine olanak tanır (Chaisson vd., 2014). Üçüncü nesil sekanslama ve haritalama teknolojilerini birlikte kullanarak, neredeyse tüm kromozom kollarını kaplayabilen ve büyük ölçüde geliştirilmiş yapısal analize olanak sağlayabilen süper-bitişikler/contigler ("iskeletler/scaffolds") oluşturmak mümkündür (Cao vd., 2014; Pendleton vd., 2015).

Üçüncü nesil teknolojiler, yüzlerce mikrobiyal genomun doğruluk oranı yüksek *de novo* assembl analizlerini ve birçok bitki ve hayvan genomunun contig (bitişiklik) oranı yüksek yapılarını oluşturmak için kullanılmış olup evrim ve çeşitlilik analizlerine yeni bir bakış açısı sağlamıştır (Berlin vd., 2015; Gordon vd., 2016).

Piyasada bulunan üçüncü nesil DNA sekanslama teknolojileri, Pacific Biosciences (PacBio) Tek Molekül Gerçek Zamanlı (SMRT) sekanslama, Illumina Tru-seq Sentetik Uzun Okuma teknolojisi ve Oxford Nanopore Technologies sekanslama platformlarıdır. Bu teknolojiler ortalama 5.000 baz çifti ile 15.000 baz çifti arasında uzun okumalar üretebilir ve bazı okumalar 100.000 baz çiftinin üzerindedir. İkinci nesil yöntemlerin aksine, bu yöntemler sekanslama için kütüphane oluşturma aşamasında bir amplifikasyon adımı içermez, bu sayede tek molekül sekanslamayı mümkün kılar (Lee vd., 2016).

### **2.2.2.1. Pacific Biosciences (PacBio) Tek Molekül Gerçek Zamanlı (SMRT) sekanslama**

Pacific Biosciences (PacBio) veya SMRT (Tek Molekül Gerçek Zamanlı) sekanslama, üçüncü nesil bir sekanslama yöntemidir. Bu sekanslama işleminde, sekanslanacak DNA, SMRTbell kütüphanesi olarak adlandırılan tek sarmallı dairesel bir DNA olarak bulunur. SMRTbell kütüphanesi, saç tokası (hairpin) adaptörlerinin (SMRTbell adaptörleri) çift sarmallı DNA (dsDNA) kalıp molekülünün her iki ucuna bağlanmasıyla oluşturulur (Şekil 8).



**Şekil 8.** SMRTbell kütüphanesi (Travers vd., 2010)

Saç tokası adaptörleri (yeşil), çift sarmallı bir DNA molekülünün (sarı ve mor) ucuna bağlanarak kapalı bir daire oluşturur. Polimeraz (gri), bir ZMW'nin tabanına sabitlenmiştir ve bazıları okunan zincire (turuncu) dahil eder.

Sekanslama reaksiyonu, çok gözenekli bir yapıya sahip olan ve her bir gözeneği sıfır modlu dalga kılavuzları (ZMW – zero mode waveguides) olarak adlandırılan "SMRT Hücresi (SMRT Cell)" çipinde (Şekil 9) gerçekleşir ve bu gözeneklerin her biri yaklaşık 70 nm çapında ve 100 nm derinliğindedir. PacBio sekanslama platformlarından daha eski olan PacBio RS II sistemi her bir SMRT Hücresi başına 150.000 ZMW'ye sahip iken, daha yeni olan Sequel sistemi her bir SMRT Hücresi başına bir milyon ZMW'ye sahiptir. Her bir ZMW'de immobilize edilmiş ayrı bir DNA polimeraz molekülü, tek bir SMRTbell kalıbının sekanslanmasını sağlar (Rhoads ve Au, 2015).

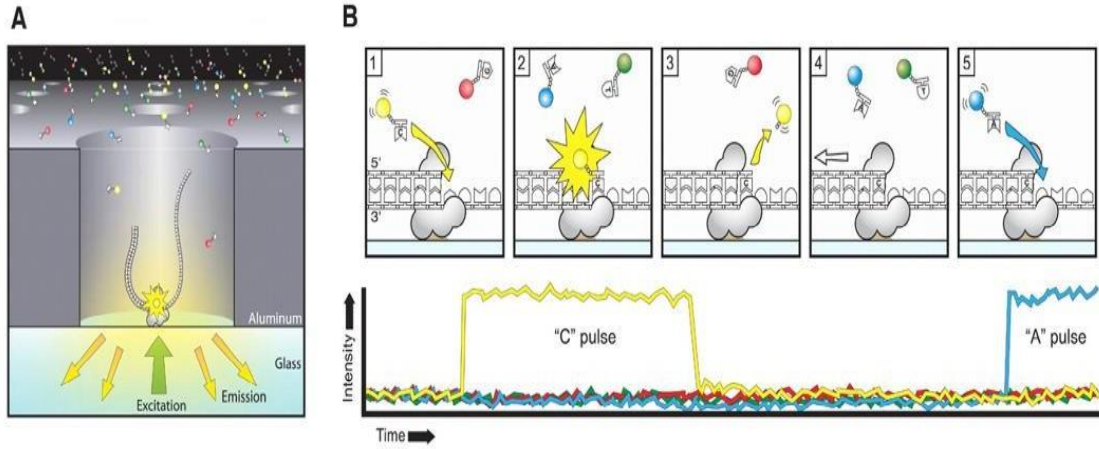


**Şekil 9.** SMRT hücresi (Media resource center – PacBio)

Her SMRT hücresi 150.000 ZMW içerir. Bu kuyucukların yaklaşık 35.000–75.000'i, 0,5–4 saat süren bir çalışmada okuma üreterek 0,5–1 Gb sekansla sonuçlanır.

Her ZMW'de, SMRTbell'in saç tokası adaptörlerinden herhangi birine bağlanabilen ve replikasyonu başlatabilen tek bir polimeraz tabanda sabitlenmiş haldedir (Şekil 10A). SMRTbell kütüphanesi SMRT Hücresine yüklenir, polimeraz SMRTbell adaptörüne bağlanır ve replikasyon başlar. Farklı emisyon spektrumları oluşturan dört floresan etiketli nükleotit, SMRT Hücresine eklenir. Bir baz, polimeraz

tarafından tutulduğunda, bazı tanımlayan bir ışık darbesi üretilir (Şekil 10B) (Rhoads ve Au, 2015).



**Şekil 10 A/B.** Işık darbeleri ile sekanslama (Korlach vd., 2010)

- A.** Bir SMRTbell (gri) bir ZMW'ye yayılır ve adaptör tabanda sabitlenmiş bir polimeraza bağlanır.
- B.** Dört nükleotidin her biri farklı bir floresan boyayla (G, C, T ve A için sırasıyla kırmızı, sarı, yeşil ve mavi olarak gösterilir) etiketlenir, böylece farklı emisyon spektrumlarına sahip olurlar. Bir nükleotit, polimeraz tarafından tespit hacminde tutulduğu için, bazı tanımlayan bir ışık darbesi üretilir.

SMRT Hücrsindeki tüm ZMW'lerde açığa çıkan ışık darbeleri bir 'film'e kaydedilir. PacBio RS II, genel olarak 0,5–4 saat uzunluğunda sekans filmleri üretir (Rhoads ve Au, 2015. Her bir ışık darbesi bir nükleotit sekansı olarak yorumlanır ve her bir ZMW'den elde edilen sekans bilgisine "Sürekli Uzun Okuma" (Continuous Long Read (CLR)) adı verilir (Hu vd., 2021).

SMRTbell kapalı bir daire oluşturduğu için, polimeraz hedef dsDNA'nın bir sarmalını kopyaladıktan sonra adaptörün bazlarını ve ardından diğer sarmalı birleştirmeye devam edebilir. Polimerazın ömrü yeterince uzunsa, her iki sarmal da tek bir CLR'de birden çok kez sekanslanabilir. Bir DNA sarmalının bir kez sekanslanması "geçiş (pass)" olarak adlandırılır ve bir DNA birden çok kez sekanslanırsa buna "geçişler (passes)" denir. Böyle bir senaryoda CLR, adaptör sekanslarını tanıyarak ve keserek çoklu okumalara (subreads) bölünebilir. Tek bir ZMW'deki çoklu alt okumaların oluşturduğu sekans birliği, daha yüksek doğrulukla okunan bir dairesel sekans (Circular Consensus Sequence (CCS)) bilgisi verir. Böylece her bir alt okumadaki hatalar diğer alt okumalar tarafından düzeltilecektir.

Her bir polimeraz enziminin SMRTbell kütüphanesini etkili bir şekilde sekanslayabileceği sınırlı bir ömrü vardır. Bu nedenle daha uzun DNA kalıplarına

kıyasla daha kısa DNA kalıplarında, daha fazla geçiş üretilebilecektir (Hu vd., 2021; Rhoads ve Au, 2015).

PacBio sekanslamanın önemli bir zayıflığı, bir CLR'nin hata oranının, nispeten yüksek (11-15% civarında) olmasıdır (Korlach). Hatalar CLR'lerde rastgele dağıtıldığı için, yeterli sekans geçişleriyle CCS okumaları üretilerek hata oranı azaltılabilir. Ortalama 15 geçişlik bir sonuç >99% doğruluk sağlar (Eid vd., 2009). Bununla birlikte, bir CLR'nin toplam uzunluğunun polimerazın ömrü ile sınırlı olduğu göz önüne alındığında, sekanslama geçişlerinin sayısı ve CCS okuma uzunluğu bir öngörülemezdir (Detter vd., 2014).

Yeni nesil sekanslama ve PacBio sekanslamanın güçlü ve zayıf yönleri birbirini tamamlar, bu da her iki yöntemi birlikte kullanan ve yenilikçi bir strateji olan hibrit sekanslamayı teşvik eder. Bu yaklaşımlar genellikle uzun okumadaki hataları düzeltmek için yüksek verimli ve yüksek doğruluklu kısa okuma verilerinin kullanılmasını içerir. Kısa okumalar, alignment/assembly/tespitleri tek nükleotit çözünürlüğe kadar hassaslaştırılabilirken; PacBio uzun okumaları alignment ve scaffoldları güvenli düzeyde oluşturabilir ve genomik varyantların kaba tespitlerini sağlayabilir.

Uzun okuma teknolojisi olarak PacBio teknolojisinin çeşitli avantajları vardır. Oluşturulan okumalar çok uzun olabilir. Okumaların ilk 5%'inin 135 kb'den uzun olabileceği bildirilmiştir. Bu yöntemle veriler gerçek zamanlı olarak toplanır, bu sayede ikinci nesil yöntemlere göre daha hızlı geri dönüş süresi sunar. Kısa okuma yöntemleri sıklıkla PCR ile amplifikasyon işlemi gerektirse de PacBio DNA amplifikasyonu gerektirmez ve bu nedenle PCR ile ilişkili sorunları (AT ve GC açısından zengin bölgelerde amplifikasyon zorluğu) barındırmaz.

PacBio yönteminin önemli bir dezavantajı yüksek düzeyde hata oranı (~14%). Yüksek hata oranının azaltmak için, kısa okuma ve PacBio yöntemlerini birleştiren hibrit sekanslama yaklaşımları kullanılmıştır (Berbers vd., 2020). HiFi sekanslamanın yüksek doğruluğu, PacBio sekanslamada görülen hata oranlarının azalmasına yardımcı olmuştur. Bir başka dezavantaj ise, Oxford Nanopore sekanslama ile karşılaştırıldığında, başlangıç için gereken önemli finansal yatırım ve cihazların nispeten büyük boyutudur (Hu vd., 2021).

Genel olarak, PacBio sekanslama, yüksek hata oranı ve düşük verim ile çok uzun okumalar sağlar. Altıncı nesil polimerazı ve dördüncü nesil kimyayı (P6-C4 kimyası) kullanan PacBio RS II, ikinci nesil sekanslama platformlarından daha uzun ortalama okuma uzunluğu sağlar, ancak daha yüksek hata oranına ve daha düşük verime

sahiptir. Ayrıca PacBio sekanslama, diğer birçok yöntemden daha hızlı ancak daha maliyetlidir (Rhoads ve Au, 2015).

### 2.2.2.2. Oxford Nanopore Teknoloji (ONT) sekanslama

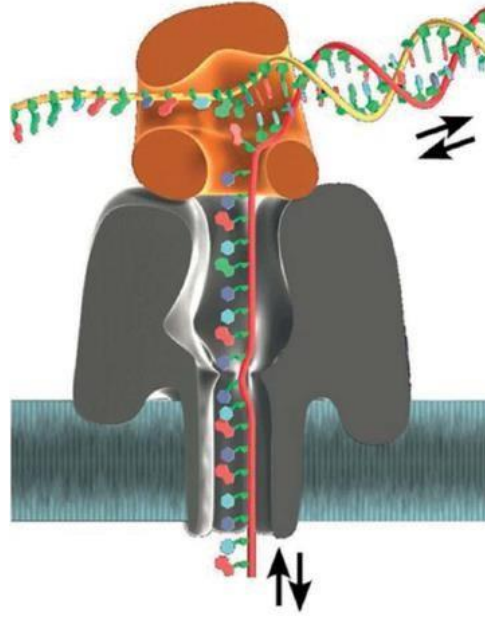
Oxford Nanopore Teknolojisi (ONT) sekanslama, 1 Mb'den daha büyük okumalar üretebilir (Miga vd., 2020). ONT sekanslama, tek sarmallı nükleik asidin (DNA veya RNA) bir stafilokokal  $\alpha$ -hemolizin ( $\alpha$ HL) protein porundan geçişine dayanır (Clarke vd., 2009). Çift sarmallı DNA'ya bağlanan adaptörler, DNA'nın protein porları tarafından yakalanmasını kolaylaştırır. Adaptörün 5' ucuna önceden bağlanmış olan bir motor enzimi kanal boyunca bir iyon akımı uygulayarak tek sarmal halindeki ipliği por boyunca hareket ettirir. Her bir nükleotidin pordan geçişi ile iyon akımında karakteristik bir değişim meydana gelir ve bu değişim sensörler tarafından algılanır. Kütüphaneler, nanoporların gömülü olduğu bir zar içeren bir akış hücresine yüklenir (Hu vd., 2021).

MinION (Şekil 11) ile ONT, nanopor sekanslama için mevcut ilk cihazı geliştirdi. Bu sekanslama cihazları küçük bir cep telefonu boyutundadır ve bir dizüstü bilgisayarın USB'sine takılabilir, bu da onları "sekanslama işlemi" için ideal hale getirir.



Şekil 11. Bir dizüstü bilgisayara bağlı ONT MinION sekanslama cihazı (Bleidorn, 2015).

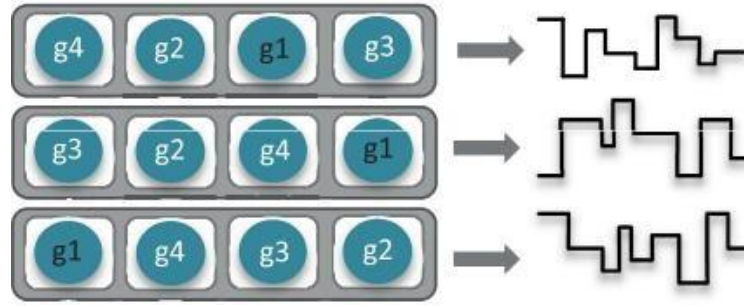
MinION sekanslama cihazı, her biri saniyede yaklaşık 10 baz çifti tespit eden nanoporlar içeren 512 kanala sahiptir. Sekanslama kütüphanesi oluşturmak için her bir DNA molekülünün bir ucuna bir saç tokası adaptörü bağlanırken diğer uca bir motor proteini bağlanır. Motor proteini, DNA molekülünü tek sarmallı hale geldiği nanopor boyunca hareket ettirir (Şekil 12).



**Şekil 12.** Bir DNA polimeraz (kahverengi) tarafından sağlanan hız kontrolü ile bir nanopora yerleştirilmiş DNA. Bir  $\alpha$ -hemolizin nanoporu (gri), tamponlanmış potasyum klorür çözeltisi içeren ve iki bölme ayıran bir lipid çift tabakasına gömülüdür (Schneider ve Dekker, 2012).

DNA'nın ikili sarmalı, bir saç tokası adaptörü ile bağlandıkları için birbirinden ayrılmaz. İdeal durumda, bir sarmal nanopor içinden geçer, ardından saç tokası adaptörü ve molekülün ikinci sarmalı gelir. Bu sayede sekanslanmış DNA çift ipliği için bir konsensus sekansı oluşturmak mümkün olur.

Standart uzun okumalar, ONT sekanslamasının akış hücresi kapasitelerinde farklılık gösteren üç platformunda üretilebilir. Taşınabilir cep boyutunda bir cihaz olan MinION her bir kanalda dört nanopora sahip 512 kanallı bir adet akış hücresi bulundururken, GridION beş akış hücresine kadar bulundurabilir. Belirli bir zamanda, sekans oluşturan kanal başına yalnızca bir aktif por olduğu ve diğer üçünün devre dışı olduğu göz önüne alındığında, kanalların sayısı aynı anda sekanslanabilen DNA moleküllerinin sayısına karşılık gelir (Ip vd., 2015).



**Şekil 13.** Bir akış hücresi, her biri 4 kuyudan (beyaz) oluşan 512 kanal (gri) içerir. Her kuyu bir por (mavi) ve bir sensör içerir. Cihaz, aktif kuyu gruplarından gelen veri akışını kaydediyor (Ip vd., 2015).

PromethION, 12.000 nanopora sahip 3000 kanal içeren 48 adede kadar akış hücresini barındırabilir. Sonuç olarak, yaklaşık altı kat daha fazla kanala sahip olan PromethION, akış hücresi başına altı kat daha fazla verim sağlayabilir ve akış hücresi başına 50-100 Gb uzun okuma verisi üretebilir (Logsdon vd., 2020).

Uzun okuma teknolojisi olarak ONT'nin genel avantajları PacBio'nunkilere benzer. Polimorfik genlerin daha iyi analiz edilmesi, gerçek zamanlı veri elde edebilme, DNA amplifikasyon işlemi sırasında ortaya çıkan hataların oluşmaması gibi. PacBio veya ikinci nesil sekanslama platformlarıyla karşılaştırıldığında, ONT cihazları düşük maliyetli, taşınabilir ve önemli ölçüde daha küçük boyutlu olma gibi avantajlara sahiptir; bu sayede, düşük gelirli ayarlar veya saha uygulamaları için çok yararlı olabilir (Quick vd., 2016).

Uzun okuma teknolojisi olarak ONT'nin genel dezavantajları, verilerin, kısa okuma teknolojilerinde görülene göre daha yüksek düzeyde hata (2-15%) ile sonuçlanmasıdır. Daha yavaş nükleotit işleme hızı sağlamak için nanopor proteininde yapılan değişiklikler, daha iyi veri toplamaya ve hata oranlarında gelişime olanak sağlamaktadır (Hu vd., 2021).

### 2.3. Genom Assembly (Hizalama)

Bir genoma ait DNA'nın sekanslanması işlemi, farklı baz çifti uzunluklarında, genomik DNA'nın daha kısa parçalara bölünmesi halinde gerçekleştirilir. Sekanslanan DNA parçalarının uzunluğu, kullanılan teknolojiye göre değişiklik gösterir. İkinci nesil sekanslama teknolojileri kısa okumalar üretirken, üçüncü nesil sekanslama teknolojileri uzun okumalar üretmektedir. Ancak her iki okuma teknolojisi neticesinde de,

sekanslanırken parçalara ayrılmış olan DNA parçalarının bir araya getirilmesi ve anlamlandırılması gerekmektedir. Sekanslanan DNA parçalarının bir araya getirilip anlamlı bir şekilde birleştirilmesi işlemine “assembly” denilmektedir.

İlk olarak sekanslanan DNA parçaları bir araya getirilerek bitişik diziler (contig) oluşturulur. Contiglerin bir araya gelmesi ile iskele (scaffold) meydana gelir. Scaffoldların bir araya gelmesi ile birlikte genom anlamlı bir şekilde birleştirilmiş olur.

DNA sekanslamanın temel amacı, DNA sekans verisinin büyük ve sürekli halini oluşturmak olmuştur. Sekanslanmak istenen DNA'nın büyüklüğü genellikle okuma sonucu elde edilen sekanslama uzunluklarından çok daha büyüktür. Bu nedenle bir tür örnekleme yaklaşımı gerekir. Baskın yaklaşım, bir DNA sekansını rasgele parçalara ayırmak, "shotgun sequencing" olarak tanımlanan bu parçaları sekanslamak ve ardından "assembly" olarak tanımlanan orijinal DNA sekansını hesaplamalı olarak yeniden bir araya getirmek olmuştur. Bu, başlangıçta kozmid veya diğer klon kaynakları için tasarlanmıştı, fakat daha sonra, önce bakterilerin ve sonra büyük karmaşık ökaryotların genomları için de uygulanabilir hale gelmiştir (Staden, 1979).

Eski teknoloji okuma uzunluklarıyla (yaklaşık 800 bp), tüm assembly algoritmaları, okumalar arasındaki örtüşmeleri kullanmanın bir varyasyonu olarak çalıştı ve ardından bu örtüşmeleri doğrusal bir hale getirerek bütüne ulaştı. İkinci nesil teknolojilerin ürettiği çok daha kısa okumalar ve çok daha yüksek derinliklerle, bu "okuma merkezli" yöntem hem hesaplama bakımından olanaksız hale geldi, hem de çok sayıda örtüşmeyi çözmek için buluşsal yöntemler bulmak da zorlandı. Bununla birlikte, 1980'lerin sonlarında Pevzner ve meslektaşlarının (Pevzner vd., 2012) ve 1990'ların ortalarında Idury ve Waterman'ın (Idury ve Waterman, 2009) öncü çalışması, daha eski uzun okuma teknolojileriyle bile assembly işlemi için farklı bir yaklaşım sunmuştu. Yeni yaklaşım, sabit uzunlukta alt kümelerden oluşan (bu alt kümeler  $k$ -mers olarak adlandırılır) bir grafik veri yapısı olan de Bruijn grafik veri yapısına dayanmaktadır. De Bruijn grafik yöntemi orijinal olarak kombinatoriyal matematik için geliştirilmiştir. DNA assembly uygulamasında de Bruijn grafiği, sekans setinde gözlenen her  $k$ -mer için bir düğüme (node) ve iki  $k$ -mer bir okumada bir arada gözlenirse düğümler arasında bir kenara (edge) sahiptir.

### 2.3.1. Genom hizalama işleminde de Bruijn grafik yönteminin kullanılması

Sanger sekanslamadan bu yana gelişen sekanslama teknolojileri, sekanslama işlemi sonucunda milyonlarca okuma üretebilmektedir. Eğer sekanslanan veriye ait referans sekans bulunursa bu sayede veri haritalanabilir. Ancak referans genom bulunmadığı durumlarda, parçalara ayrılmış veriyi tekrar birleştirerek uzun contigler oluşturmak gerekir. Buna parça hizalama (fragment assembly) denir ve biyoinformatik alanının zorlu konularından biridir. Parçaları birleştirip uzun contigler elde ederek nihai olarak tüm genomu hizalamak hedeflenir.

Sekanslama teknolojilerindeki gelişmeler ve üretilen verilerin miktarca çok büyük olması, parça hizalama algoritmaları için bazı zorluklar meydana getirmektedir. Örneğin, bakteri genomu büyüklüğündeki bir veriyi birleştirecek pek çok algoritma bulunurken, memeli genomu büyüklüğündeki veriyi analiz edecek araçlar kısıtlıdır. Bu araçlar da önemli miktarda kaynak kapasitesine ihtiyaç duymaktadır (depolama vs.) (Luo vd., 2012; Simpson vd., 2009). Örneğin, 20 Gb büyüklüğünde *Picea glauca* genomunun grafiğini oluşturmak ve sıkılaştırmak 38 saat sürmüş ve 4.3 TB hafızaya, 1380 CPU çekirdeğe ihtiyaç duymuştur (Biol vd., 2013). Yapılan farklı bir çalışmada 22 Gb büyüklüğünde *Pinus taeda* genomunu hizalamak için 3 ay ve 800 GB hafıza gerekmiştir (Zimin vd., 2014).

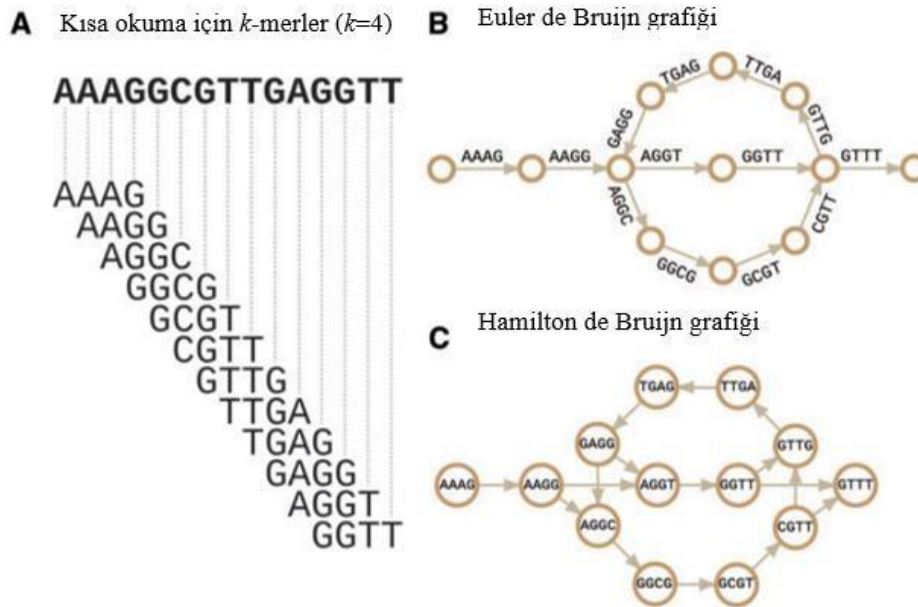
Kısa okuma parça hizalama algoritmalarının çoğu, okumalardaki bilgiye ulaşabilmek için de Bruijn grafiğini kullanır. Bir okuma setinde her bir  $k$ -mer, grafiğin bir köşesini oluşturur. Eğer bu  $k$ -merlerden iki tanesinin  $k - 1$  sayıda karakteri örtüşürse, bu iki  $k$ -mer arasında bir çizgi çizilir. Bir pozisyon kaydırarak ardışık kenarlardan oluşan ardışık  $k$ -merler ile bir assembly oluşturmak genomu yeniden bir araya getirmeye olanak sağlar (Chikhi vd., 2016).

Düğümleri ve kenarları ifade etme bakımından de Bruijn grafiği iki çeşittir: Hamilton ve Euler de Bruijn grafikleri (Compeau vd., 2011; El-Metwally vd., 2013).

Hamilton grafiğinde her bir  $k$ -mer bir düğümdür (node). Ardışık  $k$ -merler birbiri ile örtüştüğünde, örtüşen kısım grafiğe kenar (edge) olarak eklenir. Bir önceki  $k$ -merin  $k-1$  tane karakteri ile bir sonrakinin  $k-1$  tane karakteri örtüşüyorsa, bu iki  $k$ -mer birbirine kenar ile bağlanır. Bu şekilde grafik oluşur. Hamilton grafik yöntemi genellikle SOAPdenovo gibi assembler programlarında yaygın olarak kullanılmaktadır (Luo vd., 2012). Hamilton grafiğinde sekansları hizalamak için izlenen yolda tüm düğümler yalnızca bir kez ziyaret edilir. Fakat düğüm sayısı ya da genom karmaşıklığı arttıkça

grafığı çözmek için gereken hesaplama zamanı süresiz olarak farklılaşır. Çünkü genom büyüdükçe ve karmaşıklaştıkça scaffold veya tekrar tespit aşamaları çok komplike hale gelecektir. SOAPdenovo gibi bazı assembler programları bu karmaşıklığı basitleştirebilmişlerdir (Held ve Karp, 1962; Li vd., 2010; Thomason, 1989).

Euler grafik yönteminde ise düğümler ve kenarlar farklı şekilde temsil edilmektedir. Grafikte  $k$ -mer sekansları kenarları, örtüşen ardışık  $k-1$  merler ise düğümleri temsil etmektedir. Euler grafik yönteminde tüm kenarlar sadece bir kez ziyaret edilerek assembly işlemi tamamlanmış olur. Bu durum zaman bakımından daha verimli bir çözümdür. Genom assembly işleminde Euler grafiğinin Hamilton grafiğinden daha avantajlı olduğu söylenebilir (Cole vd., 2001; Compeau vd., 2011; Pevzner vd., 2001).



**Şekil 14.** De Bruijn grafik yönteminde okumaların kısa  $k$ -merlere bölünmesi. Hamilton yaklaşımında  $k$ -merler düğümleri temsil ederken, Euler yaklaşımında kenarlardır (Sohn ve Nam, 2018).

### 2.3.2. SOAPdenovo assembler

Yeni nesil sekanslama platformlarının kullanımının artmasıyla birlikte *de novo* genom hizalama çalışmalarının da sayısı artmaktadır. SOAPdenovo; okuma hatası düzeltme, de Bruijn grafik yapısı oluşturma, contig hizalama, paired-end okumalarının haritalanması, scaffold oluşturma ve boşlukların doldurulması modüllerinden oluşan bir assembly programıdır.

Yeni nesil sekanslama verilerinde assembly işleminden önce okumalardaki hataları tespit etme ve yeniden düzenleme önemli bir basamaktır (Salzberg vd., 2012; Zerbino ve Birney, 2008). SOAPdenovo programında hata düzeltme modülü kısa Illumina okumaları (35-50 bp) için tasarlanmıştır.

De Bruijn grafiği oluşturulurken çok sayıda benzersiz lineer  $k$ -mer, bağımsız olarak saklanmak yerine bir grup olarak birleştirilir. Bu sayede önemli miktarda bellek kullanımı gerektiren de Bruijn grafiği oluşturma basamağında önemli miktarda bellek tasarrufu sağlanmış olur (Ye vd., 2012).

De Bruijn grafiği yöntemine dayanan assembly analizlerinde önemli bir diğer etken de  $k$ -mer büyüklüğü seçimidir.  $k$ -mer uzunluğunun büyük olduğu durumlarda tekrar bölgelerinin çözünürlüğü artar fakat  $k$ -mer uzunluğunun küçük olduğu durumlarda ise düşük derinliğe sahip okumalar assembly edilir ve sekanslama hataları ortadan kaldırılır. Her iki durumdaki avantajlardan da faydalanabilmek için çoklu  $k$ -mer stratejisi benimsenmiştir (Peng vd., 2012). Önce küçük  $k$ -merler kullanılarak sekanslama hatalarını ortadan kaldırmak için bir grafik oluşturulur. Daha sonra daha büyük  $k$ -merler ile okumaları önceki grafiğe haritalayarak yeniden bir de Bruijn grafiği oluşturulur. Böylece büyük tekrar bölgeleri tespit edilir (Luo vd., 2012).

Scaffold oluşturma işleminde heterozigot contig çiftleri, contig derinliği ve local contig ilişkileri kullanılarak tespit edilir. Bu şartlar altında scaffold için daha yüksek derinliğe sahip heterozigot çiftler tutulur. Böylece heterozigotluğun scaffold uzunluğuna olan etkisi azaltılmış olunur. Kimerik scaffoldlar ise daha büyük ekleme boyutu olan kütüphanelerdeki bilgi kullanılarak düzeltilmiş olan daha küçük ekleme boyutuna sahip kütüphaneler kullanılarak oluşturulur (Luo vd., 2012).

### 2.3.3. Minia assembler

Minia ultra düşük hafıza gerektiren bir DNA sekans assembly programıdır. Girdi olarak genellikle Illumina ile oluşturulan kısa genomik sekans setini alır ve çıktı olarak birleştirilmiş sekansların bir contig setini verir. de Bruijn grafiğinin kısa ve öz bir temsiline dayanmaktadır. Hesaplama için diğer assembly programlarından önemli miktarda daha az kaynağa ihtiyaç duymaktadır. Hafıza kullanımını geliştirmek amacıyla basamaklı Bloom filtresi metodunu kullanmaktadır (Chikhi ve Rizk, 2013).

Hedef genomu assembly ederken her bir gigabaz başına ortalama 2 GB RAM kullanmaktadır. Örneğin bir insan genomunu hizalamak için 5.7 GB RAM kullanılmıştır (Chikhi ve Rizk, 2013).

Minia, masaüstü bir bilgisayarda tam bir insan genomu için contig oluşturabilen ilk yöntemdir (Chikhi ve Rizk, 2013b).

### 2.3.4. Kısa okuma *de novo* assembly analizlerinde karşılaşılan zorluklar

#### 2.3.4.1. Hesaplama zorlukları

İlk olarak assembly işlemi sırasında veya öncesinde sekanslama hatalarının düzeltilmesi gerekir. Bu hatalar düzeltilmezse bu durum contig ve scaffold oluşumlarının da hatalı olmasına sebep olacaktır. Sekanslamada meydana gelen hatalar kullanılan platforma ve kütüphane hazırlama metoduna göre çeşitlilik göstermektedir. Örneğin, Illumina platformları genellikle  $\leq 1\%$  rastgele okuma hataları üretir ve hatalar genellikle okumaların 3' kısımlarında meydana gelir (Alic vd., 2016; Liu vd., 2012; Mardis, 2008a).

İkinci olarak PCR, klonlama, GC eğilimi, sekanslama hataları ve kopya sayılarındaki çeşitlilikten kaynaklanan, eşit olmayan okuma derinliği sorununun çözülmesi gerekir (Sims vd., 2014; Xie ve Tammi, 2009). Eşit olmayan okuma derinliği genellikle assemblyde kırılmalara sebep olan boşlukların oluşmasına sebep olur. İdeal bir  $k$  uzunluğu seçmek bu sorunu çözmeye yardımcı olabilir (Sohn ve Nam, 2018).

Genomlarda tekrarlayan bölgelerin topolojik karmaşıklığı karşılaşılan zorluklardan bir diğeridir. Eğer okumalar tekrar bölgelerini kapsayacak kadar uzunsu bu sorun çözülür. Fakat Illumina platformları genellikle kısa okumalar üretir ve okumalar tekrar bölgesini kapsayacak uzunlukta değildir (Sohn ve Nam, 2018).

Dördüncü olarak yüksek hesaplama maliyeti gerektiren algoritmik karmaşıklık bulunmaktadır. Tüm genom *de novo* assembly çalışmaları, de Bruijn grafiğine bağlı olarak RAM, depolama ve hesaplama için gereken uzun zamana ihtiyaç duymaktadır. Bakteri genomu gibi küçük genomların *de novo* assembly işlemleri birkaç dakika sürerken memeli genomu gibi büyük genomlar için bu işlem birkaç gün veya hafta sürer ve onlarca veya yüzlerce GB RAM hafızasına ihtiyaç duyar (Sohn ve Nam, 2018).

### 2.3.4.2. Sekanslama hatalarının düzeltilmesi

Illumina platformu yüksek doğrulukta okumalar üretmesine rağmen meydana gelmiş olan hatalı okumalar aynı zamanda yanlış assembly sonucuna sebep olabilir. Sekanslama hataları daha sıklıkla sentromerler ve telomerlerin de dahil olduğu heterokromatin bölgelerde ya da yüksek tekrara sahip, GC-AT bakımından yoğun bölgelerde meydana gelmektedir. Bu yüzden assembly esnasında ya da daha öncesinde bu hatalar düzeltilmelidir.

Sekanslama hataları düzeltilirken önemli olan, hataları tanımlamak ve bu hataları heterozigot alellerden ayırt edebilmektir. Hata düzeltme yöntemleri üç kategoride incelenebilir:  $k$ -mer hesaplama, sonek ağacı veya sekans tabanlı yöntemler, çoklu sekans hizalama tabanlı yöntemler ve hibrit hata düzeltme yöntemleri dahil diğer yöntemler (Alic vd., 2016; Allam vd., 2015; Laehnemann vd., 2016).

Çoğu hata düzeltme aracı  $k$ -mer hesaplama yöntemini kullanır. Bu yöntemde düşük derinliğe sahip  $k$ -merler hatalı olarak kabul edilir. Sonek ağacı tabanlı yöntemlerde,  $k$ -mer sonek ağacında düşük frekanslı dallar hatalı  $k$ -merler olarak tespit edilir (Alic vd., 2016; Allam vd., 2015). Bu iki yöntemin tersine daha hassas olan çoklu sekans hizalama yönteminde, sekanslama hataları okumaların birbiriyle doğrudan hizalanması ile tespit edilir ve doğrudan düzeltilir.

Eğer bir genom, çoğaltma işlemi esnasında ideal bir şekilde çoğaltılırsa okuma parçaları bütün bölgelerde eşit bir dağılım gösterir ve  $k$ -mer derinlik histogramı normal bir dağılım oluşturur (Kelley vd., 2010; X. Li ve Waterman, 2003). Fakat sekanslama hataları meydana gelirse karşılık gelen  $k$ -mer derinliği üssel olarak azalan bir eğri oluşturur.

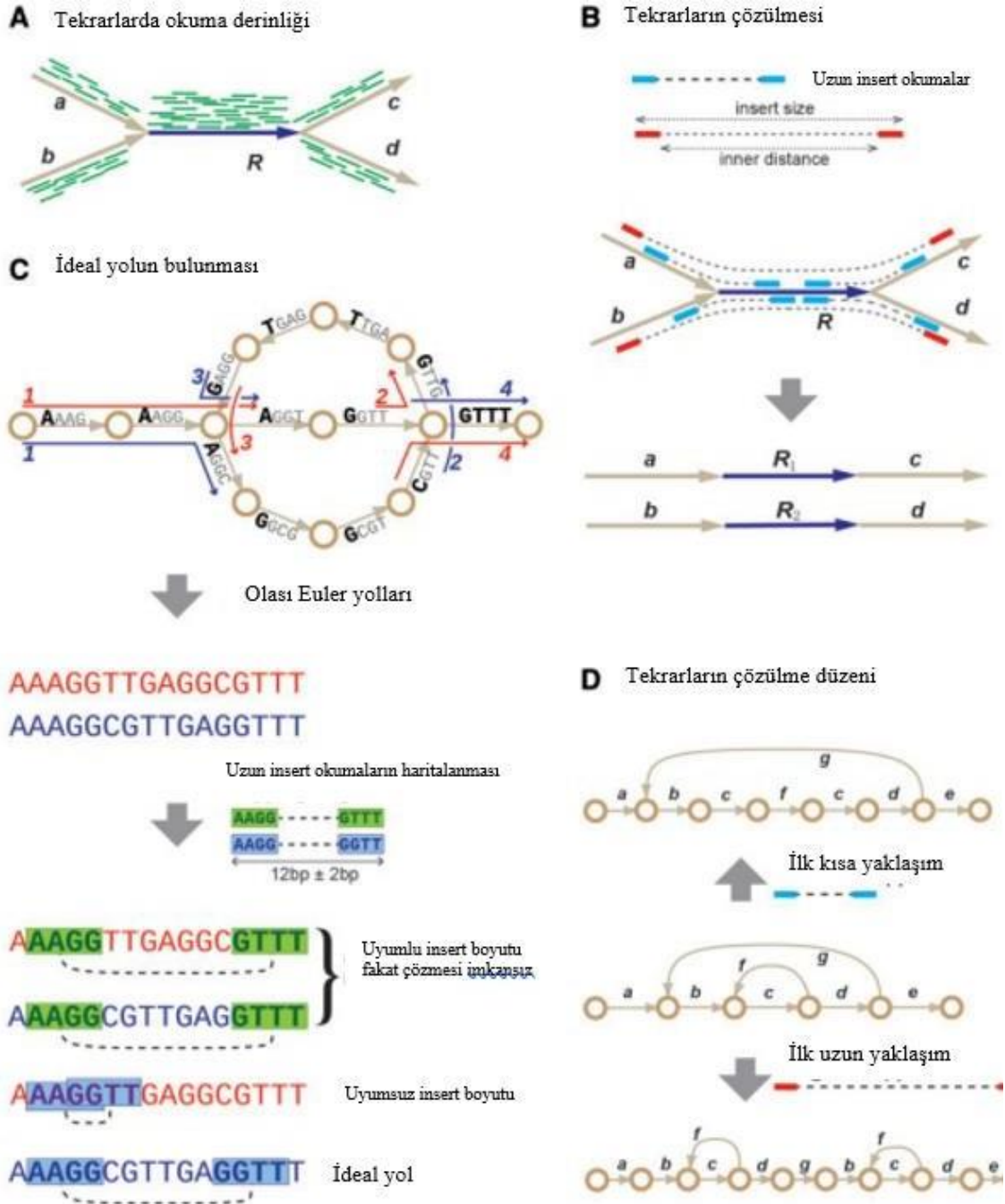
### 2.3.4.3. Tekrar bölgeleri

Karmaşık bir  $k$ -mer grafiğinde tüm düğümlerden geçen en kısa yolu tespit etmek genom assembly problemini çözmek için yeterli değildir. Bu sorun genomun sentromerlere, telomerlere ya da satelit bölgelerine yakın yerlerde pek çok tekrar eden bölgeye sahip olmasından kaynaklanmaktadır. Yoğun tekrar yapıları, oluşturulan assemblyde genellikle çok sayıda belirsiz yol ve boşluk oluşmasına sebep olur

(Chaisson vd., 2015; Cordaux ve Batzer, 2009; Koito ve Ikeda, 2013; Peng vd., 2012; Treangen ve Salzberg, 2011). Bazı tekrar problemleri okuma derinliği ile veya uzun okumalar ile çözülebilir (Treangen ve Salzberg, 2011).

Okuma derinliği istatistikleri sıklıkla tekrar eden yapıları çözmek için kullanılır. Okuma derinliği, tekrar bölgelerinde diğer bölgelerden daha yüksek olduğu için bölgenin kopya sayısı tahmin edilebilir (Şekil 15A). Ancak bu yöntem tekrar yapılarını çözmek için yeterli değildir çünkü okuma derinliğini etkileyen pek çok diğer etken, kopya sayısının yanlış tahmin edilmesine neden olabilir.

Tekrar eden yapıları çözmek için tamamlayıcı bir yöntem, her bir parçanın tekrar bölgesinin dışından itibaren eşleştiği okuma büyüklüğüdür (Şekil 15B). Eğer grafikte tekrar bölgelerinden dolayı birden fazla yol meydana gelirse bu durum okuma boyutu ile çözülebilir (Şekil 15C). Fakat geniş okuma uzunlukları dikkatli kullanılmadıkça çünkü geniş okuma bölgesi içeren bir assembly grafiği oluşturmak karmaşık bir işlemdir. Eğer grafikte uzun okuma ilk olarak uygulanırsa uzun aralıktaki tekrarlar çözümlenir ancak ekleme boyutundan daha kısa aralıktaki tekrar yapıları kalır. Bu gibi durumlarda grafiğin karmaşıklığı azalmak yerine duruma göre daha karmaşık bir hale gelebilir. Sonuç olarak öncelikle daha kısa okumalar uygulanmalı ardından tekrar eden yapıları uygun bir şekilde çözmek için daha uzun okumalar kullanılmalıdır. Bu sırayla ilerlemek tekrarlanan yapıların daha az karmaşık bir grafik oluşturmasını sağlar (Şekil 15D) (Sohn ve Nam, 2018).



Şekil 15. Tekrar bölgelerinin tespiti (Sohn ve Nam, 2018)

#### 2.3.4.4. Eşit olmayan okuma derinliği

Yeni nesil sekanslama teknolojilerindeki GC eğilimi, PCR sapmaları gibi bazı kısıtlamalar eşit olmayan okuma derinliğine sebep olur. Bu durum da taslak genomda boşlukların oluşmasına sebep olur. Veri eksikliğinden dolayı bu problem yeterince

çözülemediği. Daha derin okumalar oluşturmak bu sorunu çözebilir fakat bunu yapmak hesaplama süresini ve maliyetini önemli ölçüde arttıracaktır. Çoklu  $k$ -mer yaklaşımı ile bu sorun kısmen de olsa çözülmüştür (Gnerre vd., 2011; Luo vd., 2012). Kısa  $k$ -merler okuma derinliği az olan yerlerde okumaları birbirine bağlamak için kullanılırken, uzun  $k$ -merler diğer bölgelerde okumalar için kullanılır. Çoklu  $k$ -mer yaklaşımı, SOAPdenovo2 gibi bazı kısa okuma assembly programlarında uygulanmaktadır. Eşit olmayan okuma derinliği sorunu ile ilgili çoklu  $k$ -mer yaklaşımının dışında uygulanmakta olan bir strateji bulunmamaktadır. Düşük okuma derinliği sorunu kısmen de olsa boşluk doldurma aşamasında ele alınabilir.

### 2.3.4.5. Hesaplama maliyeti

#### RAM belleği

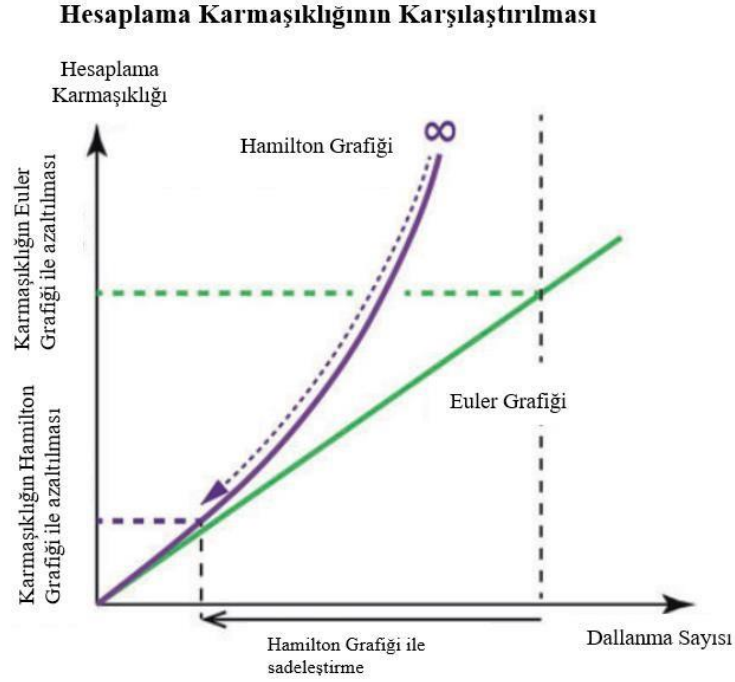
$k$ -mer yaklaşımı  $l$  uzunluktaki bir okumadan  $(l - k + 1)$  tane  $k$ -mer oluşturduğu için büyük genomları assembly ederken büyük RAM belleğine ihtiyaç duyulmaktadır. Örneğin, 3 Gb büyüklüğünde insan genomunda 75-merlik tablo oluşturmak için yaklaşık 0.5 terabayt belleğe ihtiyaç duyulmaktadır (Sohn ve Nam, 2018). Veri yapısını optimum hale getirmek veya algoritmalarındaki karmaşıklığı azaltmak için  $k$ -mer yoğunluğunu düşürmek (Luo vd., 2012; Ye vd., 2012), FM-index (Ferragina ve Manzini, 2000; Simpson ve Durbin, 2012), Bloom filtrelemesi (Chikhi ve Rizk, 2013b; X. Li ve Waterman, 2003; Melsted ve Pritchard, 2011; Pell vd., 2012; Simpson, 2014) veya süper okuma (Zimin vd., 2013) yaklaşımları gibi yaklaşımlar büyük genomların assembly işlemi için RAM belleğinden tasarruf etmeye olanak sağlayabilir.

#### Hesaplama süresi

Pek çok düğümün ve kenarın bulunduğu yüksek karmaşık yapıya sahip  $k$ -mer grafiğinde, Hamilton yaklaşımının Euler yaklaşımından özellikle büyük genomlar için daha yavaş olduğu bilinmektedir. Buna rağmen SOAPdenovo da dahil çoğu assembler programı Hamilton de Bruijn grafiği yöntemini kullanmaktadır.

$k$ -mer grafiğinin karmaşıklığı, assembler programlarının hızını anlamak için önemli bir faktördür. Teorik olarak, düşük karmaşık yapısına sahip bir grafik söz

konusu olduğunda Hamilton de Bruijn grafik yöntemi, Euler de Bruijn grafik yönteminden daha hızlı olabilir. Grafikteki dallanmaları temizlemek hesaplama karmaşıklığını, Hamilton yönteminde üssel olarak azaltırken, Euler yönteminde doğrusal bir şekilde azaltmaktadır (Şekil 16).



**Şekil 16.** Hamilton-Euler grafiklerinde hesaplama karmaşıklığının karşılaştırılması (Sohn ve Nam, 2018)

Ancak grafikten bazı yolların temizlenmesi, N50 istatistikleri ile temsil edilen doğruluk ya devamlılık değerinin kaçınılmaz olarak azaldığını göstermiştir. Bu durumda hesaplama süresi ile doğruluk arasında bir ters orantı olduğu söylenebilir (Sohn ve Nam, 2018).

#### 2.4. Genom Maskeleyme

Çoğu bitki genomu gen aileleri, satelit tekrarları, sentromerik tekrarlar ve transpozonlar gibi tekrar bölgelerinden oluşmaktadır. Genomik tekrarlar iki ana sınıfa ayrılmaktadır; ardışık tekrarlar ve transpozonlar. Ardışık tekrarlar mikrosatelit, minisatelit, satelit gibi art arda tekrar eden ve düşük karmaşıklığa sahip kısa motiflerdir. Transpozonlar ise ilk olarak McClintock tarafından keşfedilmiştir ve bunları “değişken genler” ya da “stabil olmayan genler” olarak tanımlamıştır (McCLINTOCK, 1950).

Transpozonlar yer deęiřtirmeleri için ihtiya duydukları gen bölgelerine sahiptirler ve genom boyunca daęılmış ok sayıda kopyaları bulunmaktadır (Girgis, 2015).

Transpozonlar genom boyunca “kes-yapıştır (DNA transpozonları)” veya “kopyala-yapıştır (RNA transpozonları)” mekanizmaları ile hareket ederler ve sıklıkla genomu büyüklük ve etkileřim bakımından etkilerler. Örneęin, mısır ve arpa genomunun 85%'ini transpozonlar oluřturmaktadır (Bourgeois ve Boissinot, 2019; Petersen vd., 2019; Ramakrishnan vd., 2022; Wicker vd., 2005).

Assembly iřlemi tamamlandıktan sonra genomdaki tekrar bölgeleri tespit edilir ve maskelenir. Tekrar bölgeleri, anotasyon esnasında gen tahmini iřlemine karmařık hale getireceęi ve hatalara sebep olacaęı için, anotasyondan önce tanımlanmalı ve maskelenmelidir. Açık okuma çereve (ORF) bölgeleri gen tahmin programlarında transpozon veya normal gen bölgelerini ayırt edemedięi için yanlış exon birleřimlerine sebep olur ve bu durum da hatalı bir gen anotasyonu oluřturur. Sonuç olarak bazı transpozonlar gen bölgeleri ieriyor olsalar dahi gen anotasyonu iřlemine dahil edilmezler (Hanada vd., 2009; Jiang, 2013).

Maskeleme terimi basite, tekrar olarak tanımlanan her bir nükleotidin ‘N’ olarak atanmasıdır. Soft masking yapan bazı araçlarda ise nükleotidler küçük “a,t,g ve c” olarak atanır. Bu sayede sekans hizalama ve gen tahmin analizlerinde bu nükleotidler tekrar bölgesi olarak tanınır. Tekrar bölgelerinin maskelenmedięi durumlarda, transpozonlardaki ORF bölgeleri buraların gen bölgeleri olarak tanımlanmasına sebep olur ve exon bölgelerine eklemeler gerekleřir. Bu da nihai gen anotasyon iřleminin hatalı olmasına sebep olur. İyi bir maskeleme iřlemi protein kodlayan genlerin doęru tahmin edilebilmesi için önem tařımaktadır (Yandell ve Ence, 2012).

Transpozonlar genoma özgüdürler ve evrimsel olarak genomik sekanstan daha hızlı evrimleřtikleri için tür spesifiktirler. Örneęin, pirin için hazırlanmış bir tekrar kütüphanesi mısır genomunu 25% maskeleyebilmektedir. Tür spesifik kütüphane oluřturmak bu oranı artırmaktadır (Jiang, 2013).

#### **2.4.1. RepeatMasker**

RepeatMasker programı tekrar bölgelerini tespit etmek ve anotasyon etmek ve bu nükleotid sekansları maskelemek için tasarlanmıştır. Tekrar bölgeleri anotata edilir ve “Ns, Xs veya küçük harfler (a.t.g.c)” ile deęiřtirilir. Ayrıca RepeatMasker

programının yeni versiyonları, tekrar bölgelerinde protein sekanslarını tespit edebilmektedir. Programı yürütebilmek için bir tekrar kütüphanesi dosyası seçilmelidir. Repbase (Jurka, 2000; Jurka vd., 2005) kütüphanesi, oluşturulmuş en geniş tekrar kütüphanelerinden biridir ve insan, kemirgen, zebra balığı, *Drosophila* ve *Arabidopsis thaliana* gibi canlılara ait tekrar kütüphanelerini içermektedir. Kütüphane dosyası bulunmayan organizmalar için, RECON (Bao ve Eddy, 2002) veya RepeatScout (Price vd., 2005) programları kullanılarak *ab initio* temelli kütüphane oluşturulabilir (Tarailo-Graovac ve Chen, 2009).

#### 2.4.2. Red

Red (REpeat Detector) programı, genomda DNA sekansındaki tekrar elementlerini tespit edebilmek için geliştirilmiş bir programdır. Red, tanımlanmış öğrenme verilerine bağlı olarak Hidden Markov Model (HMM) uygulamaktadır ve denetimli öğrenme modeline bir örnektir. Diğer programlarda yaygın olarak kullanılan yöntem, verilerin manuel olarak anotasyon edilmiş sekanslar ile eğitilmesine dayanmaktadır. Fakat Red aday tekrar bölgelerini; *k*-mer'leri sayarak, bir sinyal tanımlama tekniği kullanarak ve ikincil türev testi ile tanımlamaktadır. Ve bu aday bölgeler HMM yöntemi ile eğitilmektedir.

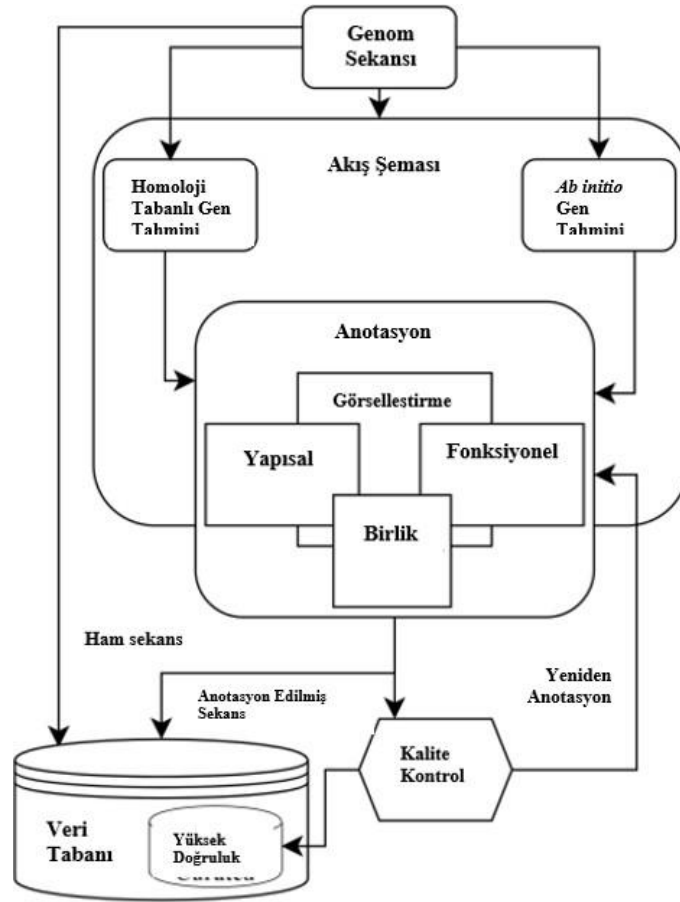
Sisteme sekanslar FASTA formatı ile yüklenmektedir ve genomu assembl edilmiş ya da edilmemiş sekanslar yüklenebilmektedir. Analiz sonucu çıktı olarak aday bölgelerin genomik lokasyonları ve genomda tespit edilmiş olan maskelenmiş tekrarlar verilmektedir. Analiz yapabilmek için dört modül içermektedir; skorlama modülü, etiketleme modülü, eğitme modülü ve tarama modülü (Girgis, 2015).

#### 2.5. Anotasyon

Sekanslama teknolojilerindeki gelişme ile birlikte DNA sekans verileri önemli ölçüde arttı (Mardis, 2008b). Zamanla, artan bu verilerin anlamlandırılması ve analiz edilmesi ihtiyacı oluştu. Anotasyon, temel olarak A, T, C ve G nükleotidlerinin sahip olduğu genetik bilgiyi anlamlandırmaya çalışma işlemidir. İyi bir anotasyon için assembly edilmiş genomun kalitesi oldukça önemlidir. Assembly analizini tamamlamak

ve assembly bütünlüğünü oluşturmak için pek çok istatistiksel metod geliştirilmiştir (Yandell ve Ence, 2012). Sekanslama teknolojilerinin gelişmesiyle birlikte düşük maliyetli ve yüksek kaliteli assembly oluşturmak günümüzde kolay ve ulaşılır hale gelmiştir.

Gen bölgelerini bulmak ve tanımlamak anotasyon işleminin önemli kısmıdır. Bu yüzden gen bulma işlemini başarılı ve doğru bir şekilde tamamlayacak yaklaşımlar gerekmektedir. Ve bu yaklaşımlar birbirinden bağımsız fakat birbirini tamamlayıcı çoklu analiz araçları ve yöntemlerinden oluşmaktadır (Brent, 2005; Mishra vd., 2020).



Şekil 17. Anotasyon akış şeması (Ejigu ve Jung, 2020)

### 2.5.1. Yapısal anotasyon

DNA'nın exon, intron, pomotor, transpozon bölgeleri şeklinde karakterize edilmesine yapısal anotasyon adı verilmektedir ve nihai amaç gen bölgelerinin tespit edilmesidir. "Gen" teriminin tanımı yapılacak olursa, fonksiyonel ürünler üretebilmek

için gerekli olan sekans bölgesi denilebilir (Spieth ve Lawson, 2006). Bir genin fonksiyonel ürünü protein veya RNA'dır. Protein üreten genlere protein kodlayan genler, protein kodlamayan diğer genlere ise kodlamayan genler denilmektedir. Kodlamayan genler ribozomal RNA (rRNA), transfer RNA (tRNA), microRNA (miRNA), small nuclear RNA (snRNA ve snoRNA) ve long noncoding RNA (lncRNA) gibi RNA molekülleri üretmektedir (Ejigu ve Jung, 2020).

Yapısal anotasyon aynı zamanda pseudogenleri de tanımlamaktadır. Bu genler önceleri işlevsiz ve çıkmaz sokak olarak kabul görmüş olsa da günümüzde gen regülasyonunda görev aldıkları bilinmektedir (Xiao vd., 2016).

### **2.5.1.1. Tekrar bölgeleri**

Yapısal anotasyonun ilk adımı tekrarların maskelenmesidir. Ökaryotik genomlar milyonlarca tekrara sahiptir. Örneğin, insan genomunun yaklaşık 2/3'si tekrarlardan oluşmaktadır (de Koning vd., 2011). Tekrar sekansları rastgele veya bitişik şekilde konumlanmış olabilir ve genellikle sentromerlerde bulunmaktadır (Barra ve Fachinetti, 2018). Alternatif olarak transpozonların farklı formlarında, kısa ve uzun elementler (SINEs ve LINEs) ve DNA transpozonları gibi, genoma dağılmış olabilirler (Bourque vd., 2018). Tekrar tespitindeki gelişmelere rağmen tekrar elementlerinin temel özelliklerini tanımlamak hala zor bir durum.

Tekrar maskeleyme araçları daha önceden tanımlanmış tekrarların listesine dayanan bir veri tabanı ile çalışmaktadır ve RepeatMasker buna güzel bir örnektir.

Yapısal anotasyonda maskelemenin ardından, gen tanımlamadan önce gelen ikinci adım transkript ve protein temelli hizalamadır. Bu hizalama işlemi için BLAST (Altschul vd., 1990) veya BLAT (Kent, 2002) araçları kullanılabilir. RNA-seq temelli hizalama işlemi için ise TopHat (Kim vd., 2013) veya HISAT (Kim vd., 2015) gibi programlar kullanılabilir. Fakat bu işlemler zorunlu değildir.

### **2.5.1.2. Genlerin tahmin edilmesi**

Protein kodlayan genlerin ve diğer regülatör elementlerin tanımlanması gen anotasyonunun merkezinde yer alan bir aşamadır. Gen tahmini karmaşık bir

süreçtir özellikle de ökaryotik DNA için (Yandell ve Ence, 2012). Kodlanmayan bölgeler olan ve exonların arasında bulunan intronların değişen boyutları ve alternatif splaysing mekanizmasının çeşitliliği gen yapılarının tahmin edilmesini zor bir hale getirmektedir. Gen tahmini için pek çok program mevcuttur ve bunları üç kategoriye ayırmak mümkündür; *ab initio* yöntemler, homoloji tabanlı yöntemler ve kombine yöntemler.

*Ab initio* yöntemler istatistiksel yöntemlere dayanır ve Hidden Markov Model (HMM) gibi yöntemlerle promotör bölgeleri, kodlayan veya kodlanmayan gen bölgeleri ve intron-exon bölgeleri gibi bölgeleri tespit etmeye çalışır.

İkinci bir yöntem olan homoloji tabanlı yöntemlerde genom sekansı, expressed sequence tags (EST), complementary DNA (cDNA) veya protein sekansları ile hizalanarak bunlar üzerindeki benzerlikler ile genler tahmin edilmeye çalışılır.

Üçüncü yöntemde ise *ab initio* ve homoloji tabanlı yöntemler kombine edilerek birlikte kullanılır.

Ayrıca gen tahmin programları alternatif splaysing bölgelerini tahmin edebilmelidir çünkü alternatif splaysing gen ekspresyonunun düzenlenmesinde, transkriptom ve proteom çeşitliliğinde başrole sahiptir (Modrek ve Lee, 2002). Bu bölgeleri tespit etmek için programlar çeşitli yöntemler kullanmaktadır. Sekanslanan genomlarda intronların yaklaşık 99%'u GT nükleotidleri ile başlayıp AG nükleotidleri ile bitmektedir. Bu durum çoğu gen tahmin sisteminde splays bölgesini tespit etmek için zorunlu olarak kullanılmaktadır (Brent ve Guigó, 2004).

#### 2.5.1.2.1. AUGUSTUS

Bir gen tahmin programı olan AUGUSTUS, hizalama tabanlı yöntemlerden gelen bilgileri, *ab initio* gen tahmini ile birleştirebilmektedir (Stanke vd., 2006, 2008; Stanke ve Waack, 2003).

Genom anotasyon işleminin ilerleyen aşamalarında proteinlerin ailelere ve alt ailelere sınıflandırılması önemli bir görevdir. Bir protein ailesi üyeliği, gen tahmini için önemli bir aşamadır. Bu bilgiye protein aile veri tabanlarından, InterPro (Hunter vd., 2009) gibi, kolaylıkla ulaşılabilir.

Protein ailesi tespiti için kullanılan kaynaklar protein sorgu (query) sekanslarını sınıflandırmak için kendi yöntemlerini sunarlar, fakat bu yöntemler ökaryotik genomik

sekanslara doğrudan uygulanamazlar. Ökaryotik sekanslarda öncelikli olarak gen yapısı belirlenmeli ve protein kodlayan genler tespit edilmelidir. AUGUSTUS, mevcut *ab initio* modelini homoloji tabanlı protein sorgulama yöntemi ile birleştiren ve “hibrit” olarak adlandırılan yöntemi kullanmaktadır.

Bir protein modelinin gen tahmini ile birleştirilmesi sayesinde belirli bir protein ailesinin üyelerinin tanımlanması ve gen tahmininin doğruluğunun artması sağlanmıştır (Keller vd., 2011).

### 2.5.1.3. Yapısal anotasyon için kullanılan veri tabanları

Anotasyon işlemlerinde kanıta dayalı tahminlerde kullanılması için destekleyici veriye ihtiyaç duyulur. Günümüzde, yüksek miktarda EST ve cDNA sekansı ulaşılabilir olduğu için homoloji tabanlı yöntemler genom anotasyonunda merkezi bir rol oynamaktadır (Mathé vd., 2002). Homoloji tabanlı yöntemler DNA, RNA veya protein sekans verilerinin hizalanmasına dayanmaktadır. Bu veriler biyolojik veri tabanlarından kolayca elde edilebilmektedir. Diğer bir yöntem olan *ab initio* yöntemde ise genler ve gen yapıları matematiksel modellemeler kullanılarak tanımlanmaktadır. Bunun için *ab initio* gen tahmin eden programlar yüksek kalitede gen modelleriyle veya kodon sıklığı, intron-exon uzunluk dağılımı gibi organizmaya özgü genomik özelliklerle eğitilmelidir (Y. Wang vd., 2015). Dahası *ab initio* yöntemler tahmin doğruluğunu arttırmak için EST, RNA seq ve protein verilerine ihtiyaç duymaktadır. Veri tabanları bu tür verileri kolayca sağlayabilmektedir.

Nükleotid ve proteinlerin sekansları ve yapılarına GenBank (Sayers vd., 2019), European Nucleotide Archive (ENA) (Brooksbank vd., 2014), National Center for Biotechnology Information (NCBI) (Geer vd., 2010) gibi kamuya açık veri tabanlarından kolayca ulaşılabilir. UniProt (Bateman, 2019) ise bilimsel topluma yüksek kalitede ve ücretsiz erişim imkânı sunan, fonksiyonel olarak ilişkilendirilmiş protein sekanslarını sağlayan bir veri tabanıdır. Protein anotasyonu için diğer önemli bir veri tabanı olan InterPro (Mitchell vd., 2019) ise protein aileleri, domainleri ve bağlanma bölgeleri, aktif bölgeleri, korunmuş bölgeler ve tekrarlar gibi önemli bölgelerin bilgisini sağlamaktadır.

Ayrıca NONCODE (Fang vd., 2018), Pseudogene.org (Karro vd., 2007), Dfam (Hubley vd., 2016) ve miRbase (Kozomara vd., 2019) gibi kaynaklardan kodlanmayan RNA, pseudogenler, transpozonlar ve micraRNA verilerine ulaşılabilir.

## 2.5.2. Fonksiyonel anotasyon

Biyolojik bilgi ile yapısal anotasyon sonucu tanımlanmış gen ve proteinlerin ilişkilendirilmesi, fonksiyonel anotasyon olarak adlandırılmaktadır. Protein kodlayan genler, fonksiyonel anotasyonun en önemli odak noktasıyken kodlanmayan genlerin ve translate edilmeyen transkriptlerin öneminin keşfedilmesi ile birlikte bu odak değişmiştir (Mudge ve Harrow, 2016).

Fonksiyonel anotasyon, temel olarak bir sekansın benzerini tanımladıktan sonra bir genle ilişkilendirilen fonksiyonel özelliklerin belirlenmesi sürecini içerir, bu da BLAST gibi araçlarla gerçekleştirilebilir.

Fonksiyonel anotasyon aynı zamanda genlerdeki varyasyon bilgisine erişmek için de kullanılır. Genomik varyantların anotasyonu, sekansa dayalı genomik analizlerin çok önemli ve karmaşık bir parçasıdır. Varyant anotasyonunun amacı, varyantların fonksiyonel etkilerinin tanımlanması ve öncelikli hale gelmesidir (Cutting, 2014). Genetik varyasyonların fenotip üzerindeki moleküler etkisi, genomik sekansların yapısal ve fonksiyonel bilgisinin varyantlarla ilişkilendirilmesi ile anlaşılabilir (Butkiewicz ve Bush, 2016). Gen varyasyonu, aynı biyolojik türlere ait bireylerin genomik DNA'larında meydana gelen ve tek nükleotid polimorfizmi (SNP) olarak adlandırılan tek nükleotidlik değişimler veya insersiyon, delesyon, translokasyon ve inversiyon gibi yapısal yeniden düzenlemeler sonucu meydana gelmektedir. İnsersiyon ve delesyon, kopya sayısı varyasyonları (CNVs) olarak bilinen varyasyonlara sebep olurlar (Pavlopoulos vd., 2013).

### 2.5.2.1. Otomatik fonksiyonel anotasyon

Manuel anotasyon halen altın standart olarak kabul görse de bu yaklaşımı uygulamak zordur. Yeni nesil sekanslama teknolojilerinin gelişmesi ile birlikte artan genomik verilerle başa çıkabilmek için otomatik anotasyon yöntemlerinin kullanılması

gerekmektedir. Otomatik fonksiyon tahmini için protein veri tabanının yüksek skorlu hizalamalar için arandığı BLAST gibi lokal hizalama araçları kullanılabilir. Burada, bilinen sonuç sekansına bağlı olarak bilinmeyen sorgu sekansına fonksiyon atanır. BLAST benzeri araçlar üzerinden fonksiyon transferi, tek bir atadan evrimleşen ve benzer sekanslara sahip proteinlerin fonksiyonlarını korudukları varsayımına dayanmaktadır. Yani kullanılan araçlar, sekanslar arasındaki ortolog ve paralog ilişkileri tanımlayarak evrimsel ilişkileri tespit ederler.

Ortologlar, karşılaştırılan genomların son ortak atasında bulunan tek bir atasal genden köken alan genlerken; paraloglar, aynı genomda kopyaların çoğalması ile oluşan genlerdir (Koonin, 2005).

### **2.5.2.2. Fonksiyonel anotasyon için kullanılan veri tabanları**

Gen Ontolojisi (GO) kaynağı, genlerin fonksiyonu için en kapsayıcı ve geniş kullanımlı bilgi tabanıdır. GO gen fonksiyonu için üç bakış açısını kapsamaktadır: bir gen ürününün moleküler seviyede aktivitesini belirten moleküler fonksiyon, gen ürününün lokasyonunu bildiren hücresel bileşen ve genin kullanım fonksiyonunu belirten biyolojik proses. Protein ve RNA gibi gen ürünleri, biyolojik kavramların kapsamlı olarak anlaşılabilmesi için tutarlı bir şekilde tanımlanmalıdır. GO konsorsiyumu bu ihtiyacı karşılamak için ontoloji standartlarını geliştirmekte ve sürdürmekte, gen ürünlerini anotasyon etmekte ve bunun için araçlar geliştirmektedir (Blake vd., 2015).

KEGG (Kyoto Encyclopedia of Gene and Genomes) (Kanehisa ve Goto, 2000), GENES ve PATHWAY veri tabanlarında saklanan üst düzey fonksiyonel bilgi ile genomik veri arasında bağlantı kurma görevi görür. Yüksek seviyede fonksiyonları anlama ve genomik seviyede bilgiden faydalanarak biyolojik sistemin faydalarını anlamayı sağlar (Kanehisa vd., 2016).

### 2.5.3. Anotasyon yöntemleri

#### 2.5.3.1. *Ab initio* anotasyon

*Ab initio* anotasyon yöntemi, verinin bir algoritma veya model oluşturmak için eğitilmesine dayanan *ab initio* gen tahminine dayanmaktadır. Yöntemde, sorgulanan genomik sekansın istatistiksel analizler ve diğer gen sinyalleri ile tahmini gerçekleştirilir.

*Ab initio* anotasyon yöntemlerini RNA-seq verisinden elde edilen transkriptom verileriyle birleştirmek de özellikle gelişmiş ökaryotik canlılar için yaygın bir kullanım haline gelmiştir (Ejigu ve Jung, 2020).

#### 2.5.3.2. Homoloji tabanlı anotasyon

Moleküler evrim prensibine göre, genomda fonksiyonel olarak önemli olan bölgelerin evrimi diğer hücrel moleküler bölgelere göre daha yavaş gerçekleşmektedir. Böylece özellikle yakın akrabalık ilişkisi bulunan türlerde, hayatta kalmak ve hayati fonksiyonları yerine getirebilmek için gereken gen sekansları korunmuş durumdadır (Clark vd., 2019). Bu durum homoloji tabanlı anotasyonlarda anlamlı eşleşmeler tanımlayarak genlerin tahmin ve anotasyon edilmesi için kullanılmaktadır. Homoloji tabanlı anotasyonda genleri tahmin etmek için kodlayan sekanslar kullanılmaktadır. Benzer sekans bölgelerinin homolog proteinler kodladığı varsayılarak genellikle protein sekansları bazen de mRNA, cDNA, EST gibi transkriptler kullanılmaktadır.

#### 2.5.3.3. MAKER

Kullanımı kolay bir anotasyon aracı ilk olarak kullanması ve yüklemesi kolay olmalı ve minimum kaynağa ihtiyaç duymalıdır. İkinci olarak hesaplama ve anotasyon motorunu sağlayabilmeli; tekrar bölgelerini tespit edebilmeli, EST ve proteinleri genoma hizalayabilmeli ve alternatif splysing, UTR bölgeleri özellik bakımından zengin verileri anotasyona dahil edebilmelidir. Üçüncü olarak her bir genom farklıdır bu

yüzden anotasyon aracı kolay yapılandırılabilir ve eğitilebilir olmalıdır. Eğer bunu sağlayamazsa anotasyon süreci sınırlı olarak gerçekleştirilir.

Diğer önemli bir özellik ise çıktı formatının hem kapsayıcı hem de veri tabanı için hazır olmasıdır. Bu görev, Genel Model Organizma Veri tabanı (GMOD) projesi tarafından kolaylaştırılmıştır; bu proje genel bir genom veri tabanı şeması ve genom görselleştirme araçları sunmaktadır. Ancak GMOD, veri tabanı içeriği üretmez, bir anotasyon aracı tarafından üretilen verileri uygun hale getirir. Bu yüzden GMOD tarafından sağlanan özelliklerden faydalanabilmek için, anotasyon araçları veri çıktılarını GMOD uyumlu Generi Feature Format (GFF3) formatında yazmalıdır. GFF3 formatında dosya yazmak karmaşık bir işlemdir. Bu dosyalar EST ve protein hizalanmalarını, tekrarları ve gen tahmin bilgilerini içermelidir. Ayrıca anotasyon olmayan EST ve protein hizalanmalarını göstermelidir böylece yanlış negatifler tespit edilebilir.

Kullanımı kolay anotasyon aracı olarak tasarlanan MAKER, yukarıda belirtilmiş olan kriterlerin tümü göz önünde bulundurularak tasarlanmıştır. Gelişmekte olan genom projelerine, bağımsız olarak protein kodlayan genlerin anotasyonunu yapma ve bir GMOD veri tabanı oluşturma olanağı sağlamak amaçlanmıştır. MAKER tekrarları tespit eder, EST ve proteinleri genoma hizalar, gen bölgelerini tahmin eder ve bu verileri protein kodlayan gen anotasyonuna entegre eder. Dahası, çıktıları herhangi bir ek işleme gerek duymadan doğrudan GMOD tarayıcılarına ve veri tabanlarına yüklenebilir (Cantarel vd., 2008).

### 3. MATERYAL VE YÖNTEM

#### 3.1. *Salvia officinalis* Bitkisinin Genomik DNA İzolasyonu

Genomu sekanslanmak üzere adaçayı bitkisi Selçuk Üniversitesi, Ziraat Fakültesi, Bahçe Bitkileri Bölümü'nden temin edilmiştir. Adaçayı genotiplerine ait yaprak doku örnekleri, sıvı azot ile toz haline gelene kadar ezilerek ve 200 mg doku örneğine 800 µL CTAB ekstraksiyon çözeltisi 2% CTAB, 100mM Tris-HCl (pH8.0), 20mM EDTA (pH8.0), 1.4 M NaCl, 1% PVP-40) eklenmiştir. Karışıma 100 µL merkaptto etanol eklenerek, homojenize edilmiştir. Örnek, hücre duvarlarının yıkılmasını amacı ile 65 °C'de 1 saat süreyle inkübe edilerek, ardından hacimce (25:24:1) oranında fenol:kloroform:izoamil alkol örneğe eklenerek, oda sıcaklığında 10 dakika santrifüjlenmiştir. Üst faz, temiz bir 2 ml'lik deney tüpüne aktarılarak, 600 µL, hacimce 24:1 oranında hazırlanmış kloroform:izoamil alkol örneğe eklenmiştir. Oda sıcaklığında 10 dakika süreyle gerçekleştirilen santrifüj işleminin ardından, üst faz 1.5 ml'lik temiz bir tüpe aktarılarak, kendi hacminin altıda biri oranında 100%'lük izopropanol ile karıştırılmıştır. DNA peletinin elde edilmesi için deney tüpü oda sıcaklığında 30 dakika inkübe edilmiştir. Pelletin (DNA çökeltisi) izopropanolden temizlenmesi amacı ile inkübasyonun ardından örnek 10 dakika oda sıcaklığında santrifüjlenerek, üst faz atılmıştır. Deney tüpünün çeperinde elde edilmiş olan DNA peletinin üzerine 70%'lik etanol eklenerek yıkama yapılarak, yıkamanın ardından oda sıcaklığında 5 dakika süreyle santrifüj işlemi gerçekleştirilmiştir. Etanol fazı santrifüjün ardından atılarak, 100 µL, sterilize distile su, DNA'nın süspanse edilmesi amacı ile eklenmiştir. DNA ekstraksiyonu sırasında, DNA ile birlikte izole edilmiş olması muhtemel olan RNA'nın degradasyonu, ekstrakte edilmiş olan DNA örneklerine 1 µL RNase enzimi katılıp, tüplerin 37°C'de 30 dakika inkübe edilmesi ile sağlanmıştır. Ekstrakte edilmiş olan DNA örneklerinin konsantrasyonu ve saflığı, Qubit cihazı ile ölçülmüştür. Elde edilen DNA örnekleri, ileriki analizlerde kullanılmak üzere -20°C'de muhafaza edilmiştir.

### 3.2. *Salvia officinalis* Genomunun Sekanslanması ve Scaffold Oluşumu

*Salvia officinalis*'in izole edilmiş olan DNA'sı Illumina Novaseq 6000 pair-end sistemi ile sekanslanmıştır. Yaklaşık 450 bp parçalardan oluşan bir sekanslama kütüphanesi oluşturmak için TruSeq DNA PCR-Free kit kullanılmıştır. Kit kullanılırken TruSeq DNA PCR-Free Sample Preparation kılavuzu takip edilmiştir. Sekanslama sonucunda 151 bp büyüklüğünde okumalar elde edilmiştir. Novaseq 6000 adaptörleri Trimmomatic (Bolger vd., 2014) v0.38 kullanılarak trim edilmiştir. Trim edilen sekansların KmerGenie (Chikhi ve Medvedev, 2014) programı ile assembly uygulamalarında yapılması gereken öncül bir analiz olan *k*-mer analizi gerçekleştirilmiştir. *k*-mer analizi gerçekleştirilen okumalar, Minia assembly algorithm v3 (Chikhi ve Rizk, 2013b) programı ile contigler olarak assembly edilmiştir. Sonrasında bu contig assembly, biyoinformatik araçlarla ilerleyen analizlerin gerçekleştirilebilmesi için Galaxy (<https://usegalaxy.eu/>) platformuna yüklenmiştir. RagTag software toolset v2.1.0 (Alonge vd., 2022) ile contiglerden scaffold elde edilmiştir. Scaffold oluşumunda, contiglerin yerleşimi ve düzeni için adaçayı kromozomlarının sekansları referans olarak kullanılmıştır. Adaçayı kromozom sekanslarına <https://ngdc.cncb.ac.cn/gwh> adresinden GWHBJVP00000000 ID numarası ile ulaşılmıştır. Scaffold haline gelmiş olan assembly için RepeatMasker v4.1.2 (Smit vd., 2015) ile yumuşak maskeleyme işlemi gerçekleştirilmiştir.

### 3.3. Gen İçeriğinin Değerlendirilmesi

Elde edilen assembly gen içeriği bütünlüğü BUSCO (Benchmarking Universal Single-Copy Orthologs) v5.3.2 (Simão vd., 2015) ile analiz edilmiştir. BUSCO yazılımı bu analiz için embryophyta\_odb10 (Creation date: 2020-09-10, number of genomes: 50, number of BUSCOs: 1614) veri seti kullanılarak yürütülmüştür.

### 3.4. Yapısal ve Fonksiyonel Gen Anotasyonu

Scaffoldda protein kodladığı varsayılan gen yapıları AUGUSTUS v3.4.0 (Stanke vd., 2008) ile tespit edilmiştir. Karşıt zincirlerde bulunan gen yapıları, genlerin overlap olmasını önlemek amacıyla bağımsız olarak tahmin edilmiştir. Varsayılan gen yapılarında tahmin edilen gen modellerinin ve atanan fonksiyonların doğrulanması için taxonomic scope 71274 (asterids) uygulanarak eggNOG (evolutionary genealogy of genes: Nonsupervised Orthologous Groups) Mapper v5.0.2 (Huerta-Cepas vd., 2017) aracı kullanılmıştır. Gen yapılarının çıktısı ilerleyen analizlerde terpenoid biyosentezinde görev alan genlerin tespit edilmesi için kullanılmıştır. Homoloji analizi, *Salvia* türünde terpenoid metabolizmasında kodlayan sekanslara (GeneBank Accession 869088 – 894017) erişilerek gerçekleştirilmiştir. Analiz için E-value eşik değeri 1E-10 belirlenerek blastx algoritması uygulanmıştır. InterProScan v95.0 (Paysan-Lafosse vd., 2023) aracı varsayılan homologların fonksiyonel sınıflandırması ve doğrulaması için kullanılmıştır. Tanımlanan gen setinin fiziksel lokasyonlarının görsel grafikleri MapChart v2.32 (Voorrips, 2002) kullanılarak oluşturulmuştur. Adaçayına ait RNA-Seq verileri kanıt temmeli anotasyona da kullanılmıştır.

Adaçayına ait protein kodladığı varsayılan gen yapıları ayrıca fonksiyonel anotasyon ve genomiks aracı olan OmicsBox (Blast2GO) (<https://www.blast2go.com/>) yazılımı ile homolojiye dayalı olarak analiz edilmiştir. *Salvia hispanica* protein sekansları UniProt (<https://www.uniprot.org/>) veri tabanından indirilerek, *S. hispanica* local veri tabanı oluşturulmuştur. Illumina NovaSeq 6000 ile sekanslanmış olan *Salvia officinalis*'in varsayılan gen yapıları AUGUSTUS programı ile çıktı olarak alınmıştır. Blastp algoritması kullanılarak peptit hizalaması gerçekleştirilmiştir. OmicsBox 3.0.29 aracı ile GO mapping ve fonksiyonel anotasyon analizleri gerçekleştirilmiştir. OmicsBox tarafından sağlanan parametreler kullanılarak sorgulanan adaçayı sekansları ile oluşturulan yerel veri tabanı arasında önemli eşleşmeler tespit edilmiştir.

### 3.5. Protein Kodlayan Genlerden SSR Markör Geliştirme

Protein kodladığı varsayılan gen yapıları GMATA (Genome-wide Microsatellite Analyzing Tool Package) (X. Wang ve Wang, 2016) programına yüklenerek basit sekans tekrarlarının (SSR) tespiti, primer tasarımı ve markör haritalama analizleri

gerçekleştirilmiştir. GMATA programının çalışması için gerekli olan Perl, R ve Java programları yüklenerek gereken çalışma ortamı sağlanmıştır. Basit sekans tekrarlarının tespiti için AUGUSTUS programından çıktı olarak alınan protein kodladığı varsayılan gen yapıları, yazılımın ‘SSR identification’ modülüne FASTA formatında yüklenmiştir. SSR tespiti için; Min-length (nt): 2, Max-length (nt): 6, Min. repeat times: 5 parametreleri kullanılmıştır. Uygulanan parametreler sonucu çıktı olarak alınan (.ssr) dosyası tekrar birimlerinin sayısını, motifini ve tespit edilen SSR lokuslarının mutlak pozisyonlarını belirtmektedir.

### **3.5. *S. officinalis* Exomunda Exonic Tek Nükleotid Polimorfizmlerinin (SNPs) ve Terpenoid Biyosentezinde Görev Alan Genlerin Tespiti**

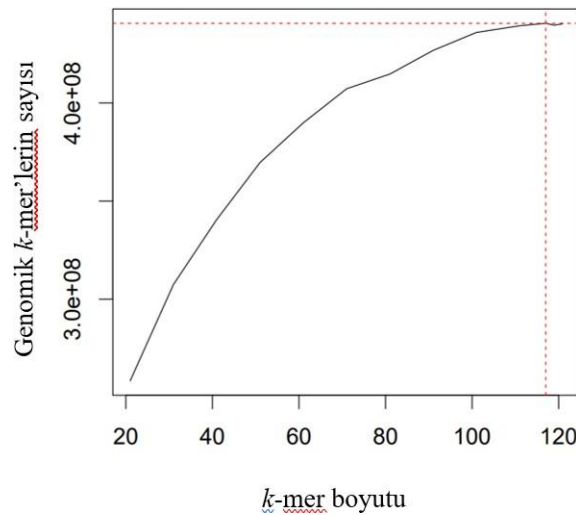
Varyant tespiti için nucmer align (Marçais vd., 2018) metodunu içeren DNAdiff wrapper kullanılmıştır. *S. officinalis*'in tüm exomu ve terpenoid biyosentezinde görev aldığı varsayılan genlerin kodlayan sekansları, yapısal ve fonksiyonel anotasyon süreci boyunca adaçayının var olan genom sekansları ile kıyaslanarak tanımlanmıştır. Adaçayının genom sekanslarına <https://ngdc.cncb.ac.cn/gwh> adresinden, GWHBJVP00000000 ID numarası ile erişilmiştir. Tanımlanan tek nükleotid varyantları lokus için en az 20 bp yakınında ve non-variable, değişken nükleotidin her iki tarafından kusursuz şekilde hizalanmış olacak şekilde filtrelenmiştir. Tek nükleotid varyasyonlarının lokasyonlarının görselleştirilmesi için peptid domainleri ve potansiyel amino asit çeşitleri değerlendirilmiştir. Domain araştırması, peptid sekanslarındaki kodlanan ORFs (open reading frames) bölgelerinde bulunan, terpenoid biyosentezinde görev aldığı varsayılan gen bölgelerinde InterProScan (Paysan-Lafosse vd., 2023) aracı kullanılarak yapılmıştır.

#### 4. ARAŞTIRMA SONUÇLARI VE TARTIŞMA

Bir genomun *de novo* sekanslanması bitki ıslah ve fonksiyonel genomik çalışmalarında verilerin geliştirilmesi ve analiz edilmesi için önemli bir gelişmedir. Bu tez kapsamında adaçayı genomunun *de novo* sekanslanması ve varyant analizi için referans genom verisi kullanılması, adaçayı genomunda bulunan terpenoid biyosentezinde görev alan kodlanan sekanslardaki tek nükleotid polimorfizmlerini tespit etmek ve lokasyonlarını belirlemek için gerçekleştirilmiştir. Bu sayede geliştirilen markörlerin terpenoid biyosentezinde genotipe bağlı olarak meydana gelen ve fenotipe etkisi olan farklılıkların anlaşılması için yürütülecek olan ilerleyen çalışmalarda potansiyel fonksiyonel markör olarak kullanılması amaçlanmıştır.

##### 4.1. Sekanslama Sonucu ve Assembly

Paired-end sekanslama sonucu olarak 118.8 Gb ham data üretilmiştir. Okumalar GenBank veri tabanında BioProject PRJNA995944 ID numarası ile SRA (sequence read archive) dosyası olarak saklanmaktadır. Ham veriden adaptör sekanslar trim edilmiş ve okumalar phred kalite skoru 30 (Q30) olmak üzere filtrelenmiştir. Trim edilme ve filtreleme işlemlerinin ardından ham verinin 89.2%'si olan 106 Gb büyüklüğünde veri elde edilmiştir. *k*-mer analizi sonucu en uygun *k*-mer değeri  $k = 117$  olarak belirlenmiştir (şekil 18).



Şekil 18. K-mer analizi sonucu *k*-mer'lerin sayısı ve boyutu

Adaptör sekansları trim edilen ve Q30 filtrelemesinden geçen okumalar ilk olarak contigler olarak assembly edilmiştir. Sonrasında elde edilen contigler referans genom kullanılarak yedi scaffold haline gelmiştir (Çizelge 1).

**Çizelge 1.** Assembly istatistikleri

<b>Assembly istatistikleri</b>	
Scaffold sayısı	7
Pseudo molekül 1 (bp)	88,498,997
Pseudo molekül 2 (bp)	77,009,490
Pseudo molekül 3 (bp)	60,691,335
Pseudo molekül 4 (bp)	53,566,787
Pseudo molekül 5 (bp)	52,731,730
Pseudo molekül 6 (bp)	54,576,965
Pseudo molekül 7 (bp)	44,470,283
Toplam assembly boyutu (bp)	431,545,587
GC (%)	35.36
AT (%)	64.6
N50 değeri (bp)	60,691,335
NG50 değeri (bp)	54,576,965
Tamamlanan BUSCO	1413
Tamamlanan ve tek kopya BUSCO	1386
Tamamlanan ve duplike BUSCO	27
Parçalanmış BUSCO	82

Sonuç olarak tahmin edilen genom büyüklüğünün (475.3 Mb) 90.8%'ını oluşturan, yedi pseudo molekülü de kapsayan 431.5 Mb büyüklüğünde scaffold assembly üretilmiştir (Maksimović vd., 2007). Temel assembly istatistikleri Çizelge 1'de gösterilmiştir. Scaffoldlara FASTA formatında <https://doi.org/10.6084/m9.figshare.24099927> (Scaffolds\_1-7.fasta) linkinden erişilebilir.

## 4.2. Protein Kodlayan Genlerin Tespit Edilmesi ve Gen İçeriğinin Değerlendirilmesi

HMM (hidden Markov model) baz alınarak protein kodlayan gen içeriğinin tespit edilmesi için AUGUSTUS (Stanke vd., 2008) programı kullanılmıştır. Varsayılan gen yapıları ve özellikleri gff dosya formatında <https://doi.org/10.6084/m9.figshare.24100032> linkine CodingGeneStructures.gff3 dosya ismi ile eklenmiştir.

Protein kodlayan gen lokuslarının anotasyonundan önce gen içeriğinin bütünlük kalitesi değerlendirilmiştir. Tez kapsamında adaçayı genomunun sekanslanmasının amacı terpenoid biyosentezinde görev alan sekansların potansiyel değişken lokuslarını tespit edebilmek olduğu için yüksek derecede bir gen bütünlüğüne ulaşabilmek önemlidir. Bu yüzden gen içeriğinin bütünlüğü değerlendirilirken kalite değerlendirmesi evrimsel olarak tek kopya ortologları olması beklenen genler baz alınarak gerçekleştirilmiş ve hatalı duplikasyonlar minimum düzeyde tutulmaya çalışılmıştır (Kelley ve Salzberg, 2010). BUSCO analizi sonucuna göre, embryophyta veri tabanından bulunan 1614 BUSCO grubundan 92.7%'si assembly scaffoldlarında 1413 tamamlanmış (87.6%) ve 82 parçalanmış (5.1%) BUSCO olarak temsil edilmektedir. Tamamlanmış BUSCOların 98.1%'i tek kopya olarak tanımlanmıştır (Çizelge 1). Duplike BUSCO oranının yüksek olması hatalı assembly sonucu ile ilişkilendirilmektedir. Ancak çalışmamızda BUSCO analiz sonuçları duplike BUSCO oranının minimal düzeyde olduğunu ve yüksek gen bütünlüğünü göstermektedir (Çizelge 1). Assembly scaffoldlarında tanımlanan BUSCO'ların ID'leri, açıklamaları ve lokasyonlarına (<https://doi.org/10.6084/m9.figshare.24103341>) linkinden ulaşılabilir. Ayrıca BUSCO analiz sonuçlarını temsil etmesi amacıyla elde edilen sonuçların bir kısmı Çizelge 2,3,4,5,6,7 ve 8'de verilmiştir.

**Çizelge 2.** Scaffold 1 BUSCO analiz sonuçları

BUSCO ID	Durum	Scaffold Numarası	Gen	Gen	İplik	Uzunluk
			Başlama Noktası	Bitiş Noktası		
5at3193	Tam	01_RagTag	44721998	44747763	+	3876
438at3193	Tam	01_RagTag	69915044	69933035	+	2089
469at3193	Tam	01_RagTag	27166480	27201884	-	1896

840at3193	Tam	01_RagTag	27582008	27593567	-	1683
938at3193	Tam	01_RagTag	25577822	25589320	+	1311
2325at3193	Duplike	01_RagTag	70545583	70562076	+	1142
2329at3193	Tam	01_RagTag	74588687	74593507	+	1195
3052at3193	Duplike	01_RagTag	714314	720953	-	852
3493at3193	Tam	01_RagTag	88221010	88237568	+	1412
3942at3193	Tam	01_RagTag	75694888	75701142	-	1095
13044at3193	Parça	01_RagTag	1931418	1936659	-	513

**Çizelge 3.** Scaffold 2 BUSCO analiz sonuçları

BUSCO ID	Durum	Scaffold Numarası	Gen	Gen	İplik	Uzunluk
			Başlama Noktası	Bitiş Noktası		
1521at3193	Tam	02_RagTag	27572784	27582785	-	842
1914at3193	Tam	02_RagTag	15152746	15163117	-	894
1998at3193	Tam	02_RagTag	10467044	10479901	+	1023
2050at3193	Tam	02_RagTag	39485252	39495906	-	1253
2325at3193	Duplike	02_RagTag	9449391	9466474	-	1216
2661at3193	Duplike	02_RagTag	27086059	27093926	-	1055
2926at3193	Tam	02_RagTag	16014408	16025569	+	1009
45095at3193	Parça	02_RagTag	8717577	8722735	+	345
47174at3193	Parça	02_RagTag	76596245	76597708	+	351

**Çizelge 4.** Scaffold 3 BUSCO analiz sonuçları

BUSCO ID	Durum	Scaffold Numarası	Gen	Gen	İplik	Uzunluk
			Başlama Noktası	Bitiş Noktası		
23at3193	Tam	03_RagTag	24231070	24256754	+	3895
364at3193	Tam	03_RagTag	42573563	42585883	-	1433
453at3193	Tam	03_RagTag	31949927	31981321	+	1597
727at3193	Tam	03_RagTag	46322681	46334282	-	1424
1259at3193	Tam	03_RagTag	19577680	19584824	-	1479
1578at3193	Tam	03_RagTag	35617410	35644267	+	1339
1589at3193	Parça	03_RagTag	57311993	57328929	+	899
1816at3193	Tam	03_RagTag	22570470	22580316	+	1575
1822at3193	Tam	03_RagTag	24925453	24936708	-	915
1916at3193	Tam	03_RagTag	26803468	26811779	+	948
2694at3193	Tam	03_RagTag	45030723	45041770	-	866

Çizelge 5. Scaffold 4 BUSCO analiz sonuçları

BUSCO ID	Durum	Scaffold Numarası	Gen	Gen	İplik	Uzunluk
			Başlama Noktası	Bitiş Noktası		
1388at3193	Tam	04_RagTag	6783537	6795658	-	1462
1440at3193	Tam	04_RagTag	4295362	4309877	-	1338
2203at3193	Tam	04_RagTag	14526607	14535660	-	1159
2653at3193	Tam	04_RagTag	49047718	49057419	-	1228
4483at3193	Tam	04_RagTag	27316389	27321335	+	950
6095at3193	Tam	04_RagTag	5689076	5699277	+	645
6412at3193	Tam	04_RagTag	12013678	12022376	+	1047
7277at3193	Parça	04_RagTag	14564357	14572516	-	602
7511at3193	Parça	04_RagTag	23814672	23816992	-	218
7542at3193	Tam	04_RagTag	15455389	15463721	-	894

Çizelge 6. Scaffold 5 BUSCO analiz sonuçları

BUSCO ID	Durum	Scaffold Numarası	Gen	Gen	İplik	Uzunluk
			Başlama Noktası	Bitiş Noktası		
2at3193	Tam	05_RagTag	10749625	10786390	-	3763
1332at3193	Tam	05_RagTag	39073832	39086400	+	914
1664at3193	Tam	05_RagTag	45811272	45831875	+	1591
2661at3193	Duplike	05_RagTag	45222250	45231504	+	1047
2834at3193	Tam	05_RagTag	8067554	8082541	-	1516
2921at3193	Tam	05_RagTag	50283879	50289239	-	1194
3441at3193	Tam	05_RagTag	6272578	6276590	-	1112
3704at3193	Parça	05_RagTag	25254099	25267323	-	743
3935at3193	Parça	05_RagTag	39348623	39359190	+	625
4251at3193	Tam	05_RagTag	20307358	20311671	+	1072

Çizelge 7. Scaffold 6 BUSCO analiz sonuçları

BUSCO ID	Durum	Scaffold Numarası	Gen	Gen	İplik	Uzunluk
			Başlama Noktası	Bitiş Noktası		
2028at3193	Tam	06_RagTag	43157230	43167439	-	1076
2661at3193	Duplike	06_RagTag	54037378	54043182	+	1054
2661at3193	Duplike	06_RagTag	54013958	54022874	+	1058

2770at3193	Tam	06_RagTag	14310702	14318660	-	1014
3595at3193	Tam	06_RagTag	54169187	54173972	+	1073
5306at3193	Tam	06_RagTag	25873944	25882445	-	900
5921at3193	Tam	06_RagTag	30076333	30084280	+	957
9385at3193	Tam	06_RagTag	32255252	32278858	+	1175
10307at3193	Tam	06_RagTag	50194288	50200926	-	720
10341at3193	Tam	06_RagTag	50063295	50072233	+	1248

**Çizelge 8.** Scaffold 7 BUSCO analiz sonuçları

BUSCO ID	Durum	Scaffold Numarası	Gen		İplik	Uzunluk
			Başlama Noktası	Gen Bitiş Noktası		
5135at3193	Tam	07_RagTag	28380281	28405940	+	1188
5793at3193	Tam	07_RagTag	19318268	19350562	-	1388
6430at3193	Tam	07_RagTag	13078082	13089766	-	780
6707at3193	Tam	07_RagTag	35982248	35988699	+	772
6756at3193	Tam	07_RagTag	35579786	35595301	-	813
8339at3193	Tam	07_RagTag	32815303	32831630	-	675
8731at3193	Tam	07_RagTag	23263026	23275500	+	890
10289at3193	Tam	07_RagTag	11235270	11249009	-	809
12643at3193	Tam	07_RagTag	43799926	43805481	-	733
13359at3193	Tam	07_RagTag	31513270	31519634	+	661

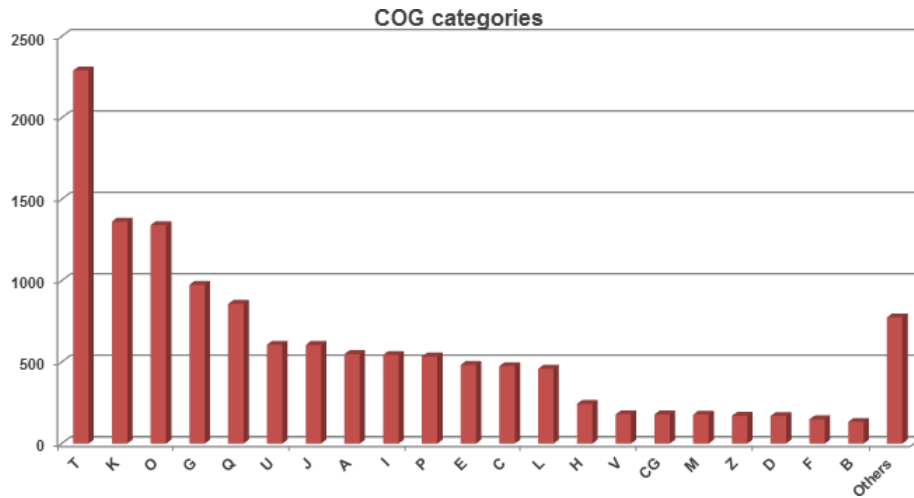
### 4.3. Fonksiyonel Anotasyon

Fonksiyonel anotasyon süreci eggNOG (Huerta-Cepas vd., 2017) ile gerçekleştirilmiştir. Hem tahmin edilen gen yapıları doğrulanmıştır hem de ortolog gruplarıyla eşleşen fonksiyonlar aktarılmıştır. Yapılan fonksiyonel anotasyon analizi sonucunda adaçayı assembly scaffoldlarında 21,443 protein kodladığı varsayılan gen bölgesi tahmin edilmiştir. 21,443 ortolog gen dışında, 20,191'i 124 COG (Clusters of Orthologs) kategorisine atanmıştır ve 948 gen çoklu kategoriye atanmıştır. 9421 gen COG kategorisine atanmış olup 'S' olarak fonksiyonu bilinmeyen bir ortolog gen üyesi olarak tanımlanmıştır (<https://doi.org/10.6084/m9.figshare.24103341>). eggNOG ile fonksiyonel anotasyonu gerçekleştirilmiş olan adaçayı scaffoldları, yüksek E-value değerlerine göre seçilerek sorgulanan bazı genlerdeki tanımlanan fonksiyonlar Çizelge

9’da temsilen gösterilmiştir. En yaygın COG atamalarının dağılımı Şekil 19’da gösterilmiştir.

**Çizelge 9.** eggNOG ile gerçekleştirilen *ab initio* gen yapısı anotasyonu sonucu adaçayı genomunda tespit edilen genlerin analiz sonuçlarından bir kısmı

Sorgulanan Gen ID	E-Value	COG_fonksiyonel kategorisi
03_RagTag.g34031	9.94e-316	U
02_RagTag.g22948	9.85e-313	E
06_RagTag.g67620	9.5e-316	S
01_RagTag.g11640	9.49e-315	J
06_RagTag.g69109	9.49e-315	Q
02_RagTag.g23903	8e-314	D
05_RagTag.g52097	8.97e-313	-
04_RagTag.g45324	8.95e-313	S
01_RagTag.g2524	8.64e-315	C
06_RagTag.g61854	8.62e-312	T



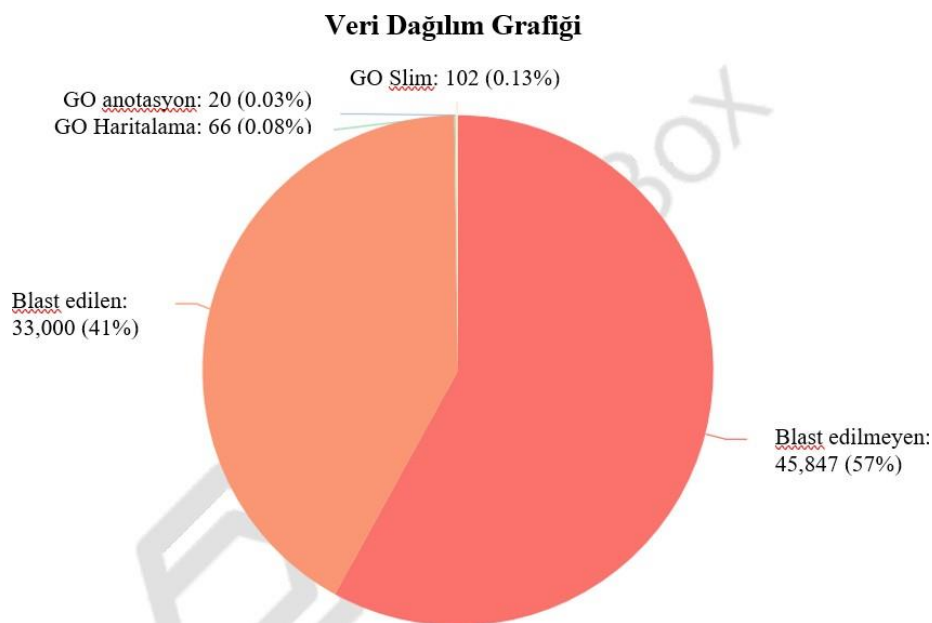
**Şekil 19.** Fonksiyonel anotasyonu yapılmış gen yapılarının COG kategori dağılımı gösterimi  
**T:** Sinyal transdüksiyon mekanizması, **K:** Transkripsiyon, **O:** Posttranslasyonel modifikasyon, protein değişimi, şaperonlar, **G:** Karbonhidrat transport ve metabolizması, **Q:** Sekonder metabolitlerin biosentezi, transport ve katabolizması, **U:** İntraselüler taşıma, salgı ve kesecikli taşıma, **J:** Translasyon, ribozomal yapı ve biyogenez, **A:** RNA işleme ve modifikasyonu, **I:** Yağ taşıma ve metabolizması, **P:** İnorganik demir taşıma ve metabolizması, **E:** Amino asit taşıma ve metabolizması, **C:** Enerji üretimi ve dönüşümü, **L:** Replikasyon, rekombinasyon ve tamir, **H:** Koenzim taşıma ve metabolizması, **V:** Savunma mekanizması, **M:** Hücre duvarı/zarı/zarflı biyogenez, **Z:** Sitoskeleton, **D:** Hücre döngüsü kontrolü, hücre bölünmesi, kromozom bölünmesi, **F:** Nükleotid taşıma ve metabolizması, **B:** Kromatin yapısı ve dinamiği. En az 100 genin değerlendirildiği COG değerlendirmeleri grafikte barlarla gösterilmiştir. 100 genin altında olan değerlendirilmeler ‘Others’ barı altında gösterilmiştir.

Kodlanan genler ve peptitler <https://doi.org/10.6084/m9.figshare.24100083> linkinde Gene\_Structures\_Nucleotide.fasta ve Gene\_Structures\_Peptide.fasta dosya isimleriyle bulunmaktadır.

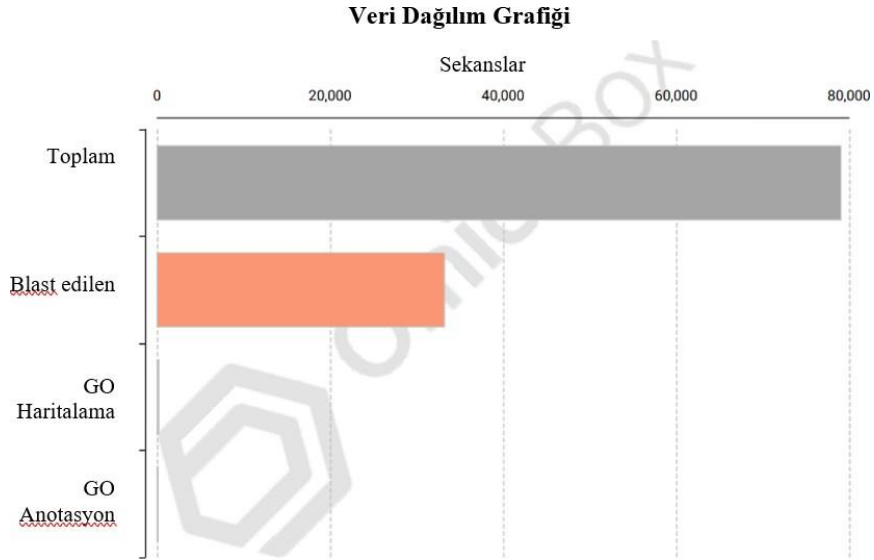
OmicsBox ile gerçekleştirilen fonksiyonel anotasyon sürecinde *Salvia officinalis*'e ait protein kodlayan genler ile *Salvia hispanica* protein peptit sekansları hizalanmıştır. Yapılan hizalama sonucunda 33,000 hit ile 41% oranında sekanslar benzerlik göstermiştir (Şekil 19-20). Analiz sonucunda blast edilen sekansların uzunluğu, benzerlik oranı, hizalanan sekansların uzunluğu, hizalanan sekansların başlangıç ve bitiş nükleotid konumları elde edilmiştir. Sonuçlardan bir kısmı, yüksek E-value değerine göre seçilerek temsilen Çizelge 10'da verilmiştir.

**Çizelge 10.** Adacıyı ve chia tohumunun protein kodlayan genleri arasındaki blast analiz sonuçlarının kısmi gösterimi

Sekans İsmi	Sekans Uzunluğu	Blast Sayısı	Hit	Blast E-Value	Min	Blast Benzerlik Oranı
GWHBJVP00000001_RagTag.g11983	162	10		9.9E-96		90.92%
GWHBJVP00000004_RagTag.g45697	257	10		9.9E-95		61.70%
GWHBJVP00000003_RagTag.g37805	934	10		9.9E-94		57.09%
GWHBJVP00000004_RagTag.g44844	140	10		9.9E-94		90.70%
GWHBJVP00000005_RagTag.g58352	229	2		9.9E-91		84.05%
GWHBJVP00000006_RagTag.g64423	254	10		9.9E-91		70.26%
GWHBJVP00000007_RagTag.g72918	207	10		9.9E-90		56.14%
GWHBJVP00000003_RagTag.g34769	230	6		9.9E-9		45.61%
GWHBJVP00000006_RagTag.g67897	527	1		9.9E-9		43.61%
GWHBJVP00000007_RagTag.g70654	240	10		9.9E-83		72.73%

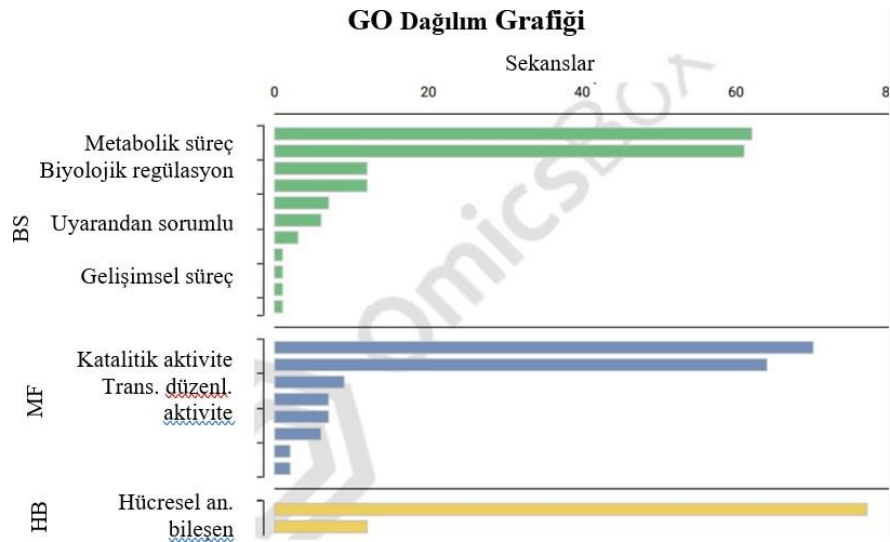


**Şekil 19.** Hizalama sonucu veri dağılım grafiği 1



Şekil 20. Hizalama sonucu veri dağılım grafiği 2

OmicsBox ile yapılan Gen Ontolojisi analizi sonuçlarına göre biyolojik süreç, moleküler fonksiyon ve hücrel bileşenler görevlerinde rol alan sekansların dağılımı şekil 21’de gösterilmiştir. Gen Ontolojisi dağılımına göre sekansların çoğunlukla; metabolik süreç, biyolojik regülasyon, katalitik aktivite, translasyon düzenleyici aktivite ve hücrel anatomik bileşen gibi görevlerde bulunduğu söylenebilmektedir.



Şekil 21. Gen Ontolojisi dağılım grafiği

Gen Ontolojisi analizi sonuçlarına göre biyolojik süreç, moleküler fonksiyon ve hücrel bileşen süreçlerinde rol alan sekansların fonksiyon grafikleri Şekil 22-23-24’te gösterilmiştir.



#### 4.4. Terpenoid Biyosentezinde Rol Alan Genlerin Tespit Edilmesi

Terpenoid biyosentezinde görev aldığı varsayılan genlerin tespiti protein kodlayan gen bölgelerinden gerçekleştirilmiştir. Ortolog genler ile homoloji analizi erişilebilir *Salvia* terpenoid biyosentez genleri (terpenoid biyosentez yolak genleri ve terpen sentaz genleri) ile gerçekleştirilmiştir ve homolog peptitlerin fonksiyonel analizi InterProScan kullanılarak peptit domain işaretlerine göre gerçekleştirilmiştir (Paysan-Lafosse vd., 2023). Sonuç olarak, 75 terpen sentaz ve 67 terpenoid biyosentez yolak genleri (MVA ve MEP yolları ve diğer terpenoid biyosentez genleri) olmak üzere 142 terpenoid biyosentezinde görev aldığı varsayılan gen tespit edilmiştir. Fonksiyonel tanım, yolak ve ID bilgisi ve 142 genin assembly scaffoldlarındaki lokasyonuna <https://doi.org/10.6084/m9.figshare.24103341> linkinden erişilebilmektedir (Çizelge 11,12,13,14,15,16,17). Kodlanmış peptit sekanslarına <https://doi.org/10.6084/m9.figshare.24100134> linkinden TerpenoidBiosynthesis\_Coding.fasta and TerpenoidBiosynthesis\_Peptide.fasta dosya ismi ile ulaşılabilir. Tanımlanmış terpenoid biyosentez genlerinin assembly scaffoldlarındaki, splay koordinat bölgelerini de içeren, detaylı koordinatları ve kodlayan iplik (+/-) bilgisi <https://doi.org/10.6084/m9.figshare.24100032> linkinde bulunmaktadır ve CodingGeneStructures.gff3 dosya ismi ile erişilebilmektedir.

Terpenoid biyosentez gen tanımlama çalışmalarında çoğunlukla *Salvia* cinsine ait transkriptom verisi kullanılmıştır. Bu yüzden bu cins için erişilebilir en başarılı ve düzenli veri, splay bilgilerini içeren kodlanan sekans verileridir. Tez çalışması kapsamında terpenoid biyosentez genlerinin tüm genomda tespit edilmesi, genomik scaffoldlardaki ve açık okuma çerçevesi bölgelerindeki, exon-intron bölgeleri dahil olmak üzere, genlerin yapısal gen modelleri kullanılarak gerçekleştirilmiştir. Bu sayede terpenoid sentezinde görev aldığı varsayılan genlerin kromozomlar arası dağılımını belirlemek mümkün olmaktadır.

Embryophyta evrimsel sürecinde terpen sentaz en önemli genişletilmiş sekonder metabolit gen ailelerindedir. Hem ardışık duplikasyonlar hem de tüm genom duplikasyonları bitkilerde terpen sentaz gen ailesinin genişlemesinde rol almaktadır (J. Li vd., 2021). Bitki genomunda yaklaşık 30-100 terpen sentaz kodlayan gen bulunmaktadır (Pichersky ve Raguso, 2018). Terpenoid biyosentezi ile ilişkisi olan genler yedi kromozoma dağılmış durumdadır ve hem terpen sentaz hem de terpenoid biyosentez yolak bileşenlerini kodlayan genler ardışık bloklarda bulunmaktadır (Şekil

25, Çizelge 11,12,13,14,15,16,17). Genlerin kromozomlar arası dağılımı, Şekil 25'te görüldüğü gibi, gen topluluklarının çoğunluğu kromozom 1,2,3 ve 6'da olmak üzere dört kromozoma düzensiz olarak dağılmış durumdadır. Son zamanlarda yapılan bir çalışmada 342 terpenoid biyosentezi ile ilişkili gen lavanta (*Lavandula angustifolia*) genomunda 126 ardışık blok olarak tanımlanmıştır. Lavanta, Nepetoideae alt sınıfında *Salvia* cinsleri ile birlikte bulunan bir Lamiaceae türüdür. Çalışmanın sonucu olarak, tez kapsamında bulunan sonuçlara benzer olarak, terpenoid biyosentezinde görev alan genlerin kromozomlar arasında düzensiz olarak dağıldığı rapor edilmiştir (J. Li vd., 2021). Çalışma sonuçlarına göre adaçayında terpenoid biyosentezi ile ilişkili genler transkriptom analizine göre tanımlanmıştır ve 20 benzersiz monoterpen sentaz geni, 29 benzersiz diterpen sentaz geni, 12 benzersiz triterpen sentaz geni ve 3 benzersiz sesquiterpen sentaz geni tanımlanmıştır (Ali vd., 2017). Tez kapsamında yapılan çalışmada ise adaçayında mono-, di- ve triterpen sentaz kodlayan genlerde benzer sayılar (23, 30 ve 10 gen) belirtilmiştir.

**Çizelge 11.** Scaffold 1'de Terpenoid biyosentezinde görev alan genler

Scaffold Gen ID	Yolak/TPS	Gen İsmi/Fonksiyon
01_RagTag.g36	MEP	ID11 isopentenyl diphosphate isomerase-like
01_RagTag.g15795	Triterpene_synthase	Squalene monooxygenase
01_RagTag.g15686	MVA	HMGR1 3-hydroxy-3-methylglutaryl-coenzyme A reductase-like
01_RagTag.g15596	Triterpene_synthase	Beta-amyrin synthase
01_RagTag.g15408	Terpenoid_backbone_synthesis	STE24 endopeptidases-like
01_RagTag.g15392	Diterpene_synthase	Gibberellin 20-oxidase-5
01_RagTag.g15326	Monoterpene_synthase	Myrcene/ocimene synthase
01_RagTag.g15324	Monoterpene_synthase	Myrcene/ocimene synthase
01_RagTag.g933	Diterpene_synthase	Ent-copalyl diphosphate synthase
01_RagTag.g14634	Terpenoid_backbone_synthesis	protein-S-isoprenylcysteine O-methyltransferase-like
01_RagTag.g13102	Diterpene_synthase	Ent-copalyl diphosphate synthase
01_RagTag.g11868	MEP	DXS 1-deoxy-D-xylulose 5-phosphate synthase-like
01_RagTag.g11284	Triterpene_synthase	Beta-amyrin synthase

01_RagTag.g11278	Triterpene_synthase	Beta-amyrin synthase
01_RagTag.g10549	MVA	HMGS hydroxylmethylglutaryl-CoA synthase-like
01_RagTag.g5643	Diterpene_synthase	Ent-isokaurene C2-hydroxylase
01_RagTag.g5645	Diterpene_synthase	9beta-pimara-7, 15-diene oxidase
01_RagTag.g5651	Monoterpene_synthase	Geraniol isomerase synthase
01_RagTag.g9259	Sesquiterpene synthase	$\alpha$ -humulene/ $\beta$ -caryophyllene synthase
01_RagTag.g7182	MEP	DXS3 1-deoxy-D-xylulose 5-phosphate synthase-like
01_RagTag.g5878	Diterpene_synthase	Ent-copalyl diphosphate synthase
01_RagTag.g5807	Diterpene_synthase	Ent-copalyl diphosphate synthase
01_RagTag.g5657	Diterpene_synthase	Ent-isokaurene C2-hydroxylase
01_RagTag.g5649	Diterpene_synthase	9beta-pimara-7, 15-diene oxidase
01_RagTag.g11271	Triterpene_synthase	Beta-amyrin synthase
01_RagTag.g11273	Triterpene_synthase	Beta-amyrin synthase
01_RagTag.g3246	MEP	DXS 1-deoxy-D-xylulose 5-phosphate synthase-like
01_RagTag.g14447	Diterpene_synthase	Gibberellin 20-oxidase-5
01_RagTag.g15681	MVA	HMGR1 3-hydroxy-3-methylglutaryl-coenzyme A reductase-like
01_RagTag.g15682	MVA	HMGR2 3-hydroxy-3-methylglutaryl-coenzyme A reductase-like
01_RagTag.g15684	MVA	HMGR6 3-hydroxy-3-methylglutaryl-coenzyme A reductase-like
01_RagTag.g15690	MVA	HMGR1 3-hydroxy-3-methylglutaryl-coenzyme A reductase-like
01_RagTag.g15744	Diterpene_backbone_synthesis	geranylgeranyl diphosphate synthase-like

Çizelge 12. Scaffold 2’de Terpenoid biyosentezinde görev alan genler

Scaffold Gen ID	Yolak/TPS	Gen İsmi/Fonksiyon
02_RagTag.g29554	MEP	DXR1 1-deoxy-D-xylulose-5-phosphate reductoisomerase-like
02_RagTag.g17608	Monoterpene synthase	Myrcene/ocimene synthase
02_RagTag.g28577	Monoterpene synthase	Myrcene/ocimene synthase
02_RagTag.g28559	Monoterpene synthase	Myrcene/ocimene synthase
02_RagTag.g27870	Monoterpene synthase	Myrcene/ocimene synthase
02_RagTag.g21282	Diterpene synthase	Gibberellin 2-oxidase
02_RagTag.g26340	Diterpene synthase	Ent-kaurene synthase-6
02_RagTag.g25700	MVA	HMGR1 3-hydroxy-3-methylglutaryl-coenzyme A reductase-like
02_RagTag.g25575	Diterpene backbone synthesis	geranylgeranyl diphosphate synthase-like
02_RagTag.g23678	MVA	HMGR1 3-hydroxy-3-methylglutaryl-coenzyme A reductase-like
02_RagTag.g24172	Monoterpene synthase	Myrcene/ocimene synthase
02_RagTag.g24847	MEP	HDR1 4-hydroxy-3-methylbut-2-enyl diphosphate reductase-like
02_RagTag.g24845	MEP	HDR2 4-hydroxy-3-methylbut-2-enyl diphosphate reductase-like
02_RagTag.g24179	Monoterpene synthase	Myrcene/ocimene synthase
02_RagTag.g25097	Triterpene synthase	Farnesyl-diphosphate farnesyltransferase
02_RagTag.g25701	MVA	HMGR1 3-hydroxy-3-methylglutaryl-coenzyme A reductase-like
02_RagTag.g25732	Triterpene synthase	Squalene monooxygenase
02_RagTag.g23173	Terpenoid backbone synthesis	farnesol dehydrogenase-like
02_RagTag.g22340	Terpenoid backbone synthesis	prenylcysteine alpha-carboxyl methylesterase-like
02_RagTag.g26555	Diterpene synthase	Ent-copalyl diphosphate synthase
02_RagTag.g22182	Diterpene synthase	Ent-kaurenoic acid hydroxylase
02_RagTag.g21164	Monoterpene synthase	(+)-neomenthol dehydrogenase1
02_RagTag.g20273	Sesqui and triterpene backbone synthesis	farnesyl diphosphate synthase-like
02_RagTag.g20170	Terpenoid backbone synthesis	STE24 endopeptidases-like
02_RagTag.g19453	Terpenoid backbone synthesis	all-trans-nonaprenyl-diphosphate synthase-like
02_RagTag.g28180	Monoterpene synthase	Myrcene/ocimene synthase
02_RagTag.g28451	Monoterpene synthase	Myrcene/ocimene synthase
02_RagTag.g28455	Monoterpene synthase	Myrcene/ocimene synthase
02_RagTag.g28464	Monoterpene synthase	Myrcene/ocimene synthase
02_RagTag.g17234	MVA	mevalonate kinase-like

Çizelge 13. Scaffold 3'te Terpenoid biyosentezinde görev alan genler

Scaffold Gen ID	Yolak/TPS	Gen İsmi/Fonksiyon
03_RagTag.g41023	Diterpene synthase	Momilactone-A synthase
03_RagTag.g40438	Diterpene backbone synthesis	GGPSIII geranylgeranyl diphosphate synthase-like
03_RagTag.g40077	Terpenoid backbone synthesis	DHDDS4 ditrans,polycis-polyprenyl diphosphate synthase 4-like
03_RagTag.g40075	Terpenoid backbone synthesis	DHDDS4 ditrans,polycis-polyprenyl diphosphate synthase 4-like
03_RagTag.g39219	Terpenoid backbone synthesis	CHLP1 geranylgeranyl diphosphate reductase-like
03_RagTag.g32354	Terpenoid backbone synthesis	CHLP1 geranylgeranyl diphosphate reductase-like
03_RagTag.g32695	MEP	HDR1 4-hydroxy-3-methylbut-2-enyl diphosphate reductase-like
03_RagTag.g37112	Diterpene synthase	Ent-kaurenoic acid hydroxylase
03_RagTag.g36723	Terpenoid backbone synthesis	CHLP3 geranylgeranyl diphosphate reductase-like
03_RagTag.g36721	Terpenoid backbone synthesis	CHLP1 geranylgeranyl diphosphate reductase-like
03_RagTag.g33922	Terpenoid backbone synthesis	PCME1 prenylcysteine alpha-carboxyl methylesterase-like
03_RagTag.g36076	MEP	MCT 2-C-methyl-D-erythritol 4-phosphate cytidyltransferase-like
03_RagTag.g35187	MEP	DXS 1 1-deoxy-D-xylulose 5-phosphate synthase-like
03_RagTag.g34098	Diterpene backbone synthesis	GGPSII2 geranylgeranyl diphosphate synthase-like
03_RagTag.g33982	Terpenoid backbone synthesis	DHDDS4 ditrans,polycis-polyprenyl diphosphate synthase 4-like
03_RagTag.g33112	Terpenoid backbone synthesis	FNTB protein farnesyltransferase subunit beta-like
03_RagTag.g37730	Monoterpene synthase	(+)-neomenthol dehydrogenase1
03_RagTag.g37734	Monoterpene synthase	(+)-neomenthol dehydrogenase1
03_RagTag.g38302	Triterpene synthase	Squalene monooxygenase
03_RagTag.g31940	Terpenoid backbone synthesis	PCYOX1 prenylcysteine oxidase/farnesylcysteine lyase-like
03_RagTag.g31111	Sesquiterpene synthase	$\alpha$ -humulene/ $\beta$ -caryophyllene synthase
03_RagTag.g30773	Monoterpene synthase	1, 8-cineole synthase

Çizelge 14. Scaffold 4’de Terpenoid biyosentezinde görev alan genler

Scaffold Gen ID	Yolak/TPS	Gen İsmi/Fonksiyon
04_RagTag.g42863	MEP	HDS 1 4-hydroxy-3-methylbut-2-enyl diphosphate synthase-like
04_RagTag.g43791	MEP	CMK 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase-like
04_RagTag.g44623	MEP	HDS 1 4-hydroxy-3-methylbut-2-enyl diphosphate synthase-like
04_RagTag.g46769	Terpenoid backbone synthesis	DHDDS4 ditrans,polycis-polyprenyl diphosphate synthase 4-like
04_RagTag.g47147	MVA	HMGR1 3-hydroxy-3-methylglutaryl-coenzyme A reductase-like
04_RagTag.g48003	Diterpene backbone synthesis	GGPSIII geranylgeranyl diphosphate synthase-like
04_RagTag.g48686	Monoterpene synthase	(3S)-linalool synthase2
04_RagTag.g44109	Monoterpene synthase	Geraniol isomerase synthase
04_RagTag.g43830	Monoterpene synthase	(+)-neomenthol dehydrogenase1

Çizelge 15. Scaffold 5’te Terpenoid biyosentezinde görev alan genler

Scaffold Gen ID	Yolak/TPS	Gen İsmi/Fonksiyon
05_RagTag.g60376	Monoterpene synthase	Myrcene/ocimene synthase
05_RagTag.g60307	Diterpene synthase	Ent-copalyl diphosphate synthase
05_RagTag.g52984	Diterpene synthase	Gibberellin 2-oxidase
05_RagTag.g60027	MVA	HMGR2 3-hydroxy-3-methylglutaryl-coenzyme A reductase-like
05_RagTag.g54335	MVA	HMGR2 3-hydroxy-3-methylglutaryl-coenzyme A reductase-like
05_RagTag.g57537	Diterpene synthase	Ent-kaurene synthase-6
05_RagTag.g57309	Diterpene synthase	Ent-kaurene synthase-4
05_RagTag.g60889	Terpenoid backbone synthesis	FOLK1 farnesol kinase-like
05_RagTag.g60891	Terpenoid backbone synthesis	FOLK1 farnesol kinase-like

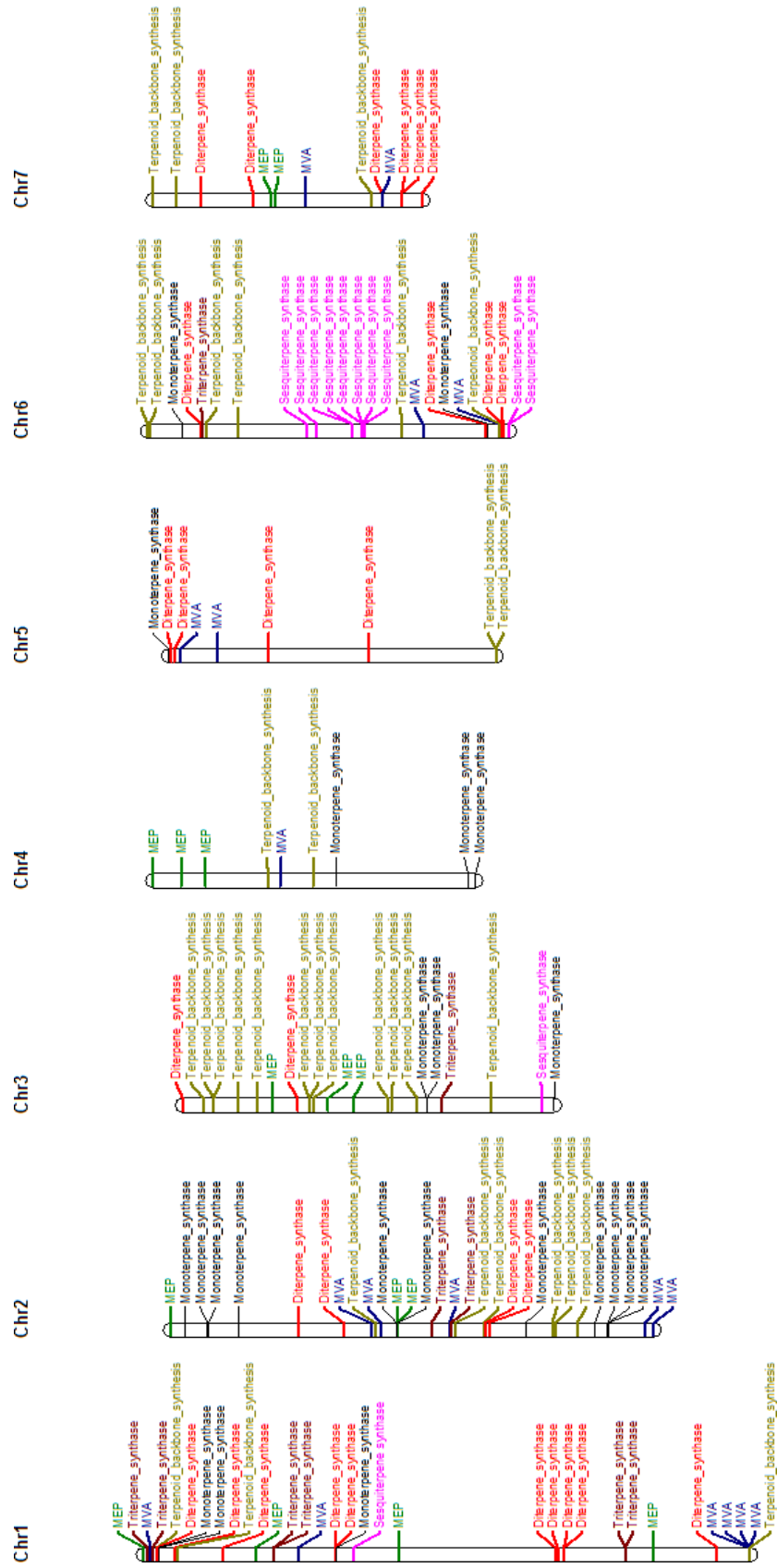
Çizelge 16. Scaffold 6’da Terpenoid biyosentezinde görev alan genler

Scaffold Gen ID	Yolak/TPS	Gen İsmi/Fonksiyon
06_RagTag.g70358	Diterpene backbone synthesis	GGPSII2 geranylgeranyl diphosphate synthase-like
06_RagTag.g61395	Diterpene backbone synthesis	GGPSII6 geranylgeranyl diphosphate synthase-like
06_RagTag.g69096	Monoterpene synthase	(E, E)-geranyl linalool synthase
06_RagTag.g68535	Diterpene synthase	Gibberellin 2-oxidase
06_RagTag.g68508	Triterpene synthase	Beta-amyrin synthase
06_RagTag.g68445	Monoterpene backbone synthesis	GPPS geranyl diphosphate synthase-like
06_RagTag.g67452	Terpenoid backbone synthesis	CMT2 protein-S-isoprenylcysteine O-methyltransferase-like
06_RagTag.g65652	Sesquiterpene synthase	$\alpha$ -humulene/ $\beta$ -caryophyllene

06_RagTag.g64527	Sesquiterpene synthase	synthase Germacrene-D synthase (TPS-6)
06_RagTag.g65488	Sesquiterpene synthase	$\alpha$ -humulene/ $\beta$ -caryophyllene synthase
06_RagTag.g65489	Sesquiterpene synthase	$\alpha$ -humulene/ $\beta$ -caryophyllene synthase
06_RagTag.g65788	Sesquiterpene synthase	Cis-muroladiene synthase
06_RagTag.g65789	Sesquiterpene synthase	Cis-muroladiene synthase
06_RagTag.g65792	Sesquiterpene synthase	Cis-muroladiene synthase
06_RagTag.g65857	Sesquiterpene synthase	Cis-muroladiene synthase
06_RagTag.g66875	Terpenoid backbone synthesis	PCME1 prenylcysteine alpha-carboxyl methyltransferase-like
06_RagTag.g63175	MVA	AACT1 Acetyl-CoA C-acetyltransferase-like
06_RagTag.g62153	Diterpene synthase	Gibberellin 2-oxidase
06_RagTag.g69329	Monoterpene synthase	Myrcene/ocimene synthase
06_RagTag.g69770	MVA	AACT1 Acetyl-CoA C-acetyltransferase-like
06_RagTag.g69798	Sesqui and triterpene backbone synthesis	FPPS farnesyl diphosphate synthase-like
06_RagTag.g69901	Diterpene synthase	Ent-kaurene synthase-5
06_RagTag.g69904	Diterpene synthase	Ent-kaurene synthase-1
06_RagTag.g61566	Sesquiterpene synthase	Germacrene-D synthase (TPS-6)
06_RagTag.g61564	Sesquiterpene synthase	$\alpha$ -humulene/ $\beta$ -caryophyllene synthase

**Çizelge 17.** Scaffold 7'de Terpenoid biyosentezinde görev alan genler

Scaffold Gen ID	Yolak/TPS	Gen İsmi/Fonksiyon
07_RagTag.g78564	Terpenoid backbone synthesis	DHDDS3 ditrans,polycis-polyprenyl diphosphate synthase 3
07_RagTag.g77798	Terpenoid backbone synthesis	FACE2 farnesylated protein-converting enzyme 2
07_RagTag.g77077	Diterpene synthase	Ent-kaurene synthase-1
07_RagTag.g73074	Diterpene synthase	Gibberellin 20-oxidase-3
07_RagTag.g75112	MEP	DXS1 1-deoxy-D-xylulose 5-phosphate synthase-like
07_RagTag.g73697	MEP	IDII isopentenyl diphosphate isomerase-like
07_RagTag.g74643	MVA	PMK 5-phosphomevalonate kinase-like
07_RagTag.g76370	Terpenoid backbone synthesis	CMT1 protein-S-isoprenylcysteine O-methyltransferase-like
07_RagTag.g71932	Diterpene synthase	Gibberellin 20-oxidase-1
07_RagTag.g76727	MVA	MVK1 mevalonate kinase-like
07_RagTag.g71454	Diterpene synthase	Ent-copalyl diphosphate synthase
07_RagTag.g71452	Diterpene synthase	Ent-copalyl diphosphate synthase
07_RagTag.g77945	Diterpene synthase	Momilactone-A synthase



Şekil 25. Adaçayında terpenoid biyosentezinde görev aldığı varsayılan genlerin genomik dağılımı

Toplamda 12 sesquiterpen sentaz geni tanımlanmıştır, bunlardan 8 sesquiterpen geni kromozom 6'da ardışık kümeler halinde bulunmaktadır (Şekil 25). İki çalışma arasında tanımlanan gen sayılarında gözlemlenen farklılıklar, tez kapsamında yapılan çalışmada genomik kodlayan sekanslar için, gen tanımlaması uygulamasında transkriptler yerine genomik yaklaşımın uygulanmasından kaynaklanmaktadır. Böylelikle ardışık gen kümelerinin tanımlanması ve lokasyonlarının belirlenmesi mümkün olmaktadır. Ardışık monoterpen sentaz ve diterpen sentaz blokları için en ilişkili kromozomlar, kromozom 2 ve kromozom 1-7 olarak görülmektedir. MVA ve MEP yolaklarının da dahil olduğu terpenoid biyosentez gen yolları için en ilişkili kromozomlar ise kromozom 1 ve 4 olarak görülmektedir (Şekil 25).

#### **4.5. *S. officinalis* Exomunda SNP Varyantları ve Terpenoid Biyosentez Genleri**

Ardışık gen bloklarındaki mutasyon ve doğal seçim sonucu kopyaların duplikasyonundaki fonksiyonel çeşitlilik, bitki sekonder metabolitlerinde en geniş sınıf olan terpenoidlerdeki muazzam ölçüdeki çeşitliliğe katkı sağlamaktadır (J. Li vd., 2021; Trindade vd., 2018). Terpen sentaz genlerinin belirgin özellikleri bitki terpen çeşitliliğindeki mutasyonların etkilerini ölçeklendirmenin belirlenmesinde etkilidir. Terpen sentazlar genellikle aynı substrattan çoklu ürünler üretmektedir ve terpen sentazdaki tek bir amino asit değişikliği genellikle çoklu yeni terpen ve ürünleri olarak yansıtmaktadır. Sonuç olarak, terpen sentazdaki tek bir amino asitteki mutasyon değişikliği bile sıklıkla önemli sonuçlar oluşturmaktadır (Pichersky ve Raguso, 2018). Reaksiyonlardaki çeşitlilikten kaynaklanan terpen iskeletlerindeki modifikasyonlar, terpen sentaz çeşitliliği neticesinde nihai ürün genişlemesine sebep olmaktadır. Bu nedenle terpen sentaz kodlayan genlerde sekans çeşitliliğini tanımlamak, belirli bir metabolik fenotipi ya da arzulan terpen son ürününün üretimini arttırılmasını belirli bir allele ilişkilendirmek için oldukça değerlidir. *Salvia* cinsinde terpenoid biyosentez geninde potansiyel çeşitliliklerin tanımlanması daha önceki çalışmada, *S. officinalis*'in erişilebilen genom sekans çalışmasında (C. Y. Li vd., 2022), belirtilmemiştir. Tez çalışması kapsamında adaçayı genomunun sekanslanması, genomik nükleotid varyasyonların tanımlanmasında karşılaştırmalı sekans analiz uygulamasına olanak sağlamıştır.

Çalışmamızda sonuç olarak referans sekansa ait, terpenoid biyosentezinde görev aldığı varsayılan genlerin exon bölgesinde 447 yer değişikliği tespit edilmiştir. Lokus için, her bir taraftan, en az 20 bp sekans uyumsuzluğu olmadan mükemmel eşleşme filtrelemesi sonucu olarak 82 tanesi benzersiz terpenoid biyosentez genlerinde olmak üzere toplamda 188 tek nükleotid polimorfizmi belirlenmiştir. Detaylı sonuçlar (<https://doi.org/10.6084/m9.figshare.24103341>) linkinde bulunmaktadır. Predominant olan transisyon mutasyonları toplam SNP sayısının 64.9%'unu temsil etmektedir. Transversiyonlar SNP'lerin 35.1%'ini oluşturmaktadır. Pürin/pürin (A/G, G/A) en fazla bulunan yer değişikliği tipi olarak, 188 SNP lokusunun 35.1%'ini temsil etmektedir (Çizelge 18).

**Çizelge 18.** *S. officinalis* genomunda tahmin edilen terpenoid biyosentez genlerinin varyant istatistikleri

Toplam exom varyantı <sup>a</sup>	Sorgulanan <sup>c</sup>	Referans <sup>d</sup>	Terpenoid biyosentez varyantı <sup>b</sup>	Sorgulanan <sup>c</sup>	Referans <sup>d</sup>
GA	3969 (16.15%)	3985 (16.22%)	AC	12 (6.03%)	8 (4.02%)
GT	1193 (4.86%)	1129 (4.59%)	AT	8 (4.02%)	7 (3.52%)
GC	997 (4.06%)	1024 (4.17%)	AG	36 (19.12%)	30 (15.96%)
AC	1155 (4.70%)	1223 (4.98%)	GA	30 (15.96%)	36 (19.12%)
AG	3985 (16.22%)	3969 (16.15%)	GT	9 (4.79%)	4 (2.13%)
AT	952 (3.87%)	993 (4.04%)	GC	13 (6.91%)	5 (2.66%)
TA	993 (4.04%)	952 (3.87%)	TG	4 (2.13%)	9 (4.79%)
TG	1129 (4.59%)	1193 (4.86%)	TA	7 (3.52%)	8 (4.02%)
TC	3982 (16.00%)	4018 (16.35%)	TC	24 (12.77%)	32 (17.02%)
CG	1024 (4.17%)	997 (4.06%)	CT	32 (17.02%)	24 (12.77%)
CA	1223 (4.98%)	1155 (4.70%)	CG	5 (2.66%)	13 (6.91%)
CT	4018 (16.35%)	3932 (16.00%)	CA	8 (4.02%)	12 (6.03%)
Toplam	24,570	24,570	Toplam	188	188

*S. officinalis* genom assembly scaffoldlarında exonik sekansların tamamındaki yer değişiklikleri tanımlanmıştır.

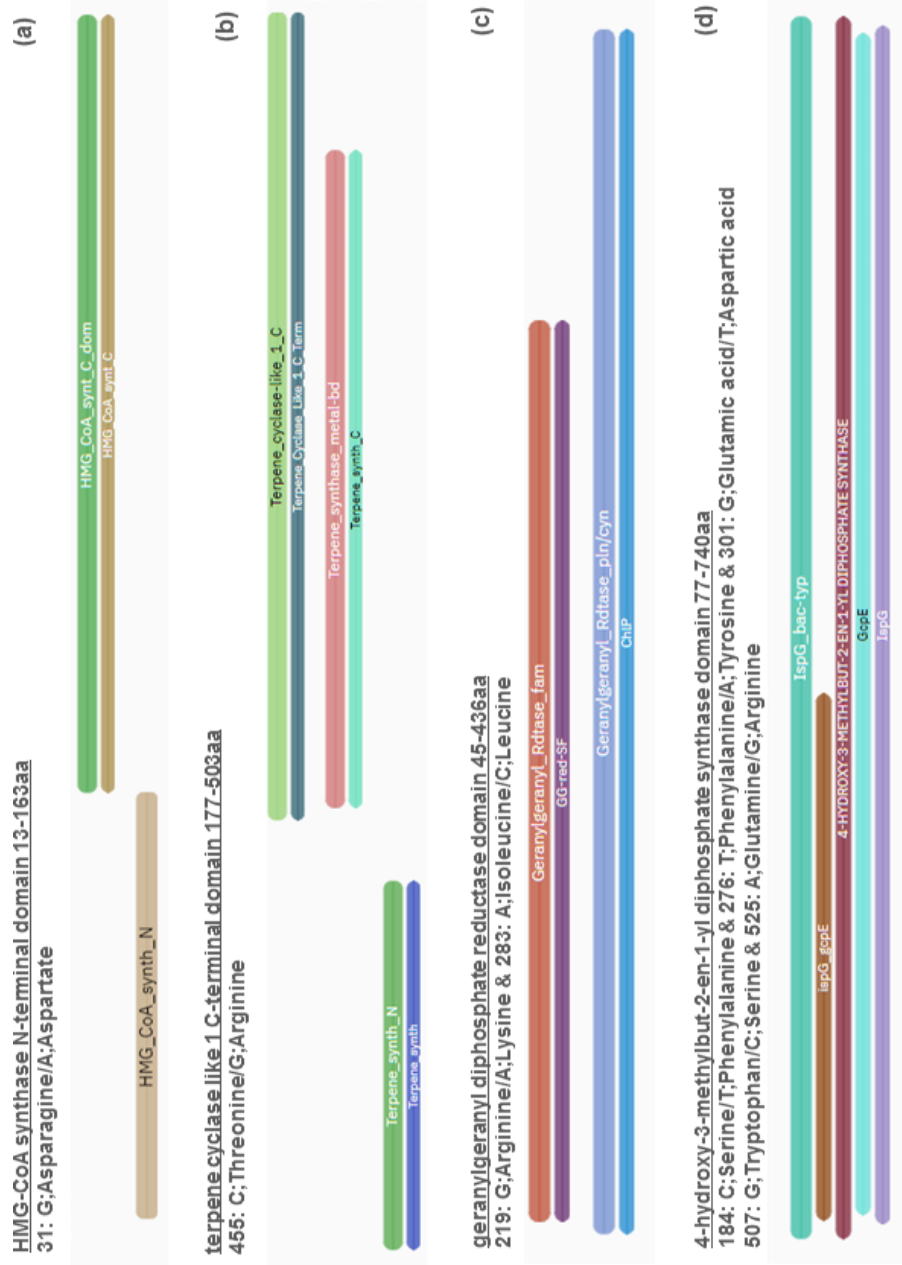
<sup>b</sup>Tahmin edilen terpenoid biyosentez genlerinin exonik sekanslarındaki yer değişiklikleri tanımlanmıştır.

<sup>c</sup>Sorgulanan sütunu tez kapsamında çalışılan ve üretilen sekans verisini temsil etmektedir.

<sup>d</sup>Referans sütunu <https://ngdc.cnbc.ac.cn/gwh> linkinden GWHBJVP00000000 ID numarası ile ulaşılabilen *S. officinalis* genom assembly sekansını temsil etmektedir.

*S. officinalis* exomunda tespit edilen tek nükleotid polimorfizmleri aynı zamanda varyantlar olarak adlandırılmaktadır. Sonuç olarak değişken nükleotid lokusu (<https://doi.org/10.6084/m9.figshare.24100155>, TotalExome\_substitutions.tabular) ve 20 baz yakınında mükemmel eşleşme gösteren 24,570 tek nükleotid polimorfizmi (<https://doi.org/10.6084/m9.figshare.24100155>, TotalExome\_gSNPs) tespit edilmiştir. Terpenoid biyosentez genlerinin exonlarındaki yer değişikliklerine benzer olarak, transisyon mutasyonları *S. officinalis* exomunda toplam nükleotid yer değişikliklerinin 65%'ini oluşturmaktadır (Çizelge 18).

Bu çalışma kapsamında enzim kodlama aktivitesine etki edecek yüksek potansiyele sahip SNP'lerin tanımlanması için özellikle varyant olarak adlandırılan exonlar hedef alınmıştır. Bu prensibin kanıtı olarak, rastgele seçilen ve tek nükleotid polimorfizmine sahip terpenoid biyosentez genleri için InterProScan (Paysan-Lafosse vd., 2023) analizi gerçekleştirilmiştir. Dört gende gerçekleştirilen peptid domain analiz sonuçları şekil 26'da gösterilmiştir. Mevalonat yolak geninin N-terminal hydroxymethylglutaryl-CoA sentaz domaini kromozom 1 üzerinde 13. ve 16. amino asitleri kodlayan peptitlerin arasında konumlanmıştır. Gen exonunda bulunan iki tek nükleotid polimorfizmi tanımlanmıştır ve kodlayan sekansın 91. pozisyonunda peptidin



**Şekil 26.** Değişken terpenoid biyosentez genlerinin bir setinde domain analizi. Peptitlerin domain tanımları ve konumları her bir grafiğin sol üst kısmında belirtilmiştir. Değişkenlik bölgeleri ve nükleotid yer değişiklikleri de ayrıca belirtilmiştir.

- Kromozom 1’de tanımlanan MVA yolak geni. Belirlenen genin ID’si 01\_RagTag.g10549 olup, tezde ek dosyaların sağlandığı linklerden gene ulaşılabilir. Hydroxymethylglutaryl-CoA domaini 13-163 bölgeleri arasında tanımlanmaktadır. Değişkenlik bölgesi konumu 31.
- Kromozom 2 üzerinde tanımlanan terpen sentaz geni. Belirlenen gen ID’si 02\_RagTag.g24179 ve tezde ek dosyaların sağlandığı linkten ulaşılabilir. Terpen siklaz benzeri 1 domaini 177-503 bölgeleri arasında tanımlanmıştır. Değişkenlik bölgesi konumu 455.
- Kromozom 3 üzerinde tanımlanan terpenoid biyosentez geni. Belirlenen gen ID’si 03\_RagTag.g36721 ve tezde ek dosyaların sağlandığı linkten ulaşılabilir. Geranylgeranyl diphosphate reductase-benzeri domaini 45-436 bölgeleri arasında tanımlanmıştır. Değişkenlik bölgesi konumu 219 & 283.
- Kromozom 4 üzerinde tanımlanan MEP yolak geni. Belirlenen gen ID’si 04\_RagTag.g42863 ve tezde ek dosyaların sağlandığı linkten ulaşılabilir. 4-hydroxy-3-methylbut-2-enyl diphosphate synthase-benzeri domaini 77-740 bölgeleri arasında tanımlanmıştır. Değişkenlik bölgesi konumu 184, 276, 301, 507 & 525.

Asparajin amino asidi ile eşleşen 31. bölgesinin Aspartat bölgesi ile değiştiği SNP tanımlanmıştır (Şekil 26a). Benzer şekilde kromozom 2 üzerinde bulunan terpen sentazın e C-terminal 340 terpene cyclase like 1 domainini kodlayan sekansta bulunan dört SNP'in her biri bölge 455'te Treonin/Arjinin değişimine sebep olmaktadır (Şekil 26b). Terpenoid biyosentez geni olduğu varsayılan gende geranyl difosfat redüktaz kodlayan sekansta konumlanmış dört SNP tespit edilmiştir. Dört SNP haricinde benzer olmayan iki tanesi, bölge 219 ve 283'de bulunan redüktaz domainini değiştirmektedir (Şekil 26c). MEP yolak geni *HADS1*'de, 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase domaininde ve 184, 276, 301, 507 ve 525 pozisyonlarında bölgeleri değiştiren, beş amino asit değişikliğine sebep olan SNP ve altı değişken lokus tespit edilmiştir (Şekil 26d).

#### 4.6. *S. officinalis*'in Protein Kodlayan Gen Bölgelerinde SSR Tespiti

Sekanslanmış olan *Salvia officinalis* genomunun protein kodlayan gen bölgelerinde SSR tespiti için GMATA programı kullanılmıştır. Girdi olarak 20689 protein kodlayan sekans programa verilmiştir. Sonuç olarak adaçayının protein kodlayan sekanslarında 2452 SSR lokusu tespit edilmiştir. Bu SSR lokuslarından 2195 tanesine primer tasarlanırken, 257 tanesine primer tasarlanamamıştır. Benzersiz markör sayısı 2125 olarak tespit edilmiştir (Çizelge 19).

**Çizelge 19.** Tespit edilen SSR'ların istatistikleri

<b>Tespit edilen SSR'ların istatistikleri</b>	
Veri olarak girilen toplam sekans sayısı	20689
Tespit edilen toplam SSR lokusu	2452
Primer tasarlanan toplam SSR lokusu	2195
Primer tasarlanmayan toplam SSR lokusu	257
Benzersiz markör sayısı	2125

En az 2 en fazla 6 tekrar olmak üzere analiz edilen SSR lokuslarında, toplam 1500 SSR lokusu ile en fazla tekrar eden motif 3mer olarak tespit edilmiştir. En az tekrar eden motif ise toplam 2 SSR lokusu ile 4mer olarak tespit edilmiştir (Çizelge 20).

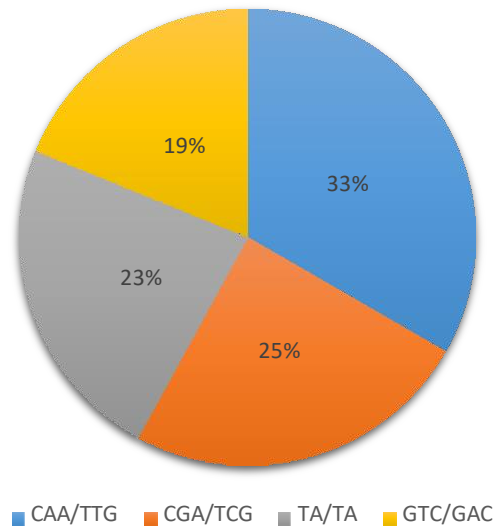
**Çizelge 20.** Tekrar motiflerinin SSR lokus sayıları

Motif (-mer)	Toplam
2	866
3	1500
4	2
5	5
6	79

Elde edilen analiz sonuçlarına göre 24 SSR ile en fazla bulunan SSR motifi ‘TCC’ ve ‘GT’ olarak bulunmuştur. En çok tekrar eden motif grupları ise ‘CAA/TTG, CGA/TCG, TA/TA ve GTC/GAC’ olarak sıralanmıştır. Sonuçların bir kısmı Çizelge 21’de ve Şekil 27’de gösterilmiştir. Çizelge ve şekilde göstermek amacıyla, 0.5’ten yüksek frekansa sahip motifler seçilmiştir.

**Çizelge 21.** En çok tekrar eden SSR motifleri

SSR motifi	SSR’ların sayısı
TCC	24
GT	24
CCT	21
GGT	20
CA	20
TGG	19
GCT	18
TGC	18
CAA	18
AGG	18

**Şekil 27.** Adaçayı exomunda  $\geq 0.5$  frekansa sahip, en çok tekrar eden motif grupları ve yüzdeleri

#### 4.7. *S. officinalis* Exomunda Terpenoid Biyosentezinde Rol Aldığı Tespit Edilen Genlerde SSR Tespiti

*S. officinalis* exomunda terpenoid biyosentezinde rol aldığı tespit edilen genlerde SSR tespit etmek amacıyla toplam 178 sekans GMATA programına girdi olarak verilmiştir. Bu sekansların düzenlendikten sonraki uzunluğu 221352 baz çifti olarak belirlenmiştir. Yapılan analiz sonucunda toplam 93 SSR lokusu tespit edilmiştir. Bu SSR lokuslarından 85 tanesine primer tasarlanmış, 8 tanesine tasarlanmamıştır. Tespit edilen toplam benzersiz markör sayısı 85'tir (Çizelge 22). Terpenoid yolağında tespit edilen SSR markörlerine ait veriler (Çizelge 23) görülmektedir.

**Çizelge 22.** Terpenoid biyosentezinde rol aldığı tespit edilen genlerde tespit edilen SSR'ların istatistikleri

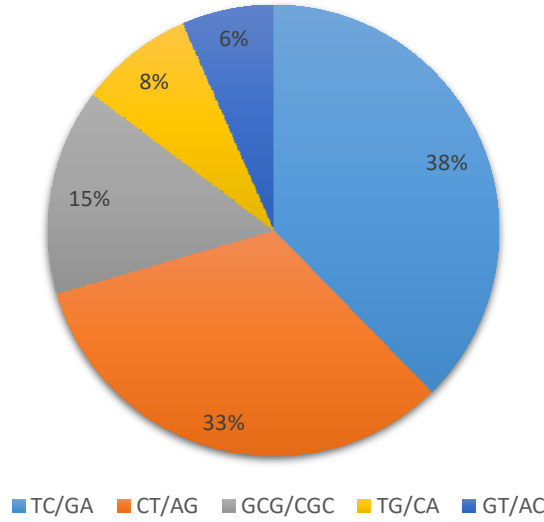
<b>Tespit edilen SSR'ların istatistikleri</b>	
Veri olarak girilen toplam sekans sayısı	178
Tespit edilen toplam SSR lokusu	93
Primer tasarlanan toplam SSR lokusu	85
Primer tasarlanmayan toplam SSR lokusu	8
Benzersiz markör sayısı	85

Tekrar eden motifler arasında toplam 61 SSR lokusu ile en fazla tekrar eden motif 2mer olarak tespit edilirken, en az tekrar eden motif ise toplam 1 SSR lokusu ile 6mer olarak tespit edilmiştir (Çizelge 23).

**Çizelge 23.** Tekrar motiflerinin SSR sayıları

<b>Motif (-mer)</b>	<b>Toplam</b>
2	61
3	31
6	1

Elde edilen analiz sonuçlarına göre, adaçayında protein kodlayan gen bölgelerinde terpenoid biyosentezinde rol aldığı tespit edilen gen bölgelerinde, en çok tekrar eden motif grupları 'TC/GA, CT/AG, GCG/CGC, TG/CA ve GT/AC' olarak tespit edilmiştir. Sonuçların bir kısmı Şekil 28'de gösterilmiştir.



**Şekil 28.** Terpenoid biyosentezinde rol aldığı tespit edilen genlerde en çok tekrar eden motif grupları ve yüzdeleri

**Çizelge 23.** Adaçayında terpenoid biyosentezinde görev aldığı tahmin edilen genlerde analiz sonucu tespit edilen SSR markörlerin primerleri ve ürün boyutu

Marker	Primer Forward	Primer Reverse	Ürün Boyutu
>SoTSSR1	GTTCGAGCCAGAGAAATCGT	TATTGATGCTGCTCCTGCTG	143
>SoTSSR2	GCATGGATGAGCAGCAGTAA	TCGTCTTGATCATGCACCTG	116
>SoTSSR3	TTGTTTTCTACGGGAGTGG	CAACTCATGCCTCGACTTCA	135
>SoTSSR4	TGGAGGTTGAAGATTGCAGA	TTCGACTTGAGCTCGTTCCT	135
>SoTSSR5	ATTTCCGGTGAGCCATTTCTG	CTGCTGCTGTTGTCCTCGTA	103
>SoTSSR6	TGAGTAGATGGCCGTTTTGC	CATCCAATGCCGAGGTAGTT	103
>SoTSSR7	GCATAAGTTGGGCGTGGTAG	GCTTCTCTCCCGTGAACTG	128
>SoTSSR8	GCGCTGGTGGAAGATATCAG	ATGACCCATCCGCCAACT	91
>SoTSSR9	GCTTGGGCTACAACAGCAAT	CCCTCCATTTGATTCTTGA	123
>SoTSSR10	CGGAGAAAGCTTCGTCTCAC	CCCCAAAAGCTCCATGATT	140
>SoTSSR11	ATTTGCCACCAAATTTCTGC	TGAATTGGGAGATCCAAAGC	102
>SoTSSR12	AACATTGGCGACACCTTCAT	AGTGCTTCTGCGTGAActCC	124
>SoTSSR13	CTTTGGTGTATTGCCCTGGT	CCAAAACATAACGCCAGCTAA	118
>SoTSSR14	GGTGAAGGAGGATTCACCAA	CTCATCCTGCTCATCGACAA	124
>SoTSSR15	CCAATGCCGTCTTCTTCACT	GATGAATGAGGCGACGAAAG	146
>SoTSSR16	ACCAACAGAGGATGCAAAGC	TTCGAGGAACATCTTCAGCTC	135
>SoTSSR17	TTCACCAACGCCTTCTTCTT	TGGAGTTGCGGATCTTCTCT	85
>SoTSSR18	TCAAGGAGGATTCACCGAAA	AGGGCATAACGAAGGGATCTT	142
>SoTSSR19	CCTCTCTCACAGTCGGGAAC	CGCCGATATGGATTTGGAG	144
>SoTSSR20	GTAAAGGCATTTGCCCTCT	TTCAACACCATAACGACGAA	116
>SoTSSR21	CAAGACCAAGAAAGGCGAAG	TGCTGTCATCATATTCACCTCA	118
>SoTSSR22	AAGCCAAATTCCTGCATGTC	CTCGATTCAGCTTTCCTCCA	138
>SoTSSR23	TGGAGGAGACGAAATCTAGC	CGGCAATCGTAAAACCTCGAA	88
>SoTSSR24	GGAGACGACTCGACAGGTTT	ATCCTCCAATCCCTTCTCGT	120
>SoTSSR25	GGGATTGTTGTGGTCAGCTT	ATCAATACTCCGCCCATCAC	81

>SoTSSR26	GGCCTCCGATCAGAAGAGA	CGGGTGATGAGGAAGGTTT	138
>SoTSSR27	CTCAATCGGAGGCAATTCTC	GGTCGCTACCGTGAATCAGT	123
>SoTSSR28	TGCGTTCTCCTCTCCAGAAT	TCTTTCAAAC TTCTCGGGTTG	137
>SoTSSR29	AGGCGAAGAGAGAGCTGGAT	GTAATCCTCCAACCCAATCG	148
>SoTSSR30	TCCGAAGCGAACAAACTACC	CCGTTTGGATGTGACCATTA	134
>SoTSSR31	ATGAAGGAGGCCAGGAAGAC	ACCCTCTTCTTCTCGGTGGT	128
>SoTSSR32	ACCCTCTTCTTCTCGGTGGT	ATGAAGGAGGCCAGGAAGAC	128
>SoTSSR33	CACTCTCGGAAAGGATGGAC	CCACCTGGTTGCAAGAGTTC	93
>SoTSSR34	AGGAGACGTACCGAAAGCAA	GCTGTTGTGACCGTGTTTCAT	126
>SoTSSR35	TGGAGCTGATTGATGACTTGA	TCTGAATCCAAGAGCTGTGAAA	140
>SoTSSR36	TTGCTGATGATCTTGGGACA	CGCCTCCCTTATCAGAAACC	137
>SoTSSR37	CCGTTTGAGCTGAAGAGAGG	CAGAAACCTCACGTGTTCCA	102
>SoTSSR38	CTGGGACTAACTTATTTCTTTCAGG	CTTGGGAAGCATTAAAACCA	148
>SoTSSR39	CCAAAAGCAATCCAGTGCTAC	CCGTGTTTCATCTCCTTCCAC	106
>SoTSSR40	AAGAACTACCACGCCTCGAA	GCTTGGAGTTGGAATTGAGC	139
>SoTSSR41	GCTGAACTCCACAAAAGGAAA	AAGAGGCCCTAACGAAAGT	150
>SoTSSR42	GCGAGGATACGAGAGGAATG	ATCCAAAGCATGACACACCA	149
>SoTSSR43	CATCTTGCTCGCATACTGA	AACCTGCGAAGCTTTGAGAA	147
>SoTSSR44	GGAGTACATCCACTTGCACA	TCCACCACCTGAAACATCAA	144
>SoTSSR45	AAACTCAATCATGCCGGAAC	TCTCGCCGGAGAAATCTACT	130
>SoTSSR46	ATCTCCCACCTCTTCAACG	AATGAGGGGCGGAATTTAGT	100
>SoTSSR47	GCCACAATCAAGAAGGGAAA	CAGGAGTGTCTGTTGCAAGG	140
>SoTSSR48	AGCCAAAAGAGGTGGGATG	TGATCTCCGGTGGTTCAAGT	126
>SoTSSR49	GCGGGGATGTACAAGGTTTA	ATGACTTCCTTCCCATGAGC	145
>SoTSSR50	TCAAGAACTTTGCTCCGACT	CCCTTATCTTTGGCACTGAAG	112
>SoTSSR51	CGGACCAGATGCAGATCCT	CCGAAGTCCAAGTTTTCCAG	133
>SoTSSR52	GATGCTTTCCGAGCTTCATC	GCGCCGTAGTAGTTTTGTTG	131
>SoTSSR53	TGCAAATGCAGAGGCTAAGA	GAGATGGAGTACGGGAAGCA	133
>SoTSSR54	TAGCAGCAAAGAAAGCCACA	GTGAGGAGGCCACTCATTCT	106
>SoTSSR55	TCATGATGGGCGATGTGTAT	CACAATCTTCAAGCCCAAT	122
>SoTSSR56	AAAGCTCGTCCCTCACAAGA	GGGGAGACTAGGGTTTCTGC	124
>SoTSSR57	AGTTGCTGGGTCAGTCTTGG	AGTTTCGTGATGTCCCTGCT	114
>SoTSSR58	CAGCTGCTTTTATTGCCTCA	ATATCGACTGATGCCCAACC	149
>SoTSSR59	CCTCAATGACATCCGTCTG	TTTGTGCGTCGACCATACTT	140
>SoTSSR60	TCTCGCTTAGCTCATCCTG	TGAGGAAAGAGGCAAGGAAC	148
>SoTSSR61	TTGTCGATGGAATGCTGAAA	GCAATTCACGCTCAAAACT	148
>SoTSSR62	TGGATTTCAACCTTTTGCAG	CACCACCTCCAATTTCTTCC	80
>SoTSSR63	CATGGGGAGGAAATCTTGA	GTGAGGCTCTTTCGATTTGG	137
>SoTSSR64	TGGATTTCAACCTTGTGCAG	GATGTATGGTGGCGAATCCT	98
>SoTSSR65	AGAGTCGAAGCGCACAGTTT	CCTCTTTGTGCATCATTTGC	127
>SoTSSR66	GGATTGGCCTACCATTTTGA	GTGGCGGTGTTGTCTGAGTA	135
>SoTSSR67	TAAAGGAAACGGTGCAACAA	TGCTCGTTATCTCCGAGTTTT	114
>SoTSSR68	GCACAACATGACCACGACTT	GGGGTACTCTGAGCAAACGA	137
>SoTSSR69	GTTTGCAGAAAGGTTGAGG	ATAGGACGGCGGTTTCTTTT	95
>SoTSSR70	TCGACCTCACCACTCTCTCA	TCCTCCTCTTCCACGAGCTA	132
>SoTSSR71	CCTGCCCAAGCATGAAATAC	GCTTCGCTGCTTTGTGTAAC	135
>SoTSSR72	GGGCACATAAAGACATGCAA	TGAAGAACCCAGCTCTCTG	127
>SoTSSR73	GCTCCACTCGAATCAAAACC	TGCACGATTGACTCATTTCC	143

>SoTSSR74	TGAACCATCCCGAGTACCAT	TGCGTTGAGTTTACCATCTTTG	110
>SoTSSR75	TCGACTTCAAGGGTTACATGG	CGACTCGTGGATGAGAAGC	98
>SoTSSR76	ATTCAGGCTTCTGAGGTTGC	AAGATCGGCGAGACTGAAGA	81
>SoTSSR77	GAGGATGCTAAACCCGACAG	AAACGGGTTCTGCTATCAC	147
>SoTSSR78	GGTTTCTGGGGTTCCATTG	TTCTGCCACTGTTTTGTGGA	90
>SoTSSR79	AGCCGTTTCTCTGTCCTCTG	AACAATTGGGTGGTTTGGAG	115
>SoTSSR80	AGGCACGGCTACAGAAGAAA	GTGCTTTGCTCGCAACTCTT	128
>SoTSSR81	GTGGACAGACCAATGGGAAT	GCCCCAAAAGTATCCTCTCC	139
>SoTSSR82	CTTTTGGGGCAAGAGATGAA	TAACTCGGCTTCGTCTGGAT	101
>SoTSSR83	TCAATGATATCCGTGGTGATG	CGAAACGGCAGCTTCTTTAC	110
>SoTSSR84	CACTCGACCTCAAGATGCTG	AACTGGATCTTACCGGAAA	116
>SoTSSR85	CGAGTTCGACCACGTCAT	CGCTCGCCATGCAAATTA	131

---

## 5. SONUÇLAR VE ÖNERİLER

SNP markörleri yüksek genom kapsamı, kodlayan sekanslarda sıklık, genotiplemede üstün doğruluk ve yeniden üretilebilirlik gibi avantajlarından dolayı bitki ıslah ve fonksiyonel genomik çalışmaları diğer markör sistemlerinin yerine geçiyormuş gibi görünmektedir. SNP'lerin bir diğer çok faydalı özelliği ise bitki genomu ve moleküler ıslah için, sekanslama teknolojilerinin gelişmesi ile birlikte, geniş popülasyonlarda kısa ampikon sekanslama ile alel tespitine olanak sağlamasıdır (Eriksson vd., 2020). Tez kapsamında yürütülen çalışmada genomun sekanslanması ile adaçayı terpenoid biyosentez genlerinin kodlayan sekanslarında tek nükleotid varyantlarını tespit etmek mümkün hale gelmiştir. *S. officinalis*'in başlıca esansiyel yağ bileşenlerinin genotip alel ilişkisi daha önce araştırılmış ve gösterilmiştir (Jug-Dujaković vd., 2012). Bu çalışma, terpenoid üretiminin karmaşık doğasının aksine, arzu edilen bir özellik için çeşit geliştirmek yerine genomik seçim yaklaşımının mümkün olduğunu göstermektedir. Bu sayede yeni çeşitler endüstrinin ihtiyaç duyduğu özelliklere göre, adaçayı terpenoidleri ham materyal olarak kullanılarak oluşturulabilir. Bu gerçek göz önüne alındığında yüksek oranda çeşitlilik gösteren terpenoid bileşenleri, antikanser ilaçlardan biyoyakıtlara, farklı endüstrilerde ham materyal olarak geniş kullanım aralığına sahip olmaktadır. Biyosentez genlerindeki alellik varyasyonlar, değerli potansiyel moleküler araç haline gelmektedir. Tez kapsamında gerçekleştirilen de novo genom birleştirme sonuçları ve tüm genom anotasyon verileri aynı zamanda farklı türlerdeki bitki genomlarının tüm genom sekans verilerinin anotasyonu ve anlamlandırılması çalışmalarında kullanılabilecek yeni biyolojik büyük veriler üretilmiştir. Bu veri setleri karşılaştırmalı genomik, filogenomik ve fonksiyonel genomik çalışmalarında araştırmacılara önemli bir kaynak olacaktır.

**EKLER**

**EK-1** Adaçayı exomunda analiz sonucu tespit edilen SSR markörlerin primerleri ve ürün boyutu

Marker	Primer Forward	Primer Reverse	Ürün Boyutu
>SoSSR1	AGAAGCTGAACGAAAAGCAA	GATCTCCCCAGGATCTCTTC	111
>SoSSR2	ATGGGTGTCGAAGATCACAT	ATCGGATGAATTGCTGAAGA	174
>SoSSR3	AAACCTGTGGTTCGTGCTAA	TCATCATCACCGTCATCTTCT	191
>SoSSR4	ATGCTGCGGAACACTGCT	ACCCGGATTTCTTCTTCTTG	142
>SoSSR5	CGAGCTCTCCAAATACCAGA	GACATATTCATGGCGAGGAG	216
>SoSSR6	CGATTCTTCTCGGATAGTG	CGGTGCTCTCGTCACTCT	128
>SoSSR7	CCTCTCCACAGCAAGTAGA	TGTGAATGAAGGAAGCAACA	115
>SoSSR8	GCAGGAAGAAGGTGAATGAA	GGAATCGCCAACTGTAGTC	192
>SoSSR9	CTTTTGAGCAGAAACCGATG	ACGAGCTCACACTCTTCGAC	238
>SoSSR10	GGCATTTCGCCAAATATCAC	CGTGGCTTCAATTGGTTCTA	232
>SoSSR11	GCAAGATGTTTTGAGGGCTA	GTTTGTGTTGGACGAATCCTG	250
>SoSSR12	AATGCCTCATTGTGTTGGAT	CTGGTGCTGTCCAATCTTTC	155
>SoSSR13	TGGTGATAGGGTACGTGGAC	GACCCGCATTTCTTATACCC	246
>SoSSR14	CCTCTGCAGAGCAACTAGA	AGGGCTCTTGAGGCATATTT	169
>SoSSR15	GAAGATTTTGCCGAATTGGT	ATCATCCTCTCTGCAGTTGG	132
>SoSSR16	TTTTGCCGAATTGGTAGATG	ATCATCCTCTCTGCAGTTGG	127
>SoSSR17	CTTAATAGGGTTGGGGTTGG	TGCGTAGACCATCAAGCATA	157
>SoSSR18	GGTAACAAATGCCTCGTTGT	TTCATTCATAAACCGCACCT	192
>SoSSR19	TATCACCGGATAGACCTTGG	GCTCCTGCTTATTCTGCTTG	167
>SoSSR20	GATGGGTTGAACAAGAATCG	CATTTTCACACTCAGCAGCA	246
>SoSSR21	CTGTTTTCCCAAGCAGAGAG	CACATCAAACCGTTTCTTC	161
>SoSSR22	AAATGGGCGTCAGTTACAAA	TCCCAATCCCACAACATATC	208
>SoSSR23	GCAGACGGTTCCATCAATTA	GTCTCTATCCTGGCCTCTC	241
>SoSSR24	ACAGCCAATTCCAAACGTC	GAAAATCCGGAAGGGACTAA	232
>SoSSR25	TCATTCCTACTGACGGGATT	TAAACTGGAGGCGTTTCTTG	196
>SoSSR26	ACGGATTGTGTTGTGCGAGTT	CCTCTCTCACCTTCTCTC	207
>SoSSR27	TGGCAGTTTCTTCTTTGCTT	TATGAGCTCGCTCTTGAGGT	134
>SoSSR28	ATGTGGCTTCCCACCTAACA	TCCTGCATGGCTCATAGAAT	233
>SoSSR29	TGCACCTTCTTCTTTGCTTC	AGAGCCTCAAACGGATAACA	236
>SoSSR30	GCTGCAGTTTCTTCTTTGCT	GCTCCTGAGGTCGAGTAACA	111
>SoSSR31	TAGCTGACATCGCCTCTTCT	ATCATCGGCTTTGACACACT	250
>SoSSR32	AGAGCCATCGCAAATGAA	TCTCCTAGATGCCACTCCTG	198
>SoSSR33	CCACCGACTTCAACATCAA	GGCTTGGCTTTACCAACAC	226
>SoSSR34	TTCTGCATTCCCTTCTTTTG	AACAACAGCTTCAGGGTCAG	105
>SoSSR35	GACGAGTCAGGATCCCTCTT	CGAATTCTTCGAGCCTACAG	100
>SoSSR36	CGGACTGGATTGGTGAACATA	TCTCAGCAGGCAACTCTACC	141
>SoSSR37	TGGCAGTCTCTTTGCTTCTC	TGATTCCAATTCCCAAATA	157
>SoSSR38	TTCTCATCAAAGGGCTGTTC	TTTGGCTTTGTCCATCTCTC	194
>SoSSR39	TGGGGAGATAGAGTGGGAAT	TTCTTCTCCTTGCTCTTTGG	164
>SoSSR40	CGTCCAAATCGTTTCCT	GCATAAGCCTCGCATAGAGA	180
>SoSSR41	GAAGTGTGGATAAGGGAGCA	GGAAGTTGGTTGCACGATAG	162

>SoSSR42	CCGGATCAGTCAGATACCAC	TTTGCTTCTCCTTGGTTGTC	179
>SoSSR43	CGGAGAAGGAGATGATGAAA	TAGCAATAATTTCGGGTTTCG	234
>SoSSR44	GGAGGTCAAATGAGAGACGA	GTACCAGCAAATCCCCTTTT	197
>SoSSR45	CAATTGGAGAGCCCTATTCC	CTTCCCCTGCAATCTTGTTA	244
>SoSSR46	AACGCTACTGCTTCTTACGG	GCGGAAAACAGATTCAAAGA	207
>SoSSR47	TCTCTTCTCTCCCGGAAT	CTTTCCTCCGGATACGAAAT	245
>SoSSR48	CAAACCAAACGATCCAAAG	CTGTGTTTGGTGGTGGAGAT	164
>SoSSR49	AGCAAAGCCAATCATTTCTG	AACCAAGTGTAGGGTGGTGA	212
>SoSSR50	CAGGAACCGGACTTCTGTAA	CAACGATGTGGTAGGAAACC	213
>SoSSR51	ATGACTCCAGTGTTGGGAGA	GAAGCAGAGAAGCCCACATA	133
>SoSSR52	GCTATTTTGGTGAAGCCTGA	GGTGGAGGAGGTAGTCCATT	164
>SoSSR53	ATAGCCAGATTGAGCCACAG	CAATTCTCCACCAATTCAGG	172
>SoSSR54	AAACCTCCCACCATGAAACT	CCCTAGTTTCCGATCCTGTT	213
>SoSSR55	AGTGGAGAGAGGAGCAATGA	TTACCGTTAGGACCCCATTC	172
>SoSSR56	GCAGATCAGTGGAGAGAGGA	TACTGTCAGCACCCCATCT	160
>SoSSR57	TCTCTTGAACGGAAGTTTGC	CCATCTCGTGGTCCAATTAC	151
>SoSSR58	CAACCACCAGAAAAGGAAAC	GCTGGGGACACTTAGGGTAG	212
>SoSSR59	TCAACCAAGCCACCTAAAGA	TCAGGTGCCTCGTATCTCTC	218
>SoSSR60	GTACCACAAAACCCACGAC	TTCTCCATCACATCCTCCAT	228
>SoSSR61	TTACTAAACGGAGCCAGAG	TCCTCAGCTTGTTCACTTCC	178
>SoSSR62	GTGAACACGCTGGTGAAAG	TCAAGCCACTATCTGCTCCT	198
>SoSSR63	GAATTGGGCTTTGAGAGTGA	TGGCAGCCCTTATATCAGAC	140
>SoSSR64	GTCGGATTTGAGGAAAGACC	AGCAAGCCTTCCATCTCTCT	127
>SoSSR65	GTCGGATTTGAGGAAAGACC	CAAATGCCCTCTTCTCTCTG	236
>SoSSR66	GGAGGAGTACGAGGTTGGTC	GCAATAATGGCTTTTCCTGA	169
>SoSSR67	CCTCTCCGTCAATCTATCCA	GCAATCTATCGGACTCTTC	243
>SoSSR68	CACCACCCTATTTCCATCC	TCGTCGTCATCATTGTCATC	146
>SoSSR69	ATCGAAGAGCTGCAAGAAAA	TTCCAGTTGAGTCCAAAACC	217
>SoSSR70	CACCATCTCCACAAAACACA	AGTCTTCCCTGCGTCACAATC	156
>SoSSR71	ATTCATCTTCGAGCTTCGTG	TTCTGTTCCTCTTCCTCTCT	149
>SoSSR72	GAGGAGGAAGAGGAACAGGA	GCACCCTTTGTTCATCAC	167
>SoSSR73	AGAGAGAAAGCAGCAGTGGA	TTGTTGTTGACCATGTTTGG	185
>SoSSR74	CACCAGCACCGTCTTCTAC	GGGCAGCACTCATAATCATC	164
>SoSSR75	TGAAGGGCTAAGATTTGTGG	ATCGTCAGTTTGCATGAGGT	223
>SoSSR76	CGTCTCAAAATGAGGGAGAA	TCCTAACGAACGAAGTGGAG	219
>SoSSR77	GTTTGATTGCATCTCCATCC	CAAGGAGAGCTTGTTCCTGG	145
>SoSSR78	GCCCTTTAGCAATGAAATGA	CTTCCCCTCCAACATCCT	157
>SoSSR79	ATTCAGCAAATTGGAACACC	TGCATCGTACCTGAGCCTAT	194
>SoSSR80	AGCAAGCAATTCAAGGACAC	CGGTGGGGTAATTGTTGTAG	214
>SoSSR81	CAGAAGAAACGCCTTGTGAT	TGACATTCTTACCCGAGTT	241
>SoSSR82	TTTCGCCATGTCTCAATTC	ACCACGATCTCTCAATCCAA	190
>SoSSR83	GAGACGTGAGGGTGTATTG	ATCATCCCTCTTACATCCA	110
>SoSSR84	TCCAAAATCCTCCTCCTCTC	CTCTTGCAAAGTGAGCCTTC	180
>SoSSR85	AGCAATCGGAGCTGAAGTC	ACACACTCCACTCCCTGAAA	225
>SoSSR86	CTCCCCTATTTCCCCTCATCT	GTCGTGTGGGATTTTGTGAG	112
>SoSSR87	GTCAAGATTGCCGTTTGTGT	GGGCTGTAATCGTCAGAGAA	180
>SoSSR88	CCGTCATCTCAAATCTCCAC	GGCAGAGGTAAGCTGAACAA	189
>SoSSR89	TGGTGAAGGCAATCAAGAAT	ATCCATGTTTCGAGATGCAGT	227

>SoSSR90	TCAAATGGGAAGGTAGTGGA	GTTGAGGTCGATCAGCAACT	243
>SoSSR91	GCAAGGAGATGCGATACACT	TCTTTACCGAGCCACATAG	214
>SoSSR92	CGAATTGGACAAGAAATTGG	AATGCATAAGCCTTCCACA	249
>SoSSR93	CTGTCTTGCGCACTCTGATA	GCTCACCAGCATAGGGTAGA	197
>SoSSR94	GAGGTTTTGTACTCGCAGGA	TTGGTGTCGTTGGATTTTCT	194
>SoSSR95	ATTGAGTTGGGATGTCGATG	GCCTCCTTCTCATTGCTTTC	231
>SoSSR96	GTGCCGTAGAAATCATCACC	CTTCCTCTTCCACGACAAGA	219
>SoSSR97	AACTCGCGTATCAGCACATT	CCTGACATCTTCGTCTGCTT	170
>SoSSR98	GCTCTTGAGTTGGAGGATGA	CTTGAGTGGAAGCCTGAAGA	142
>SoSSR99	GATGAGTCGCCGATAAACC	TCTTGCGTTTTCTTTGAAGC	239
>SoSSR100	CAGAGTGTGACCGTGTGTGT	CATGAACCTCTTCCACAACC	225
>SoSSR101	AATGGAGCTAGTGGTGGTCA	CGGAAGTTCTTCTTCTGTGC	152
>SoSSR102	AACAGCAACAACAGCAACAA	TACTCGCAGCCACCTTACTC	189
>SoSSR103	GGACTCTGGTTTTCAAGCAC	TTTTGGTTGTTTTCCACCTC	129
>SoSSR104	ATGTGCAAAGGGAGAGAGG	TGATTTCCCAATCCATTCC	110
>SoSSR105	CAAATCCCGATGGGAGAA	GGATCGTTGAAAGGGAAGAT	216
>SoSSR106	GTCAAGCCTCATGTGGAAGA	GTGGAGGAGGAGAGGAAGTC	237
>SoSSR107	GGCTGAAGATGTGATCGAGT	CTGCTGAGCTTCCATCATT	226
>SoSSR108	AGAGTGGTGAGGATGGTGAA	ATCCTCCAGTGAGCCTATCC	180
>SoSSR109	ATGTGTGTCAACCCCAAACCT	CATCGTACTGACGCTGACAA	196
>SoSSR110	TACAACCTCCAGTCGCCTCTC	GAGCCCGAGAACTCCTAGAC	155
>SoSSR111	ATTCGAACCTACTCGGCCTTT	TTCACTGTCACCAACGACAC	248
>SoSSR112	GGTATGAGATTGCCAGCCTA	AGAGCTTCCCTTTGGACCATC	230
>SoSSR113	AATAGAGGGGAGGCAGTGAT	CTAGCGCAGCAACCTGTAGT	214
>SoSSR114	CGAGATGGAAGAAGTGTTC	CTGAGCCACTGCATTCTTTT	240
>SoSSR115	AACAACCCGGTAGCTCTCTT	TGATCCCCAAAATGATCCTA	243
>SoSSR116	AGTGAGATCGGTGATGAGGA	GAAGGCTATTATTGGCAGCA	193
>SoSSR117	CTGCTGCTCCTTCAACTCAT	CATCATTCCAAACGAAAAGC	224
>SoSSR118	GGATTACAGCAGTAGGAAATC	GCTGCCACTTCTTCATAACC	245
>SoSSR119	AAGCTTCCTGTTCCTGTTC	CACATTGTATCGAGCGAACTC	250
>SoSSR120	AGGACGTTGCTTTCCTAGT	ATAACGCCCTCGAGTAAAC	196
>SoSSR121	TAAGGAGCTCGTGGAAGA	CGTCCAAACTCACCTTCATT	117
>SoSSR122	ATCGTCAAAGCTCGTGAAAG	TGAGGAAGCAGAGGAGAGTG	206
>SoSSR123	GGAAGCCCTTCTTCAAGTG	TTTGTCTTGCAAACCTTGG	211
>SoSSR124	CTTCCGCTTCTGCTCTCTC	TATACCCTTCTCCGCTTTG	245
>SoSSR125	TGAACCAGTGGGGAGCTA	GCGATTGGCTATTCTTCCAT	213
>SoSSR126	ATGAACCAGTGGGGAGCTA	GAGGGTTTCTGTTGAAAA	165
>SoSSR127	GCTACTACAAGGGCGAGGA	TTGACGAAATTCTGGGTCAT	208
>SoSSR128	CTCTTACCTACTGCCACCA	AACTGGTGGTTGTGATCCAG	165
>SoSSR129	TGGCAACATTGAGTACATGC	GCTCTTCCACCCTCTTCTT	175
>SoSSR130	AAAGAGGAAAGGTTGCGAAG	CTATCCCGGTAATCTTGCT	122
>SoSSR131	TCCAATCCGAGGTTTTATCA	TGCGGAGATGAGGAATTTTA	177
>SoSSR132	ATTAGGCACCACCACGAGTA	CGGTTGAAATTATTGGTTGC	189
>SoSSR133	GCGCGAAGGAAATTAGGTTA	TGTGAGGAGCGTCTTCTCT	162
>SoSSR134	AAGCAGATGGACTGGCATAAC	ACATGAAGCTTGCAATAGGC	188
>SoSSR135	CTTGAGGATGACGTGGTTTC	CCTCCAGATTTTGTGAGT	203
>SoSSR136	AGTTGAGACAGGTCCAATGC	GTATGGGCTGCATTTTGTTC	189
>SoSSR137	ACAAGCCCTTATTTCGGTTC	GAACAGCTCTCTTTGCTTGC	248

>SoSSR138	ATATCAGGGAAGCATGTGGA	TGGAGAACTTAGGCCTTTGA	154
>SoSSR139	TTTACGGTGGAGATCAGAGG	CACCATCACCAGATTTCTCC	204
>SoSSR140	TGTCAACCACCAGTCTCCTC	GAAATTCGGGGAAATGCTT	148
>SoSSR141	ATGCAAGATTTGGACATGGT	CCCACAAACACAATTCCTTC	205
>SoSSR142	GTTAGCGACAACGGGAACTA	TTCTTCGTCTTCACCGTCAT	179
>SoSSR143	CTAGGGGAGCGGTGTAGAGT	TATTCATCGTCGGGGTCTAA	158
>SoSSR144	GATTCAAATGAGGCAAGAGG	CTGTGCTCGATTTCTCCATT	212
>SoSSR145	TCATCGAGCAGGAAGTAAGG	TTCGACTGGGGACTACTCTG	229
>SoSSR146	GAAGAAGCGGTTTTGAGACA	TGCACGTTATGGTGTGGAC	219
>SoSSR147	GGCACCCAATAGAGACACAT	TTAGGGTTTCCTGGCTTAGG	164
>SoSSR148	GGGAAGATTCAGATTCCAT	GTTGCTGCTGAGGTTGAGAT	135
>SoSSR149	CGGCCACTCCAAGAGAGA	TAGGTGAGTGGTTTGGAGGA	240
>SoSSR150	AGAACTTGCAGCCAAATCAG	CAAGAGGGCAAGAGCAGTTA	222
>SoSSR151	TGTGCTTTGTTGTTGGTGTC	GATCCGACAAGTCGTCTACC	231
>SoSSR152	GCCCCAACTCCAACTCATA	AGAAATCGTTGCCGTAACAC	134
>SoSSR153	CGTCATGAAGAGAAGGCAAG	CTATGACGCCGTAGATGAGC	233
>SoSSR154	TTGTGAATCTGCATCTGTGG	CTCCTTCAGTAGGTCGCAAG	236
>SoSSR155	GAAAGAGGAGAGGCGGTTT	GTTAAATCCCACCACAGCAG	103
>SoSSR156	CTGCCCTTTATCCAATTCA	TTGGATCCCTTCAATATCCTC	226
>SoSSR157	AGCACTACGGCAACACCTAC	AAATGCGGCGACATATACA	208
>SoSSR158	CCAGCAGACCTTACCAAGAA	GAATCGAGGAATTCAAAGCA	231
>SoSSR159	AAAGGGTTC AAGAGGCCTAA	GACACCGACGAGATGAACTG	134
>SoSSR160	CTCTATCCGGAGGGAAATGT	CTGCGTCTCTCCACAACAA	244
>SoSSR161	AATATGGCTCCACAGTCCAA	GTAGACGATGAGGCCGTAGA	164
>SoSSR162	CGGCCTCATCGTCTACTCTA	GTACCACCTCCACCAGATTG	238
>SoSSR163	GAAAGAGGAAAAGCCAAAGG	GATTCCCTTCTCTCCCTCAGC	104
>SoSSR164	CGACCGATACTTCCAATCTG	GCTGCTCTGTAAAGCTCGAT	234
>SoSSR165	TTACGAGCTGATACCCTTCG	TTCTCCTTTGTGGTCTGCAT	151
>SoSSR166	GAAGATAGGCAGCGTAATGG	GATGAGAGCCATGAGCAACT	241
>SoSSR167	ATAACGCATGTTGGTGGTG	TTTTGAATCCCTGAAGCAAC	230
>SoSSR168	AGCCAGAGTGATGGAAGTTG	CTTCCTGAAGGTCAAGCAAA	170
>SoSSR169	GTTGCTCTAGGCATTGGAGA	TTGATGTGGTCAGCGACTTA	238
>SoSSR170	TCGTTTGTGTATGCGTATGG	TTGATGATGTGTGTGTGTGTG	158
>SoSSR171	TCACCAACAACAACAACAGC	GAAGGCTGAGGCTGAATAGG	180
>SoSSR172	AAAAGCAATCAAGGTGCTGA	CTTTATCTTCCCCGGTGTTT	175
>SoSSR173	ATCCTCAAGTTCTGCATTCCG	TGAAGAGATTGCTCGTCTCC	178
>SoSSR174	GAAAAGCGAATCAGA ACTGG	CGGTACAAAGCTTAGGTGGA	221
>SoSSR175	ACAGACGAGACAGCTCCAAC	CTTTTGAAACCGAGACCAA	247
>SoSSR176	ATCTCGTCAACGTGAAGCTC	GGAATCAGGGAGGTGAAACT	245
>SoSSR177	ATGGGATTGAGCAGTAGCAA	GACAAGAGGCAGTCTTTCCA	208
>SoSSR178	GATCTCGATCCCCAACACTC	GCGCTCCTTCTCCTTATCTC	180
>SoSSR179	CCTCTTCCTAATGCCCTCTC	GGATAGTCACGTACCCGATG	133
>SoSSR180	TCCGAGATATGCAGAGGATG	CCGAAGTCGAGATTCTTCT	249
>SoSSR181	CAGACGCCGGAGATAAAAC	AGTAGGATTTGGCCTGCTCT	185
>SoSSR182	ATCGACGAGGAAATCATCAG	GCGATGGAGAGAATCCAGTA	239
>SoSSR183	TCTGGCTTTTCATCCTCAAG	AGGCCGTGATTTCTTCTTCT	208
>SoSSR184	CATGTACGGACTCTGCTCCT	GCTCTCCTTGC ACTTGACAT	145
>SoSSR185	GTGGGGTTCGGAACAGTTGA	CTTGATAAATCCGTGGTTGG	171

>SoSSR186	GAGCAACAAAGAAAGCAGGA	CCTTCTTCCCAGGATTTTCAT	242
>SoSSR187	AGCGATCAATTCTGGTGATG	GCGTAGCATTCTTCTGTGGT	157
>SoSSR188	CCAAGCAACAAAAGAAGAAGG	GTTGAGGGAGGAGGAGTTGT	187
>SoSSR189	GCAAGGCATCAAAGGATATG	GGTTGTTTCAGCAGCTTCATT	207
>SoSSR190	GCTCTGCCTCGACTCTACAT	GGTGATCTGGTAATGCAAGC	238
>SoSSR191	ATCTTCCAAGAATCCGAAGG	GCATTAGCTTCAATGCGTTT	217
>SoSSR192	ACGAGATTGGTGTCCAAATG	ATTGCTCAGTGATGGATGGT	248
>SoSSR193	CCAAGCTGAATCTCCTCAA	TCGTTGTAGGGAGTGACGAT	124
>SoSSR194	GCACAAGCTGAATCTCCTCA	GTTGTAGGGAGTGACGATGG	115
>SoSSR195	GGAGGATGGTCGATTGTATG	ACATCCTTGGACTCGTTCAG	205
>SoSSR196	CAACACCCAAGACGAAGAAT	GTTCAAGAGTGGCTTCTCCA	200
>SoSSR197	GCTCGACCAGATTCGTACAG	TCCGACTTTTGCTTGTAACC	204
>SoSSR198	GTACGGTCAAATGGAGCATC	CCATCAACCTCTTCTGCAAC	180
>SoSSR199	ATCTTGCAAACACCAGCATT	CTAAGCCCATGTTAGCTCCA	216
>SoSSR200	TCTCTCCATTCAAATTCCA	AGAGCCCACCCTCTCTTCT	193
>SoSSR201	CAGCTGAAGAGGTTGTGGAT	GGATCAGTTTGCTGAAGTGG	124
>SoSSR202	TGGCTCGTCATCAATATTCTC	CTTCTCAGATTTGCCGCTAT	140
>SoSSR203	GACGAAGACGATGAGGAAGA	TGGTACGTGACTTCTGATGC	202
>SoSSR204	TATTCGCACCATGCAAACAT	CGTTGTAACGAGCACTGTTG	236
>SoSSR205	GTCGATCCAAGCAGGAGTAA	CAGCGATGGTGATAATGTGA	184
>SoSSR206	GGTGGCTACAAGATGAGTGG	GAGACGGCTTTGAGAGAGTG	217
>SoSSR207	AACTAAACGAGATGCCGATG	CATTGGGAAGCTCATAGTGG	212
>SoSSR208	GGCGCTGAAATCCTTCTTAT	GTTTACCAGCTTGGGGTTTT	199
>SoSSR209	TATCGATCCCAATGCATCTT	GCACTGTAGCCGAGAGAAAC	209
>SoSSR210	AGCAGAAGCAGAAGCAGAAG	CGGTGAAGAGAGAGATGAGG	162
>SoSSR211	TGAAAATGAAGGGTTTGGAG	CATTTTATCGCCCAATTCAG	239
>SoSSR212	GGAGATGGAGAAGATGCTGA	GGCTCAAATCGATCAGAAAA	177
>SoSSR213	ACGTTGATTGAGGATTGGAA	GAAGGGGTAGCAAAGAAAGC	182
>SoSSR214	TCCTCTTCCCTCCAATTC	TTCTGCTGTTGTGATTTCCA	244
>SoSSR215	ATGACACACCGAAAGAAAGC	GTGGTCTTCAGGCTTTTGA	213
>SoSSR216	AGATGTTCCGGTGGTTCTTC	GCCAACCTCTAGCTTTGGAT	133
>SoSSR217	CTGCCTTGTCCTTAGCAAGA	CTTCCAAGTAGCCCTCCACT	179
>SoSSR218	CTATTCACGCCTCCAAAATG	CCTCCCACGATCACTATGTC	122
>SoSSR219	CTCTCCTCTTCTCCTCCTG	TCATCATAGCAGCTTCGACA	249
>SoSSR220	GTCCGATTGACTGATGGTTC	GCACTCCTTTCTTCTCTCC	247
>SoSSR221	AGATTCGACAACCTGCTTGC	CGCTGACCGCATCTTATACT	187
>SoSSR222	CAAATCCTTACC GCATAAC	CGAGGGTCCAGTCGTAGTAA	184
>SoSSR223	AGGTGTCGAAAGAAATGACG	GAGGAACTTGAGCTCCAACA	245
>SoSSR224	ATTCCAATGCAAGTCAAAGG	TTGGGATTCTTGTGGTTTC	165
>SoSSR225	AAGAACCACATGCTCAGCTC	CTCGTACAGAGCCTCCAGTC	239
>SoSSR226	ACCATCTTCGTTCTCCATCA	GCGAGAGCCACAAGTTTAGA	185
>SoSSR227	TTGTCAATCACCTTCCTCGT	GCTTGTTCCTTTTGCTTCTG	189
>SoSSR228	ATGCCGCTGAATATGTGATT	TGAGCTGAGCCTATGGTTCT	244
>SoSSR229	CCGCTGATGAAGATGAAGAT	ACAGCATGGGTGGTACTGAT	221
>SoSSR230	TAATTATTGTGGCGGAGAGG	GAGCTGAGCCTGTGGTTCT	236
>SoSSR231	TGGAAAAGAGAGTGCCAAAG	CGACCAAATTGTCATCCTTC	197
>SoSSR232	GAGCGCCAAAAGCATAGTAG	ATGTCAGCCTCAGCAACTTC	141
>SoSSR233	TGGAAAAGAGAGTGCCAAAG	GACCCGACAAAACCTGTCATC	201

>SoSSR234	GGCGGGGATACTTCAATAGT	CAATCAGCATCCCTATCTGC	231
>SoSSR235	CAATAAACCCGACCTCAGC	AGACTTCCCCGGTGAAGAAG	242
>SoSSR236	ATAAAGGCGGCAAAATTAGG	ACCGCCTATCAATGTGCTTA	181
>SoSSR237	ATCAACGGCTGAGAGACAAC	TGCAGATCTCCTCTTTCCAC	245
>SoSSR238	CTACAAGTCGGAGCCATTTT	CTCCACTGTACATTCACCA	187
>SoSSR239	GAGCAAACATCCCCACAAC	CCGTTGAGGAGGGAGAAG	239
>SoSSR240	CACCAATTCCACCAGAAATC	GCGGCTTTGACTTGTCTT	211
>SoSSR241	CGATGAGGAGGATCTTAGGC	CACCACAGAGACCACAAACA	153
>SoSSR242	CAAATGCTACACCTGACGAA	CGTCTCCATTGCTTTCACTT	221
>SoSSR243	TTGAAATCCTCCATCCGTTA	ATCGGAGAATCCAACCATCT	194
>SoSSR244	GCCTCAGAGTCAATTCAAGG	ATGAGCAGATGAGGAAGTCG	231
>SoSSR245	GATATGGCTCGGGAGTTACC	ATCCACGCTCATCACCTCT	223
>SoSSR246	ATATGGCTCGGGACCTACC	TCGTACATCCCAGGAAACAC	236
>SoSSR247	GATGGCATAGGAACATCAGC	ATAACTGTCTGGAGCACTTGG	221
>SoSSR248	GCCTAACTGGCCTAGGTTTC	CCTCCTGCGACATATCAATC	206
>SoSSR249	ATTGGACGAGAGCAGAACAG	AGAAAACCAACCAAGGAACC	112
>SoSSR250	TATGCCAATGTTGTGGTGAG	CAAAGCAAACCTCGCCTTAGA	232
>SoSSR251	CATCCATTCATCATCCATCA	TGTGCAAGAATTCATTGTC	140
>SoSSR252	GGGCTACCGTATCTCAACAA	AAGCTTGGACACTCCATATGTC	230
>SoSSR253	CCAGTCTCAGATTCCCTGTGG	GGTACTTGAGCATCCCCTCT	249
>SoSSR254	GCCCCAAAAGAGAAGAGAAG	ATATCTCCGGGAAGTTGAGG	235
>SoSSR255	AGAATCGTCCATCAAAGCTG	TTACCGGTGCTGATTTCTTC	113
>SoSSR256	CTGTCAACCTCATCGGATTT	CTCTCTTCCCCAGATCTTCC	232
>SoSSR257	AGAGACGCACAGCTGATTTT	CGATTGGAGAACCAAAGATG	243
>SoSSR258	CCAGAATTCGTAATGCCTTG	GAGGGCTCACCTTAAGATCC	180
>SoSSR259	TCTTCTTCAATCCCCATCAA	CTTTCTTTCTCTCGGGAGGT	169
>SoSSR260	CCCGAGAGAAAGAAAGGAAG	GTTTCTCAAAGGCCAGAC	110
>SoSSR261	GAAGGAGAAGATGGGGAAAC	TGAACCAGCCCTGAAGATAG	222
>SoSSR262	AATGAGCCCTTTTCAGCTTT	CCTGTCCCTCCCTCTATCTC	234
>SoSSR263	AATGAGCCCTTTTCAGCTTT	GTCTAAACCCTTCCCTGTCC	247
>SoSSR264	CCCTAATCAATTGCAACCAC	AGCATTGTATGGCTTGCTC	234
>SoSSR265	CTCCGTAGCCGTACCTATGA	CTGGAACCACCCTATCCTCT	178
>SoSSR266	GTTAGGCTCTCTCCGACCTC	GTGTGTGCGAGTCACAAGAA	108
>SoSSR267	CGCTGAGAAATCCTCTCCTC	CTTTTGCTGCCTGTTTTGA	249
>SoSSR268	GTTCAAGAAAGCCTGTGGAA	AAGAAATTGGTGCTCTCGTG	235
>SoSSR269	CCCAATCCACAATCAAATC	GCAGGCAGTCTAACAAGTGC	229
>SoSSR270	GCTCCCTAAACCTCCACACT	GTGGAGAAGGGATCTGGATT	206
>SoSSR271	GCTCCCAACTTCGTTGTTT	CTGAGGTGCCAATCCAAG	204
>SoSSR272	GGAGAAGCATGGGAAAGAGT	AACATGGAAATCGTTCTGGA	225
>SoSSR273	AACTCCACGGCAAAGACATA	ACATCTGGCCGTTAGCGTA	174
>SoSSR274	AATCACCTGAGAAGCAGCAC	GATCCTCTCCCTGCAGTTG	164
>SoSSR275	AGACACAGCAGCAGCAGTTA	GTCAGAGACGTCATGGGAAA	238
>SoSSR276	GACAGCGGCATACTCTCAAC	TGAAATTGCATCTGTTGCTG	201
>SoSSR277	CCTGAAGATGAAGAGGACGA	ATCTGAAGCGGGAAAGTTCT	243
>SoSSR278	AAATTTACCACCGTGTCC	ATCGACAAAGCTGCGATTAC	187
>SoSSR279	CGCCACCGGTTTATTACTC	ATATGGTGGAGGTGGGGAAG	245
>SoSSR280	CACCACCACCTTCACCTTC	CGGTGGAGATGAGTAGACGTA	164
>SoSSR281	ACCTCCACCTTCACCATCTC	GCGGCGATGAGTATTTGTAT	227

>SoSSR282	GGGGTTCAATTGTTGAGTTG	CATGCCTCGAACTTCTCATT	221
>SoSSR283	TTCGACGACAAGAAGAGGAG	TTTGAGAGTGAGACGGGAAG	200
>SoSSR284	ATCAGCAGCCAAGTTCAGTC	CTGCATCGGAATTTGGTTAT	228
>SoSSR285	CGAGGAATAATAGCCAAGCA	TGTGTCCCGGTAACTCTGAT	108
>SoSSR286	CCGAACATCCAGTAGGGATA	CCTCTGTCAGCATCCTTCTC	201
>SoSSR287	GGATAAGGAGCGAGAAAAGC	CTTGCCCGTGTCAACTCT	202
>SoSSR288	GGCTTTGATGTCTCCAACA	GATTGAAATTCCGACAATGG	231
>SoSSR289	GTCGTGGATGAAGAAGATGG	GCTGTGAGGGCTTGATAAGA	234
>SoSSR290	CAAATGGTCGAGGAAGAAAA	ATCTCCCCTTTCCAAATCAG	218
>SoSSR291	TGGCCATCTCACGCTTTC	CACTTGCTCACTGCTTCTCA	149
>SoSSR292	AGGCTCTTCTCTGACGTCT	GGACTTCAACCAACTGAGCA	242
>SoSSR293	AGACATCCCCTTCTTCAACC	CGAAGAGCTTTCTGGTCTTG	170
>SoSSR294	TGTAGAAGCCCGAATTTGTC	ATAGACCAGAGCGACTGAGC	116
>SoSSR295	TTCATTCGATGCAGCTTAC	GTGCTCTCATCACCAAATCC	179
>SoSSR296	CATACGACCGCTTTATCCAC	TCTGGGAGAGAAAATGGTGA	105
>SoSSR297	CCTCCTCATCTTCTCCTTCC	GATAAATCGAGGCGGCTAAT	232
>SoSSR298	TGTTCAAGATGCAGAGCAAG	AAGGCTGCATAGACTTGACG	104
>SoSSR299	TCAGGTGTTTGATGGACTGA	ATCACCATCGCTCTCAACAT	166
>SoSSR300	CGAGGTTGTTGAGCATCTTT	ATTCTCCTCCAACGCTTTCT	121
>SoSSR301	CGGCCTTAGCCTTACCTTTA	CCTCCGATAGCGAGATTATG	123
>SoSSR302	TTCGGATTTTGAGGAGAGTG	TTATCATGGCGAGGAGAAAC	168
>SoSSR303	GTCAACATCATCATCTTCTTCTTCC	GAAGAGAGCTTGAGCAGCAG	181
>SoSSR304	TTTTACTGGCCAAACCAATC	AGCCTCCTTCTTCTTGTGCT	230
>SoSSR305	GCTACTTACCTCCGCTTCG	GGCAGAGAGTGAGAAAATCG	169
>SoSSR306	GTGCTTACGGAAAGAGAGA	GACCGATAAATTCACACCA	211
>SoSSR307	GACGGCCCAGAACATAGAC	TGCTCCAGCTCTACCTTGAC	150
>SoSSR308	AAGCAGCAGCCTCACTTTTA	TCCTTCTCCTGCATCTTGAG	205
>SoSSR309	ATGCTCACAACTGGAGAGG	AAGGTGTCCGAGCTTCTTCT	248
>SoSSR310	TCCCACAAACTCAAAATGT	AACACGAATTTACGTCGAT	167
>SoSSR311	AGCCAAGAATCCACAACAAG	CATTTCTCTCAGCCTCCAAA	163
>SoSSR312	CCCATCAAGCAAGATCACTC	GCTTTGCTGCTGGTATTTGT	219
>SoSSR313	CCTCAAATTCAACACCTCCA	GACTTAAACGGGGTGGGATAC	141
>SoSSR314	ATTTCAGAACAATCGGACCA	GCATTATCCTCGGAGTCCTT	243
>SoSSR315	CCTTCTCTTTCATCCTCACCT	GTTGTAGGGGTTGCTCTCTG	134
>SoSSR316	CCTCTAGCGAGGAAGAATCC	CGCCTCTTTGAGAAGGTCTA	224
>SoSSR317	GAAGAACCCTGCTTAGCTC	CTCACCTTGAAATCACACC	128
>SoSSR318	TAGTGGAATTGGTGGATTGG	TAGGGAGATGAATGGGGAAT	147
>SoSSR319	TGGATTGGGATGAGAAAGAG	ACTCGAACTCGAACACGAAC	244
>SoSSR320	AGTCGGAGAGAGGAAACACC	TGCTGGCTATGTTTCATGGTA	176
>SoSSR321	GGGAGAGGATCGCCTATATT	TATGTGTCGGTTGTGATGGA	223
>SoSSR322	TGCAGAAATACCGCTCTAC	GCTCCCAAATGCCTAAACT	245
>SoSSR323	CTGCAGAAACTGATCGTCCT	CGAAGAGGAAGAGCATGAGA	229
>SoSSR324	AGGGAAATTGCAGGGTTTAC	TCAATTTTCTTGGCTTGCTC	174
>SoSSR325	AACAGGAAGATCCCAACTCC	GTGGAGGTGAGAGAGACGAA	187
>SoSSR326	TGGACAACAGCATATGGAAA	AAATCCCAACAGGTACAGCA	214
>SoSSR327	AAGACGCCAACATCTTCATC	TTGGACTTCCCCTTCTTCTT	223
>SoSSR328	AATATGGTGGTGGAAAGCTGA	TCCGATTTCTTCTCCTTTGC	230
>SoSSR329	CCTCTTAGCTCCCCTCTTCC	CGGATTCAGGTACAACGTC	248

>SoSSR330	CGCAAGGATGTTATGATGGA	GGTTGTACGCAAACGAAGAC	192
>SoSSR331	ATGACTGCTTATGGCGTTTC	CGCGGTTCTTCATCATCTAT	133
>SoSSR332	CCAATCGTGAAAAGATGACC	GTCGAGAACAATACCCGTTG	122
>SoSSR333	ATGGAGAAGATTTGGCATCA	GCTGACACCATCTCCAGAGT	237
>SoSSR334	GGGATCACTACTCGTTTTGG	ATGGATGAATCCGAGAGATG	147
>SoSSR335	AACAGACCACCTGCTTCATC	TTTTCTGCTCGTTAAATGC	239
>SoSSR336	CCCTTCTTCTTCCTCACACA	TATTCAGCGGTGGATCACTT	101
>SoSSR337	CACGCAAGTCTCCTCAAAC	ACTACTCGACGACGAAGACG	160
>SoSSR338	AAGCTCCTCCCTCTCCTCTA	GTGAAGGCTTGATCTTCTGC	112
>SoSSR339	TTGAGGAGGAATACGACGAG	CTCAACAGCCTCAAAGCAAT	202
>SoSSR340	GCTCGCCGACACTAAGATT	CCAAATTATTCTGCCACACC	236
>SoSSR341	CAAATCTAAGCGAGGGTGA	TTCTTGAATCCGACGTTTCA	214
>SoSSR342	GGGATGATAGTGACCGTTGA	CGCCATTATTGCATTCTGAT	243
>SoSSR343	AACAATCAGCGTGAAGAAGG	ATTTTCGTAATCGCGTACTGG	179
>SoSSR344	AGCCCTTCTTTGGAATTGAG	GCTAATGCCGCTATTGCTAC	186
>SoSSR345	TTCTCGAGCAGATTTTAGG	CATCGTTGTACCCTCCCTTC	238
>SoSSR346	GGCTTCCAACAACAACAATC	GAAGACCTCCTCCACCTCTC	163
>SoSSR347	GGCGTAGTGTGCAGAGC	GCATCGATAATGGCGTTC	155
>SoSSR348	GGAGGTGGTTATACGTCAGC	TCACACTCGTCGTGGTTAGA	221
>SoSSR349	GTGGCAAAAGAGCCAGTAAA	GATCAGAAGATGCTCGCTGT	100
>SoSSR350	CCTCACCTCTCTCCAAC	CAAGAATTGAGAGGCCGTAA	162
>SoSSR351	CCTCTTCCAGGTCAAATGAA	TGTTACCTTGGTAGGGTGGGA	121
>SoSSR352	GCGGAAGGAGCTCAAAT	GGGGTCTTCTTCATCTTGGT	143
>SoSSR353	TGGGGAGTAGATCAAGCCTA	AGATTGCTAGCTCCAGTTGC	168
>SoSSR354	TCCCCAACACTTTCACGAG	GAGGTGTGGACGATTTTGAC	172
>SoSSR355	GAACCTCACATTTTCTCCA	CAAACTTTTGCATTGTGCTG	231
>SoSSR356	CATGCATTAAACCCCAAGTC	CTATGAATTCCTCCGCCTTC	180
>SoSSR357	TTGGAAGCTTGAAGAGATGG	TAGGTGGTGCTGCTTGTGTA	203
>SoSSR358	CAACCTCCAAGAAATGATGG	TGTTGATTGTGGTGAACCTG	234
>SoSSR359	CTCCACCACTCAGAAAGCAC	ATCAAGTTTTTCGACGCTCAC	144
>SoSSR360	CTTCCAGACGTCGCTATCTC	TTGCTGATCTTCATCCGACT	194
>SoSSR361	GAGGTGATGGTGTACGAAG	TTGTGGTGATTCTCAACGTG	210
>SoSSR362	GGATATTCTGTTGGGTGACG	ATCGAGAAGGAGGAGAGCAT	229
>SoSSR363	CCTCTCCCCACCATAGTAA	CCCTCAAGGAGGATTTGATT	128
>SoSSR364	AAACCAGACCCCATCTGAT	TCCCACAATCACCAAGTATCC	179
>SoSSR365	GCGTTGATAACGAACCTGCT	GCTAATAATGTGCCGAGTGG	233
>SoSSR366	GCAGTACGTGAGCAATGTGA	AGATGGTAATCGCAGTGGAG	230
>SoSSR367	ACAACCCACTTGGATCTTCA	ACGAATCATCACGGTACCAC	212
>SoSSR368	CAGAGCCAAGGACAAGAAAA	ATGGATGGGTGATGAAAAGA	132
>SoSSR369	AGCAGCAGCAGCAAAC	AGTATTAAACGGCGTCGTGA	218
>SoSSR370	GTTCTACTATGCCCGTTGA	TCCTTGTCGATTCACTCTC	129
>SoSSR371	GTGGCTGATGTTTCAAATCC	GCCCACGAATTACTTGACAG	110
>SoSSR372	CAGCTACATGGACACAGCAG	TGGGAGAAGTGGTTTGAAGA	197
>SoSSR373	AGCTCATCAATCAAGCCATC	GGACCCATCTCATGAACCTC	220
>SoSSR374	GGGAAGCTAATCTGCGACT	GTCAGGTGCCTCTTCTGCT	173
>SoSSR375	TCTTCATCCTCGTTGGAGAG	ATTTGAGCATCCCATTTC	201
>SoSSR376	GAGCAGAGGCAGAGAACATC	GCAACTCCTGCACCATAATC	215
>SoSSR377	CGATGATTTATCCGATTGC	CTACATCCTCCCCATCTCCT	162

>SoSSR378	AGGAGATGGGGAGGATGTAG	CACATGCACTTGTTTTCCAC	209
>SoSSR379	GGAGAGATCCCTGCAGAACT	TCACCGGTCTTTAAATCGAG	177
>SoSSR380	AGGAGTGGGAGACGAAGATT	TGCGCCCACTATCAAATAAC	249
>SoSSR381	CAGTATGCCGGGTTTCAG	GCTCCAGATACCTCATGTCTG	205
>SoSSR382	ATCGTCCTCTACAGCAGCAA	CTCCTTGGAATTCTTCACCA	222
>SoSSR383	AATACGCCTTCTCCTCGTCT	CCATTCCATGGTGAGTGATT	191
>SoSSR384	TATGCTGTGCAGAGAAGTGG	TGCTCAACCAGAGGAAGAAC	230
>SoSSR385	TCTTCACGCTCTCAATCCTC	AGATGTTGCGGAACCAAAC	201
>SoSSR386	CATCTACTTGAGCGAACCT	CCTTTTACTGCCGCTCTTCT	169
>SoSSR387	GGCGAGTAGTGGTAAAAGCA	ACAGAGCCAGTGAACCTCGAC	141
>SoSSR388	CGTCTTTGGACTTTCCTTCA	GGGTAGCAGTAGCCTCTTCC	241
>SoSSR389	TCCTCCAAATATGGGACAGA	TTGTTCCACAGAAGAGCACA	238
>SoSSR390	GAGAGCTTGTCGAGGTGGT	CCCTATGCTGAACTCGAAGA	222
>SoSSR391	CTTGCTCATGGTGGTAGAGG	ACAGACATCACCAGCTTTCC	243
>SoSSR392	ATCTTCGCAACCTTCACAAA	CAAGATTAAGTGGCCTTGC	203
>SoSSR393	AAAAGTGGGGAAGCGTATCT	CACCAATCGGCTATAAAAGG	203
>SoSSR394	TTATCGGCCAGTCAGAGAAC	TCGTTCCATTGCTTTAGCTC	240
>SoSSR395	GTTGGGACAAGAGATCATGC	GGATGGAACCACCAGATACA	125
>SoSSR396	ATTGCCGCGCTTAGCTTAC	CCAAATCCTCTGTCCAAATG	173
>SoSSR397	CATCTCTCTCACGCTCATCC	TGAGCTTTGTTCTCATCGAA	216
>SoSSR398	CACCACCATAACCCTGTTTGT	CAGGGATCTTTTCTGGTCCT	211
>SoSSR399	CTCAGAGAGCCCAACAAGAA	TCATCCACAAACTCCTCCTC	248
>SoSSR400	GAGGATTGCTCTTTGGAAT	GCTTCCATTTCCACGTCATA	104
>SoSSR401	TCTTTGGTGGCATCTCTCTC	AGGCATCAATGAAGAGACCA	222
>SoSSR402	TTTTCAGCTCTTCGATTTGC	GAATCCATGCTCCATTCATC	246
>SoSSR403	GAGGAGGAGAGGAGCGATAC	GAAATTGGAGACAGCATTGG	138
>SoSSR404	TGGCTTCCTCCTACTGTCTG	GAGGATAGAGCGGCTAAAGG	181
>SoSSR405	GGCTCAGACACGGAGATTTA	TACCCATTAACGGAAGCTGA	236
>SoSSR406	AGTGCGAAGGGAGGAAGTAA	CTCTTGCTATTCGGCACATT	128
>SoSSR407	CATCCATCACCCATTTTCTC	GAACCCATTTCTCAAGCTCA	195
>SoSSR408	GGATCAGTCAAATTGGATGG	TTTTGAAGGTCCTGTTTTGG	206
>SoSSR409	CTTTGCATCCATTTTCCATC	TAGCTGTGAGTCTGCTCCT	232
>SoSSR410	GCTCAAAGCCATAAACTCC	CAAGAGAAGTCCAGGAAACG	212
>SoSSR411	TCCCTCCACTAATCTACCC	AGCAGCTCGAATCTAGACGA	244
>SoSSR412	ATGTCGCCATCATCATCC	GCTGCTTGTCGGTGATTACT	175
>SoSSR413	CTCCGATGGAAAGACAACAC	TGCTTTCCTCACGTTTGAGA	229
>SoSSR414	AGCTGGAGATGAGAGTGGTG	CGATTTCTCTTCTGCACAT	200
>SoSSR415	GAATCCGTTCCAATTGTTTG	AAGTTTGGGGAATCTGAACA	166
>SoSSR416	CGATGGATAAACGAAGGATG	GTGTTCTTAGGTGCAAATGG	241
>SoSSR417	GGTAGACGATGTTGGGAATG	TCCACACAAAATCCCTCTGT	201
>SoSSR418	TTCATTTACGGGGAATCTCA	GGTTTCTTACCCAGACGAGT	216
>SoSSR419	GAGGTGTTCTGTTTGAGTGCT	GCTTTCTCTCGATCTCTTCC	243
>SoSSR420	ATCATGGCGAATTTTCTCAC	AGCTGCAGATCTTGTGCTCT	185
>SoSSR421	ACGAAACAAGCAAAGGACTG	GCCCATCTCAACAATTTAGG	199
>SoSSR422	CCCCATCTTCCACTTCTTT	CTACGGAGTGGTGGAGGAG	111
>SoSSR423	TTTCTCATCCTCCTCCTCCT	GTGCGGTCGAAGATCATTAC	155
>SoSSR424	GAAAGGGAGGTTGTGTGATG	CACTGATCGCTCTTCGAAAT	169
>SoSSR425	AGCTGGATCCGAGGATAAAC	CACTGGATTGCTCCCTGTAT	186

>SoSSR426	AGTTGGGGTACGAAGAGGAG	CCTTGCTCACCAACTAGACC	138
>SoSSR427	CGACTACAGCAAGCTCGAC	TCCTCCGTCCTTCACAT	225
>SoSSR428	CGATTGATAACTGGGAGGTG	CAAACAAGTTGCTGGGATTT	176
>SoSSR429	GAGGAGGAAGAAGCCTCAAG	TCACCTCCATTCTTCACCAT	207
>SoSSR430	ATTCCAATGAGATGCCTGAA	TCAACGAAACCTGGAACATT	175
>SoSSR431	GAGTGGAGGAAGAGGAAGGA	TCCACACTCAAGCTCAAATTC	155
>SoSSR432	GGAAACTCTCTTGCATGCTT	ATCTGAGCTTGCTGCTTGAT	191
>SoSSR433	CCTCCAACAATGCAGATTT	ACGTAGAGCAGAACCACCAA	185
>SoSSR434	ACCATGTTCCAAGCTGTGTT	CTGTTGGATTGGATCAAAGG	147
>SoSSR435	CGAATTCTCGAAGAGCAGAG	GTTGTTGTCAGCGTTGGAAT	160
>SoSSR436	GAATGTCTCCGAACAAATGG	ATTGGAGCATGGTTCCTACTA	143
>SoSSR437	AATGAGAAGAGGCAGCTCAA	ATCATTCTGAGGCACATCCT	192
>SoSSR438	CTAACACAAGGGGAAGAAAGC	TTTGACCTTTGAATCCTCCA	161
>SoSSR439	AGGGGCATCTACAGAACAGA	AATCCCACCTCAGACTCTCC	249
>SoSSR440	TCCAACGTTTCAACCTTCAT	GGCAAACCCGAATCTTTATT	186
>SoSSR441	AAATATGGTGGCTCGAAGTG	GTAGTGCACCCAGCTTCTTA	245
>SoSSR442	CGCAATCTTCTGTCGATTTT	GGCATCCACTTTCACATCTC	230
>SoSSR443	TTGCCTTACACCAACAACAA	TTGCTCGATTTCGTACCTCTC	155
>SoSSR444	GTTGTCTCAATTTCCGGTGCT	AGAAATAACCGCCAGAATCC	241
>SoSSR445	CCAAGGAGCAAGAGAAACAA	AGCAACTACTGCCTCCCTTT	209
>SoSSR446	GGCTTCCTCGTCTTCTTCTC	TCGGATAGGAATTGATCACC	235
>SoSSR447	TCCCAAATCACAAACCTTAA	CCTATACTCGCAGAGGCAGA	231
>SoSSR448	GGTTTGGTTGCTGGAGGT	GCCGGTATCTAACCACACAC	241
>SoSSR449	GGGAAAAGCCAAGAGAAGAC	CTTCCCGTAGATGCTTGAAA	220
>SoSSR450	CAAAGCCACAGAAACGAACT	GCGAATGTGGAGAGAAGAAA	159
>SoSSR451	CACGATTCGAGAGAGGTGAG	TTCTTCCGCTATCAACAAGC	106
>SoSSR452	GCCAGTGAAGGTGACACTA	GTTGTTCTGCCCATCACTTG	238
>SoSSR453	TGGCAGAGGATTGGATAAAG	GAGAGCTAAGAAGCGTAGGTTG	230
>SoSSR454	GCATGATCTGTTGAGGGATCT	ATGGCACCAAATAGAAGCTG	213
>SoSSR455	GCAGAGAAAGAATCCCAACA	CAGTGTGTGAGACGACTGC	191
>SoSSR456	GCCGGAGCTACAATACTCAC	CTCGCCGAACTTGTATTTGT	228
>SoSSR457	AAGAGGGAATCTCCATCACG	GATCTCCCATTCCGAGTTTT	112
>SoSSR458	CTACCCTTCATCCGGTCTG	GGGCTTGAGAAGTAGGGAAG	244
>SoSSR459	TAACGATAGGCCTGGTTCTG	AGCAAAGGGATTGACCTCTC	178
>SoSSR460	GAGGCTTGCTGTCAAATTA	ATCAACAATTAGGCCCTTCC	193
>SoSSR461	TTACCTGAGGCGTAAATGC	GGGCCATAAGGCTACTGAAT	197
>SoSSR462	GTGGCTTTTCGAGGAGGTG	TTCTCGTGAGAGAGCTTCGT	148
>SoSSR463	AGATTTCTTCCCCAGCCTAA	TTGATAAGGAGACATTCGATCC	160
>SoSSR464	AGGCCAGAGTCAGAAAGAG	GTGACCCCAAATAGCCTTCT	160
>SoSSR465	AGGAAGTGGCAGAGGATTG	CCGAGGCATTATCAACTTTC	246
>SoSSR466	AGCAGGCTAATGGTCAGTCA	TCGCTCCCTTCTTCTTCAAT	132
>SoSSR467	GAATAACCGCTCCGATCTC	AAAAGTTCGCGATTGAGTTG	207
>SoSSR468	CCAAAACCTCACAAACCCTTG	AATCACCATCTCCATCTCCA	129
>SoSSR469	CCAAAACCTCACAAACCCTTG	GTGTTGGCTGCTGAACATCT	200
>SoSSR470	TTCCACCGATGAAGAAGAAG	TGCATGGTATCTCAAGAGCA	188
>SoSSR471	CCAAGGAATGATCGTAATGG	TCTCAACCGGAATATCCTCA	208
>SoSSR472	TGTGCAAATATGTTGGATGC	ATGTTTCCAAACTGGCAGAA	217
>SoSSR473	CTCAACACCAGGCTAGGAAA	GCCCGATTTTGCATACATAC	108

>SoSSR474	AAGGCACTGCAGGCTATCTA	CCTCTAGGCTCTCATCCACA	216
>SoSSR475	AAGGGGAGACCAATGGAGTA	TTCCTCTCCATCTCCATCCT	233
>SoSSR476	AGATTTGGGTTTGGCGATG	GCCTCAAGCTCAAGGATGTA	133
>SoSSR477	GCGTAGAAATGGATGAGCTG	TGATTACCACTGGAGGAGGA	150
>SoSSR478	ATCAAGTTCATGTCGGTCGT	CGTCGTTCTTCTTCTTCGTC	113
>SoSSR479	GCCAATGGAAGTCGTCGT	GAGGTGCTGTGGTTGAGATT	126
>SoSSR480	AACCGAAGATGATGGAAACA	TATGTCTCACCATGCAAACG	104
>SoSSR481	TCGGAAAGGGAGTTTAGTGA	CAAAGCCATCAACATTAGCC	187
>SoSSR482	CCGTCAACCTACAGCAAAAC	TCGAACACAACCCTCTTCTC	184
>SoSSR483	TGACGAGGAAGGTAGTGAGG	TAAATTAGGCCACCTGATCG	162
>SoSSR484	GTGGTGTATGGCGATGAGTT	TCCTTTCCCTTCGCATAAAT	104
>SoSSR485	GGA CTGTTGCCTCTCCTACA	ACAGAACCCTCCCAAAG	209
>SoSSR486	GTCTCGTCCACCAAACCTACG	GTGCAGCTTGAGGAGGTAAA	223
>SoSSR487	CCACTTTGAAACCTCAATGC	GAGGTCCAAGATGAGTCGAA	123
>SoSSR488	CGAAATTCTCCTGCTTCCG	GATTCTCGTTGGAATTGG	244
>SoSSR489	GTGGTGTTCCTCGTTCAGAC	TAACCAAACTCCGTGAAGC	193
>SoSSR490	CCATACGTACCCCTTCCTCT	ATACCGACCGATCAAACG	135
>SoSSR491	AGTGTGGTGGCGCAGAGA	TGTAAAGGGGATTTGGGAAT	195
>SoSSR492	GCTCAATCGCATCAGAACA	TTATGCTGAGCTCCATCCTT	101
>SoSSR493	CTCCAAGGAGATGGATCAAG	CGTTCAGCTCGTTTTCGTAT	168
>SoSSR494	GAGTGGAACGTCGTCGTATT	ACGACAGAGCAGAAACAACC	211
>SoSSR495	AAGCAAGACGCTCGAGTAAA	CTTGCTCTCCGACCTCCA	239
>SoSSR496	AATCCCAGTTGAGGGATTTT	ATCAAGGCCCAAGTTAGGAT	140
>SoSSR497	ATCCTAACTTGGGCCTTGAT	CAATTCCCATTCTCATCCAC	142
>SoSSR498	CGGTGGCTGGTTAACAAATA	ATAATTTGCGAGGTTTCACG	232
>SoSSR499	AAAACCACAGCAGAAACAGC	AAACGCCATAGGGTCTTCTT	175
>SoSSR500	ATAGACAAGGCAAAGCAACG	GCCTTCGTGGTAAACATCATC	212
>SoSSR501	CCCGTTCGAGTACCAAGATA	TCCTCTTCGATTACCTTTG	215
>SoSSR502	AGAGCCTGAGGGTTTTGG	TCCCTCCTCATCTTCTCTC	130
>SoSSR503	TCCTCACCCACCTCAATATC	GCAAGCCATCCTAAACCTCT	177
>SoSSR504	TCAGACTAACTGCCGAAAC	ACTGGCATTGCATCACTTG	177
>SoSSR505	ACAGAGTGAGTGAGCCAGCA	GATGAAGCATGTGTGGGATT	102
>SoSSR506	TACAAGAAGGGCAGCATAGG	TCGACTCCATCGTCCATATC	200
>SoSSR507	GAAGGTGAAGCGGTTACTGA	GGAGGAAGCTGATCACGATA	191
>SoSSR508	GAGTTGAAAGAAGCGATGGA	CTTGACTGATCCGAAGCTA	244
>SoSSR509	TCATGCTGAAGATGTGATCG	ACCATCTCATTGGAAGTGGA	240
>SoSSR510	CGTCTCCGACCTCAATTTT	GTCAGCTCCAGATTCTCTCGT	216
>SoSSR511	CGTCACA ACTCTCGGAATCT	CTCTGAGGCTGTCCAGA ACT	194
>SoSSR512	AGGAGGTGGTGAGGAGTAGG	CCATGACGTTCAAATCTCC	246
>SoSSR513	TGGCGCTTACTTACCTCAAC	CTAGTGACAACGGGGTGAAC	228
>SoSSR514	TGCAAAGACCATCTCTCTCC	GGGAGCTGGAAAAGTGAAGT	110
>SoSSR515	CGTCTAGTAAGAGGCGTTCCG	TCTGTGCCTTCTCAAGTTC	211
>SoSSR516	ACTCTCTCCCCACCTCTCAC	TCGAGAAGCTCTTGCTTGAT	239
>SoSSR517	AGA ACTCCACAAATGCAAGC	TGTGCTAGAAACCTTAGCCTCT	179
>SoSSR518	GGTACCCTAATGGGAATTGG	ATGCTGTATGGGCAAGGATA	248
>SoSSR519	GGGCTATGGGAATTGTAAGG	TGGAAAGCTCTAGACGATGG	219
>SoSSR520	TGAAGGTATCAGCCAAGAGG	GGACTTTCCCGTATCCGTAA	144
>SoSSR521	ACAGCCACAAGAAATTCCAC	CTTGCTGAGGTAGAGGCAGT	135

>SoSSR522	CTCCAACCCTACCAGGAAAT	TCCCGAAGAAGATGAAGTTG	149
>SoSSR523	GGAGGAAGAGGAAGATGAGG	GCAGTGTGCAATCAGCATAG	158
>SoSSR524	GAAGTCATCAAACCGCATT	ATTCCGAGCTACGGTCTTCT	246
>SoSSR525	CAAAGGCTGCAGATGAAAAT	TTTGCTCAGCTCCACTTCTC	177
>SoSSR526	TCAGGATCGCCGACTATTTA	GTCCTTCTCTCGGGCTTTAC	117
>SoSSR527	GAGGTTCTCTCCCTCTCC	TTATGCCTTCGGTGAATTTG	146
>SoSSR528	ACTCAGTTGCCAATCCAAAG	AGTGAGGGGTGGTGGTATTT	114
>SoSSR529	ACCATCGAAATGACGGACTA	GCAGAAGATAGATGGGCAGA	132
>SoSSR530	GGATTTCTACGCGAAACTGA	CAATTCTGCATCACACAAGC	232
>SoSSR531	CATGGCAGATGAGTTCACAA	AAAGCTGTTGACCCCAAGAT	185
>SoSSR532	TGGGACAAGAGCAAAGAAAG	TGATCACTTGGAACATGTCG	238
>SoSSR533	TAAGAGGAAGGCTGATGCTG	CTTTCTTCTCCACCTCCACA	113
>SoSSR534	GGAACCCTATTCGGAAACC	TTGCAGTACATCCTCCACAC	233
>SoSSR535	CTGATAGAGGCGACGTGTTT	AATGCTCTCGATGCAAGAAT	182
>SoSSR536	CAGGGGAGATGTGTTTGAAG	GCACTCCCATTGATATGCTC	194
>SoSSR537	GAGCCCTGACGACTCCTT	CCACAGAATTCACGGGTATC	233
>SoSSR538	AGCCACCAAATCCTTTCATA	GAGCGTTTGTGGTTGATT	216
>SoSSR539	CGGAAAGCTCTTCAAAC	GTTTCAGCTCCAGGTAGACG	231
>SoSSR540	GCGACTAGTTGAGGACTGGA	GATGTAGGAGTCGCGAAGAA	198
>SoSSR541	GGAGGATCTGATGCCCTTAT	ACTCACGGGAGGAGTTGC	140
>SoSSR542	AGGTGAAGATGATGCATTCG	GGTTTCAACCTAGGGATTGG	241
>SoSSR543	CTTGCACCTCCAGGTAACTC	CCTGTAGTATCCGCCACTTG	222
>SoSSR544	ATGGAGAAGTCCTGTGGTGA	TGGAAAACGACCTACCTCTTC	180
>SoSSR545	TGGAGAAGGAATTGGAATCA	CATCTCCATCTTCCCCCTTT	231
>SoSSR546	CATCTCCAACGCGCTGAT	GAGGTACTIONAACGATTGCTCA	219
>SoSSR547	CCTTCGTTATGGCTTACACC	TACACGTGCCTGCAAATA	245
>SoSSR548	TCCTCAACAGCAGCAAGAAT	TCTGCTCGATAAATGGGTTT	215
>SoSSR549	CCCGTTGAAGATGAAGAAGA	ATGGGCTATTTGAGAGCTT	210
>SoSSR550	GCACTCTCAAGGGATCACAC	GCAGCAGATCTCCACTTGAT	149
>SoSSR551	GTGCACAAAGCAGTTACGAA	AAACATTCTTGGCCTCTCCT	126
>SoSSR552	CCGTACGTCCATCCTACTCC	ATTTTGAGCTTGCTCGACTG	201
>SoSSR553	CATCTTGCTTTGAGGAATGG	CCCTCTGAAGAGTCCAACAA	155
>SoSSR554	GCTCTCAAACGCAAAGTGT	CTCCATCAGCTCATCGTCTT	238
>SoSSR555	TTCAACGCTATGGGATCAAT	CAGTGCCTGACACAACACAT	192
>SoSSR556	TCTACCTCAAATCCGTCGTC	CCTTGAGCATTGATGAAC	108
>SoSSR557	TAGGGAGAGACGTGCTTGAC	GGCAATTCGTGACTGCTACT	153
>SoSSR558	CTGTGCAGGACAAGGAAACT	TGCATGGTATAAGGGCAAAT	137
>SoSSR559	TCCAAGCACTCAAACCCTAA	CACCTTGAAAGGGAAGTGTG	204
>SoSSR560	AGAAACCCCGAATTCATCTC	TAGTTAGGGCTGTCCACCAG	177
>SoSSR561	TCGATCTCCACTCAAACCAT	GCCACATCAAATCTCCAAC	219
>SoSSR562	CTCGAAAAGGCAAAACATA	CAGCATGTGAACCATTTTGA	183
>SoSSR563	AGCCCTTGCTTCAAACCTTT	CAGCAAAGGACTCGACAAAT	210
>SoSSR564	ATTTGTCGAGTCCTTTGCTG	TCACAAATAATGGCCAAGGT	206
>SoSSR565	GTGCAGATGCAGTTCTGTTG	CGATTGCTAAAACACGATCC	141
>SoSSR566	CGAATCATCTAACCACGAG	ATTCTTCCGTCTGGCTTTCT	250
>SoSSR567	CAGCGTTGCTAAAACCTGTGA	TATCTTCTCGATCGTCTGC	248
>SoSSR568	CCACCTGTCTCTTCTGCATC	AAAGATCATCCACTGCTTGC	238
>SoSSR569	CGTTGTCATCAAACCAGACA	TAACACTTGAACCCCATGCT	132

>SoSSR570	CAGTGTAGGGAGAAGCCTGA	CTCATCGAAGTCACCATCCT	102
>SoSSR571	AAACGATGGCGTAGAAAGTG	CTTTTCCTCTTCGTCCTCT	103
>SoSSR572	GAGCACGGCCATTACAATC	AGCTAATCGGGTCAATCTT	158
>SoSSR573	CCGAATCACATTTCTTCCAC	TACTGCGAGGGAGATTGAGT	167
>SoSSR574	TGGCGAAGAAGTTATGGAAG	GGCATTGACTTTGCAATTCT	151
>SoSSR575	GTGGGTGACAGTAGGCAGAG	CAACTTGTGAATCCCACACA	212
>SoSSR576	CGTCTCAATAGCCAAACACC	ATCGACACCTTCCATCTCAA	234
>SoSSR577	TTCCAATCTTCCAGGGATAA	TGGTAACGGAAGGGATCATA	155
>SoSSR578	AGCAGCTCAAACCTGCATAGG	CCTGCAAACCTTCTGTTAGGC	250
>SoSSR579	TCAATCAAGGGATTGGGTTA	CGACTACACATTCACCAGCA	216
>SoSSR580	CCTCACTCTTCCAACACCAT	CCATGCTTGAAGGGCTCTA	218
>SoSSR581	CCAAATCCAACAAGCCATAC	GAGCTTGTATTCCACGTGCT	190
>SoSSR582	TGGAGAGAGCTTCTAAGCAAA	ATCACGAGCACGAAGAAATC	227
>SoSSR583	AGTGTAACGGCAATTCACAAG	GCAACCACTAGCAACCTCTG	244
>SoSSR584	TCTTGTGGCTCTCTCCCTAA	CAAACCTCGGAAGATCACAC	245
>SoSSR585	CGAGGTGAAGAAGAGCAAAA	CGCTCAATTCCACTCTCTGT	151
>SoSSR586	GGTGTGGAGCAGTTCATTC	ATCCAAATTCCGACTTCTCC	206
>SoSSR587	CACGTTTAAGCACATTGACG	TACGGCACATTGGACACATA	204
>SoSSR588	GAAGCTGAATCCGATGGTC	TCCGGCGATATCATAGAGAG	193
>SoSSR589	CTATGATATCGCCGGAACAC	CTGCATACCCTTTCACCTCA	224
>SoSSR590	CACCAGAAAAGGCTGTACAAA	GGATTGCTCCGCTATGTTTA	183
>SoSSR591	ACCAAACCAGCTCTGCTAC	CATCTTCTCCCGCAATCTC	119
>SoSSR592	TCGGATTAGCTCCGATACAG	CTGCTCGATTTTCATCAACCT	187
>SoSSR593	CGAGGTTGTGGTGGAAATTAG	TCAGCTTCTCCCGAGAGTA	153
>SoSSR594	TCATCAACACACCCATTTC	TGGAAGATGGAGAAGCTTTG	172
>SoSSR595	AATCACTGGTCACACGTCCT	GGGTGAGAGAGGATGAGGTT	182
>SoSSR596	GATCACGCAGAGAGATGGAC	GTAGAAGATCGCCGAAAAT	162
>SoSSR597	ATAAGATCAGGACGCAGACG	ATCGATTCCAGTCCTCCTTC	219
>SoSSR598	ATTTCCAAAGGCGGTTATTC	CCTTGTGCTCCGTTTACTTG	191
>SoSSR599	CTACTTCGACGTGGTGGTGT	GTGAAGCCTCTGCACGFACT	187
>SoSSR600	CTCGGGGATCTTGTTGATAC	TTCCCTTTTACTCGCCTCTT	233
>SoSSR601	TGAGTGTGATATGCAGCAGTAGA	GTACCATATGCCGACGTACC	206
>SoSSR602	ATATGGGCTGCTCTTGGTCT	TGCTTGGAGGCTGAGATAAC	127
>SoSSR603	GGGAAACCGATTATGAGAGC	ACATTCGATGCTACCACCAC	179
>SoSSR604	GAACAACCAGGTATCGCATC	TTTGGCTTCCATCTCAACAT	246
>SoSSR605	CAGAGCTTTCTGCTTTGAGG	TCACCTTCTCCATCTGCTTC	228
>SoSSR606	CGGACTGGTGGTCTACATC	ATTATGCCGACGGGTAGAGT	151
>SoSSR607	CATCGGATTTGCTATGCTCT	ATGGAAGCTGAAGCTGTTTG	236
>SoSSR608	GGATGAACAAGCTGGTGAAG	GGTTTTCGTCTCATCCTCT	173
>SoSSR609	TCCCTCTCTCGATCTCTCT	TAACCGCTGTCAATCTAGGC	162
>SoSSR610	CGCCATCATCTTCTCTCTTC	GCTTTCTGGGTAGGGATTTC	211
>SoSSR611	GGAGAGAAACTCCGTGAACA	ATTTCTGGTCTTCGTCCTC	224
>SoSSR612	GAATTCATCTGGCCGTACAC	CCACGTCGATGCTTACTCTT	180
>SoSSR613	ACAGCTGATGCTTTGGTAGC	CCCAAGCCTTCAAGAAAAGT	224
>SoSSR614	AACCCTAATCAACCGGTAT	TCACAATCGCTGTGACAATC	239
>SoSSR615	GATTTCTCGAAAGGGTGTGA	AATCAGCCAAGCACAAAGAC	154
>SoSSR616	AATGGAGTAAGGGGAATTGG	CGTGAATCCGTCAACTCTCT	104
>SoSSR617	GCGGCTTTTGATAAAACTCA	CCCAAAGACGTAAGCGTAGA	111

>SoSSR618	GAGTGAGCTGAGCAAGGAGT	CGTCTTCCACCATGAATCTC	221
>SoSSR619	GAAAACGAGGTGGTGAAGA	CGAACACCCTTTTCTTCAA	133
>SoSSR620	TGAGCTTCTTTGCCATTTTC	AGGGGAGGGAGAAGAACTTT	200
>SoSSR621	CAGCTCGGTCTCAAATCC	GACTGAAAGCTCATCCCTTG	223
>SoSSR622	ACTCCACCTCCTCCTCCTC	GAAGTTGAGCTTGGCCTTTT	100
>SoSSR623	GAATGAAGCCGAGATGAAGA	CGCTGAATCGAATAGAGAGC	202
>SoSSR624	CAGGTCTCGGTCTGAGTAGAA	CCTCCAATCCCTTCTCGTAT	146
>SoSSR625	TGCTTCCAATCTTCCTCAAG	ATGGCAGGCACAGTATCAAT	234
>SoSSR626	TGACGTTTTGGCTCAGTACA	ATAATCTGCGACTGCTGCT	214
>SoSSR627	GCTCACCAATACGCTCAAGT	AATCTGAACCAGGATGAACG	127
>SoSSR628	CTAGTCTGCAGGGAAATCCA	TGCTTTCGAATCTTTTCAGG	249
>SoSSR629	CTCACCTGGAGCTCGATAAG	TGATGTAGTTGCCGGAGATT	235
>SoSSR630	GCCTACAACCTCCCTCCAAT	CTCCGAATTTCTTCCATTT	250
>SoSSR631	CTCCTCATCCTCCCCTCAT	AGTATTGTGCTGCGACTGC	103
>SoSSR632	AAGCCACACAGAGAAAATGG	GGTGTTAACCGACAACCTGA	241
>SoSSR633	CCGAACGACTACGAGGAGTA	TTAATTCCTTCTCCCGATCC	151
>SoSSR634	TGAGTTAAGGCTGGAGGTTG	GTTGCCGACTAGCATGTTCT	185
>SoSSR635	AGAAAGGAGCAATGATGAGC	GTTTCATGGCGTCTGAGTTC	174
>SoSSR636	GCAGCAGTTTCAAACCCTA	CCTCAACCGATTCTGATGTT	244
>SoSSR637	CGGGTTCGATGATCAAGATAC	TCCCCAACCTCACTCCTAT	243
>SoSSR638	AGCGGTATCATGTCAGCAAT	CGTCAATTTTCTTGGACGTT	208
>SoSSR639	CAGTCAGAGATCGACCGAGT	TAGCTTGGACATCCTCTTCG	250
>SoSSR640	ATCGTGGCTTCAATCTTGTC	GATCATGCTCGTCGCTAGAA	162
>SoSSR641	AACCTCCC GGCTATACACTC	CCGGACAAACATGAACAGAT	216
>SoSSR642	AAGCAAAAGAGGGTTCACCT	GGATGATCTGATATTGTCCCTTC	148
>SoSSR643	CTCAGCAGCAGTCCAGAAAT	CAGGCTCCTCCTCTTTCTTC	197
>SoSSR644	TCCATCCTACCATGTCTCT	GATTCGCTGAATTCGTCTGT	218
>SoSSR645	CTGCCTCCACCTCTATTCA	CTCGCAAACATCACACAGAC	250
>SoSSR646	AAAACAATGGAGTGCTCTGG	GTTTCTGTCTCCCCACTTT	164
>SoSSR647	TGTGCGGAGGAGCTATAATC	GAAAATCGGCTTCAAAGTCA	195
>SoSSR648	GAGAGTCCAGCGTCGAGAG	CCGGAATTCTCTTTCTGCTT	193
>SoSSR649	GAGCTCACTGCACTCAATCC	TGCTGCTGGTAAAGAGCATA	193
>SoSSR650	ATCGTGGCAAGCAAAGTAAG	TCATTCATCTTCGTCTTCGTC	215
>SoSSR651	CATTTTCAGACAGCCCTTGT	ATCTGGTGCTGCTCTATTGC	131
>SoSSR652	ACCAGAAAAGACAGCGACAC	ACGATCACCTGAGACAGAGC	188
>SoSSR653	CCTCCCCAAATTGAAAGAGT	AAAGGGTTGGAGGAGATGAG	247
>SoSSR654	TGATGAGCTGGAGGAGAGTT	CCAAAGAGGAGAAAACACCA	140
>SoSSR655	CGGACACACTTATGAGACGA	TCCTTCGCCACTACAACCTC	163
>SoSSR656	TCCACCTATTCGAAGGATGA	AGCAACTGCTGTCTCTCCAG	236
>SoSSR657	CGGGGTTGAAGTAGCTAACA	GCAATCCCACATCAAAGAAA	245
>SoSSR658	AGAAGCCGAATTGGATGAG	CCACACGTCCTTGTCTTCT	223
>SoSSR659	TCCAATCAATACCCCAACTG	GGAGGGTTGTTGTGTGTCAT	102
>SoSSR660	CACGAAATCCCTCTCTCTCA	TTCTCCTCTTGTTGCTGTC	189
>SoSSR661	CCAAGGAAAGCCATCCTATT	ACTGTTTCTGGAGCAAATG	210
>SoSSR662	TGAAGCAAAAGAGGAGGAAA	TGAGGGAACCTTCTTCGTG	159
>SoSSR663	CCGTTTCCAAGAGTTTGATG	CTCATGCCTTGTCATCCAGT	247
>SoSSR664	CAGTGGAGAATGAGGCAGAC	TGGAATGATGATTGTGCAAG	232
>SoSSR665	CCAGTGCCTAAACTTCTTGC	GAAGCCAAATCCAAGAATCA	225

>SoSSR666	AGAGGGTCTTGGCTTGTTTT	CGGAGATCCACAGGATACAG	197
>SoSSR667	CAGGATCCATGTTTCTGGTC	GGTCGAAAAGTTCTCCAAT	192
>SoSSR668	GCCTCCGATCAGAAGAGAA	ATAAAACCAGGCATGAGCAG	239
>SoSSR669	CCCTACTACAGGCAAGGACA	GTCTCGGCTATATCCCCATT	102
>SoSSR670	GAGGCAGAAATGCGAGATAA	CTTCCAGCTGCTTCTTGTTT	208
>SoSSR671	CGTCTTTCCATTTGATGTCC	CGCTTGCAGAAATGATAGGT	228
>SoSSR672	TGCACTCTCTGGAGGTGTTT	ACCTCTTCAGCAAATGCAAC	189
>SoSSR673	GTTGCATTTGCTGAAGAGGT	AAGCCTCAAATACCCAATCC	178
>SoSSR674	TGATCCAGACAGATGGGAAT	ACTCTCTTGCTCGTGCCTTA	229
>SoSSR675	AGGAGAGCAGCAAGAGGAAT	CTGCCCCATCAAAGACATAC	224
>SoSSR676	TTTCATTTCCAACCTTCTGG	CCCTTCTCCTTCCTGTTCAT	209
>SoSSR677	TACGCATCCGAAGAGAAATC	GCATTGTGGAGATGGAAGAG	203
>SoSSR678	AGAATTTGGTAGCGCACAAG	CGGCATTACATATCGATGTTC	195
>SoSSR679	TGGAACAACCTCAACAACACG	TCCTGCCCAAGTAGAGATTG	138
>SoSSR680	ACAATGAGGGATGGTTGTTG	TCCAAGAATAAGAGGGACGA	125
>SoSSR681	ACCCACTACTGTCATCACC	AGAGAAGTGTCTCCGCTTGA	144
>SoSSR682	CAGGATAAGAACACCGAGGA	TCGGTTAGAGTCCACAGCTC	202
>SoSSR683	ACGAGCTCAAGAGGACTGTG	ACAGCAAGCAGTTGATCCTC	240
>SoSSR684	CGTGGGGTAAAAGAGGATTT	GGCCAGACTCTCTCATCA	235
>SoSSR685	CTCTACTCGAGCCACCTCTG	CAAGGAAAGCGAGAGATTCA	183
>SoSSR686	GGTAGATCCGCTTTGGAAAT	GCACCTCTTTCCTACCATA	196
>SoSSR687	CCTCCTTCCCAACTAAAGA	GAGCTGTCTGCTTCTGCTTC	126
>SoSSR688	TGAAGGTGTTTTGGTGTGTG	GAGTCGACCGTGAAGAAAAC	225
>SoSSR689	TGGGGAGATTCAAGAGAGTG	AAGCTAAATCCCCACCAAAC	113
>SoSSR690	TAATGGTGTGAGGGTTTCGT	ACCCTTTICTCTGGTTTTGC	191
>SoSSR691	TGCTTATGCAGATGGACAAA	CCGCTATCCGGAATATCATC	192
>SoSSR692	AGCGACAACAACAACAACAA	GATATCCGCACGTTCTTCC	200
>SoSSR693	TGAACATGGATGAGCTTTTG	ACCTCATCAACCGTCTTCTG	205
>SoSSR694	TGTGAATGGTGTGAGGCTTG	GGTCTGTTTGGGAAGGATCT	112
>SoSSR695	AGACTGGGGATTCCGTAATT	TATTTGCAGCTTCCACATGA	247
>SoSSR696	AAACCCACGACATCCTC	TTGGTTCTAGGGCTTTCCTT	125
>SoSSR697	GATCTGAGCAGGTGGAAAAG	GACAAATCCATGAAGCAACC	190
>SoSSR698	AAGTACGCGTTGAGGATGAG	AGTAATGCTTGGCTCCCTCT	238
>SoSSR699	AAATTCTCGACCCCTCTCC	GACCAGACAAACGGACAATC	241
>SoSSR700	CACAACCTTCTTCCCAACCAA	GGTTGTTAGGGTTGATCACG	136
>SoSSR701	AAGGACCCATGAAGTCATCC	CTAGATCCGGGCTTTCCTTC	178
>SoSSR702	CAAGTTCGGCTCTAAGGTCA	TCTACTCGAAAACGCCTCAC	130
>SoSSR703	AAAGAGGAAAAGCCAAAGGA	ACGTGTATCACGCCATACT	132
>SoSSR704	AGCTGTACACGAAAGCATCC	ACACGAACTGGAAGACAAGC	105
>SoSSR705	GAGCAGAAAGGAGAGCTTCA	GTAGCGATATGCCAGAGGAG	102
>SoSSR706	ATGTAGAGCAGGCTGTGAGG	GAGGTAGCGCTTGTAGATCG	242
>SoSSR707	ATGAACCTTCCATTGGGTTT	CGTAGATAGAGGAGCCACCA	123
>SoSSR708	ATGGCGGAGAAAGATAGGAG	CCTCAGATGAGCGTGTTTCT	144
>SoSSR709	ATCGAGTTTGTGTGGGGTAA	CGTTCAGATTCGTAGCATT	209
>SoSSR710	CAGCAATGTTGAAAGAGCAA	CAGCCTTCTTCCTTTCCTTC	183
>SoSSR711	GACAACATCCTCTCCACCAC	AACACGTGGATTGGCTTG	239
>SoSSR712	GCCAAAGAAGGATCATCAA	AGCGGAGCCTTTTAGTTCTC	203
>SoSSR713	TGCTCCTGTAAGCATCAACA	TTTTGGGATCTCCCTTCTTC	215

>SoSSR714	TGATTCCCAATCTCCACACT	GTAATCCCGATCCGAGAAAT	172
>SoSSR715	ATTTCCTCGGATCGGGATTAC	GCCACCACTGCTCAAACATAT	173
>SoSSR716	CTGGAAACCATTGAAGCAAC	TCTTCACCAGCCTTTTCTTG	163
>SoSSR717	TAAATCCGCTGTCTTTTCAAG	TCCTTCTCTTCCCTCCTTTGC	144
>SoSSR718	CCTATTCACCCAACAGCAGA	TACGGCTCAACAGCAATGTA	226
>SoSSR719	CCTGTCTGTAGGCATGAGGT	TCAAGGAAATCAAGGCAGAG	245
>SoSSR720	GTTATGGAGCCTTCCTTGGT	TCAGGACCCAAACCTCAATA	192
>SoSSR721	ATCTTTACGCGAGTGGATTG	GCTGCCACAACACTACAACAGA	214
>SoSSR722	GTTTACTCGCAGCTCGACAT	GCGGTAGTCCCTCACAAAC	177
>SoSSR723	GCGAGTTTGTGTCCGGTGTAT	CGTCTCCTTCTCCACCTTCT	104
>SoSSR724	CGGCAACAGAAGAGATCCTA	CCTCTCGCTTTTCTCTCTCT	228
>SoSSR725	GATAGTTCAGGCGCCATTAG	TCCCATACTCAAGTCCCAGT	218
>SoSSR726	CTTGTTGCAGAGGCCACTAT	CTTGGAGTCAGGGCATAAAA	191
>SoSSR727	AATGGCACCTGTCCATCTTA	GGGTTTACTAGATCGGTGA	141
>SoSSR728	TCTCTCGCCAAAATAGCATC	AAACCTCATCATCACGGCTA	234
>SoSSR729	GTCTACGCACGACCCCAAT	GTGCTGGAGTTGCTTGATG	178
>SoSSR730	GCAGAGTGAGAAAAGCTTGC	CAATCCTCCTCCTTACCTT	216
>SoSSR731	CAGCTTCTAATCCTCCTCCA	TTCCAAGCTTCTGTTTCTCG	222
>SoSSR732	GAGAAAGATGAAGGGGTCTG	CCGACTCACTACCAGAGGAA	220
>SoSSR733	AACAGCAACAGATGCAACAA	CTTCTGTCACCATTCCCATC	186
>SoSSR734	TGGGATACTGAAGAGGCTGA	TTCAACAGCCTTTGTTCCTC	193
>SoSSR735	AACAAGCCCCACTCCTTC	ATCCATCCAGCTCCCTCTT	223
>SoSSR736	CCAAGATGAGGTGAATGAGG	CATCGACTTTCCAGTCCAAC	125
>SoSSR737	GCGTCAAGCTTATCAACCAT	TGTTGCATCTCTTCCAATCA	246
>SoSSR738	TAATTGATGAAGGGGTGTGG	TAGGGTTAGGCTCCCAGAAT	153
>SoSSR739	CCGACGTCAAAAAGAGACCTA	CTAGGGTTTAGGGCGGAGAC	232
>SoSSR740	CAGTTTCTGGTTAAACGCTTG	ATTGCTCCTCCTTGCTTCTT	234
>SoSSR741	ACACTGGAAGTGGAAGTGGA	TATGAGGGTGCGTTTCTCTC	129
>SoSSR742	CATTGGAGCAGGATTTGTTC	ATCATCCACGGCTTCTCTCT	250
>SoSSR743	ACGTATGGAGAGGTGGTGAA	AGACGTAGACCACGGAGTTG	220
>SoSSR744	GGTTCACGAATCTCCACAAC	GAATAACACGCCCATGTCTC	111
>SoSSR745	GCCTCCGCTTCTTCTCTCT	TGAAATTAGACCGGCTCAAG	128
>SoSSR746	CTTGACGAAGTGGGATTGAC	ATCATTATGTGCGGATGCT	243
>SoSSR747	GAGCCACAGAGGAGAACAAA	CCAAAGCCAAGCTCTGAAT	219
>SoSSR748	AGGGGTTATCCTCAACAAGG	CCACTGATTCTGGATTTGCT	153
>SoSSR749	CATTACGTGCCGTACTACCA	GCGAATTATCTTGGCTGAAA	213
>SoSSR750	ATGCCAAATACAGAGCAACC	CACCACCACTTTAGCACCTT	213
>SoSSR751	TAACCCTTCGTCGGAGTACA	CTCATCTCGCTGCTCGTC	231
>SoSSR752	TGCCCAGATGAACTTAGGAA	TAAATCAGCTTGGGATCCAG	172
>SoSSR753	CAAGTGATGCAGCTTTGCTA	CCTGCTGTGACTTGAGCTTT	219
>SoSSR754	GAATTGAATCTGCTCCTCCA	GGGTACATAACCTCCCCAAG	163
>SoSSR755	GTCTTGGAAACAGGCAATC	ATCACAAACGTGCCCTTTTAC	155
>SoSSR756	GCTTGTAGGTCTTGGGTTGA	AATAGCGTTGAGTTCGCATC	148
>SoSSR757	GCGATTCTAAACAGGATGGA	TCCTAGCAAGAATCCCTTCA	247
>SoSSR758	TTGCCAAGGAGATCTTCAAC	GCTTATCCCACACATCATCC	213
>SoSSR759	CCAGCAACAGAAGGAGGATA	TGAATTCCTGGAGAGACAGG	137
>SoSSR760	CTTGATAATTTCTGCCTCTTC	CGTCGATGACGTGGATCT	227
>SoSSR761	ATTTCCTGCCTCTTACCAC	CGTCGATGACGTGGATCT	220

>SoSSR762	GGAGGTGGAGGTAGCTATGG	GTTCTCACCGCACTTGTAGC	135
>SoSSR763	GCTCTTCTATGGCTTGCTTG	AGTTGAATGACGGATCCAAA	213
>SoSSR764	AAGACCCTTCCTCCACCAC	GGGCCGTAATTGTAGGAGAT	160
>SoSSR765	GTTTTCGAACTGGAATGGTG	ATGGCCAACGTGTAGTAAGC	108
>SoSSR766	TCAGCTCAATTGACAATCA	CTCCATCTCCATCTTCTCA	228
>SoSSR767	GATTGCCTTCTGCCTGAATA	TCATCATCATCCTCCTCAGC	156
>SoSSR768	CTACGTGGACTTGCCTCACT	TGCTTCTCTCTCCACCTCAC	247
>SoSSR769	GTGAGGTGGAGAGAGAAGCA	TCCTTCAAAGCTTCCAACAG	160
>SoSSR770	AACCTGTTGCGAGCAGTATC	CTTACCACGATCTTCATCC	104
>SoSSR771	TCACCATCAACGAGTACCAG	GGGCTTGAGAACCCTGTAGA	206
>SoSSR772	TCCCTTTGGTGCTTTCTATG	AATGTTACCACCACCACCAC	208
>SoSSR773	GATGATGATGACGATGACGA	CAACCGCTCATTCAAATCT	249
>SoSSR774	AGGCATCGTTTAAGGCTTTT	ACACGCAGAGAAATTTGCAT	221
>SoSSR775	AAATTTCCACCAACCGAGAC	CCGCCTTCTGAGACAATAAA	212
>SoSSR776	CCAGACCTCGTCAATCAATC	CTCATGTCGCCGAGTAGAGT	105
>SoSSR777	GTCTGCCAGGAACCTGTTTA	ATCTCCACCACCAAAGACAA	199
>SoSSR778	AGATAATGCAGCAGGAGCAG	GCCTACGAGCAAGAACGATA	185
>SoSSR779	GTGGGAAACACAGGTTATG	ACTAACCATCTTGGCCTTCC	233
>SoSSR780	CTTCATTGGTGTCGTTGGTT	ACGTTGTCTGCGTTGTATCA	233
>SoSSR781	CGAATTTGAATCCGAAACC	CTGTTCAGCCTCTCGATCTC	220
>SoSSR782	TCTGGCGTTTTCTTTTGAAC	ATCTCGCAGACTCACTCCAC	144
>SoSSR783	GACGAAGAGGAATAGCCACA	AACCACTGCAACATCCATCT	250
>SoSSR784	CCGGTGAAAATGGCCTTC	TCCAGCATCTTCTCCTTCTG	182
>SoSSR785	AAGAAGGGCAAACGGTACT	CAAAGCCCACAGCTTGTATT	125
>SoSSR786	TCGCATTTGGGTTAACATCT	TAATAGGCAGCACCTCCTTG	218
>SoSSR787	TGCTTGGTAGCTCAATATCCA	GTGGCAAGAGCTCAAGGTAA	143
>SoSSR788	CAAAGACTCCTCCGATTTCC	AATGGATGGTGGTTGATGTG	228
>SoSSR789	TTTCAAGAGACTGGGAGCAC	GTTTCTGCCTAAGCGTGTGT	158
>SoSSR790	GAAGAGGAGGAGGAGGAAGA	CCAACGATCTTTTGGATCTC	225
>SoSSR791	ATTTCAAACGAGACCTGCAA	CCTCAGCAGTGCACACTTTA	205
>SoSSR792	AATTGCTCAACACGATGGAT	GTAGCCTTCAAACCTCAGCA	161
>SoSSR793	CAAGATTTTGTAGCCCTCCT	AGTAGGTGGAGAGGGTTTGG	167
>SoSSR794	ATGCAGGAAATCCAGAGACA	TGATGTCGACGTATGGTTTG	149
>SoSSR795	CCACAAGGCCAATGAAATAC	GGAGTCCAATTGTTGTCTGG	156
>SoSSR796	CATCCCAAAGAGATTGGTG	CAAAGTAAACGGCCCTAGT	221
>SoSSR797	TCTCCTTTGTTTGGTTCTCG	TCCGATTCTTCTTCATCAGC	236
>SoSSR798	CATGTCAAGAATGGCTTTGA	ACGCGTACGACCGATAAGTA	132
>SoSSR799	GCTGATGATCTTGGGACAAC	ACCCATCTCCCTCGAGATAC	250
>SoSSR800	GTGGGGAGATTTCTTCTGT	ATTCTCCATTCCCAAACCTCC	148
>SoSSR801	AATGATCTTGGGACAACACG	ACGCCCTCCCTTATCAGAAAC	133
>SoSSR802	ATGTTGTGTGCGAACAAGTG	AGCCAAGAATTGATCCCTCT	187
>SoSSR803	TGGTGGAGACAAAGAGAGGA	TCTCCAGAAGCCACAAAGAT	183
>SoSSR804	CCGTTGATGAAGCAGTATCC	ATTGCAAGAGGATTTGGTGA	189
>SoSSR805	CAGAGACCTTCTTCCGCTTC	CAGCCCCTCGAAGAAATC	188
>SoSSR806	CTAGCAGGCGAGAATGAAAG	ATTTCTTAGCGCTCCTTTGG	193
>SoSSR807	GCTAAGGCGTATCGCTTCTT	ATGGGGAGGAATTTATTGGA	238
>SoSSR808	CTCCAGATCCCAATCACAAAC	GTTATTTCCAACGCTTTTGC	161
>SoSSR809	GTGCACTAGATTGGGGAGTG	GTTGGTCGATCTCGGTAGTG	164

>SoSSR810	CATCAGCCACCGTCTGTC	GGTGCTCTGAAACCAAGAGA	211
>SoSSR811	CTTCAAGGTTTGCATCT	GCCTGAGAGGACTGTCTTGA	197
>SoSSR812	TGGAACAGGTCCTATGGAGA	TTGATGAAGTTGTCGGGTTT	215
>SoSSR813	GCTGTTGCTGAGGATTTGT	ATGAAACGTCTGCGATAAAG	182
>SoSSR814	TATTCCTCCTTGGATTGCAG	GGAGTCTCTTGGTGGTGTG	232
>SoSSR815	GGAGGCGTTACAGTACCAGA	CTTGAAATCCTCGTTGAGA	191
>SoSSR816	ATGCTCAGACCTGCAAACTC	GAACAAAGTGGGACTCGATG	138
>SoSSR817	TCCTCACCACCACCTTGTAT	CTTCTTTGGATGCGACAAGT	174
>SoSSR818	TAGGACGTTCTCCATTGCTC	ATCTTCCCATCTCCTTCCAC	234
>SoSSR819	GACCAAAGGCATCATCAAAC	CTAAGGTTTTGGGAAATGC	216
>SoSSR820	GTCGAGAGGGAGGAAAACCTC	GAAATCGGAGAAGGAAGCTC	200
>SoSSR821	AAATCAAGAGTTGCGTCTGG	CGAGCTCACGTTTAGAGGAA	245
>SoSSR822	GAGCTGCGGTAAACCTCATA	CAGTGCTTCTCGAATCCTGT	183
>SoSSR823	GGCAGAATACCAACTCGAAA	GGATCATCAGAAGTGGTTGG	123
>SoSSR824	CAGCTCGAAACCCTTGTCTA	TCTCCCATGGAAACCTACAA	184
>SoSSR825	CGCCTTATCATGCTGAGTTT	CCTTCTCCTCTCATCGATCC	213
>SoSSR826	AGAACATCAAGATCGCCAAG	ATCGCGAGAGAGACGAGAGT	247
>SoSSR827	TCCTGATTGTAACCGTCGTC	AGGCATTTCTTCGTTGTGAG	172
>SoSSR828	CCATGGGCTTATTCATTTTG	AGTTCATCATGGCGTAGAGG	167
>SoSSR829	TCAAGTCTCAGCCTCTCTCC	TCTTCTTCGAAAATCCCTGA	211
>SoSSR830	CTCCACCGCCCTAATCTC	CTAGGGTTTTGACGTGTTCCG	193
>SoSSR831	AGAGCCTTGGTAAGGGAGAA	AATTTGCCTGAAAAGCACAC	217
>SoSSR832	GCAGCTCTCCTCCTTCTT	GGGATTTTGAAAGGGAGAAA	200
>SoSSR833	AAGACACCATCGACTCTTGC	TGGAGCTTTAAGGGTTTGTG	229
>SoSSR834	TAAAGGCCCTCATTGACAGA	ATCGAGCAAGTGACATGGAT	249
>SoSSR835	CCGGAACCTAAAATTCCTC	GAGTAGCGTTTTGGGATGGT	214
>SoSSR836	AAGAAGGCTCCCTATTGGTG	GACCTCCCAAATCGAGGTAT	241
>SoSSR837	TCTAATTGCCGCTATCTTGC	TTTCTCGCGTATCCTTTGTC	238
>SoSSR838	GGGTTCCCTTTCATCTTCAT	GGAAGTACGGGGAGAGTGAT	202
>SoSSR839	TCATCACAAAATGGCTCTCA	ATCGTGCGTTTTTCATGATCT	227
>SoSSR840	GCTCATGCTCCTCCACTTC	CACGTCCTTGTTGTCGTAGA	154
>SoSSR841	AGATTATGTGCCCTTGGTTGG	CCATCGTCACCAACTCTCTC	144
>SoSSR842	TGGAGATTATGTGCCCTTGGT	CAAGGATTAGGGCTTTGATG	245
>SoSSR843	TCCTCTTCATCCTCCTCTCC	AGCTTGGGGTTGAGAGTTT	178
>SoSSR844	TGTGGAGAAACACTCCTTCC	TTCTGCGTCTCCATGAAAGT	220
>SoSSR845	CGCGATAAAGCCATAGACAA	TTTAGGCATTCCTTTGGTGA	150
>SoSSR846	GAGCATCATCATCACCATCA	TGATCAACATCAAGGAGGGTA	126
>SoSSR847	CGCTCTCCCAACTTCAAATA	AGGAGGTGTTGCTTCTTGTG	235
>SoSSR848	TCCGTCCTCGCTCTAAATAA	GGCGACCCAGTAGAAGATG	210
>SoSSR849	GGTTCAGGAAGCGTTCTACA	TTTAGCATGGACCTGTTTCGT	216
>SoSSR850	AGACCTACATCGCAGTCCAG	CCTACAAACATGGCAAGTCC	223
>SoSSR851	AATTTGAAGTGGAGGAGATGG	ATTGGAGCCACGTTAGATCA	241
>SoSSR852	CACCCAAGAATTGGTCAGAG	ATACGCCAATGATGATGAGG	184
>SoSSR853	GGAAAACAGAATCCCACAGA	GTTTCTCAGGCCCTCTATC	176
>SoSSR854	CGGACTTGGTATTCCTCTT	CCTTCGGCCCTTAGATTTAG	175
>SoSSR855	GCAAAGAATTTGGTGAGGAA	GTCGCACTGGATGTAGCTG	201
>SoSSR856	GTGCTGTTCATCTTGGGATCT	GCCTCTGGTGTGGTTTAATG	224
>SoSSR857	CATACCTAGCGACTCCGATG	GAAAAGACGCCGTATCTCAA	159

>SoSSR858	TCACCTTGGGACTTTTTCTTG	TATTGCGCATACGAGTCAAAA	137
>SoSSR859	CGCATTTCCTACTAAATCCAC	ACCAACCTCCTGTTTGTGTCAG	204
>SoSSR860	CCCTTTCCAGAACAACCTTCC	GGGTTTTGTGGAGTATGGTG	182
>SoSSR861	GAATCCGAGAACAAGAAGCA	TCCGTTCCGTTGTCATTACT	107
>SoSSR862	GCTGCTCAAACCCCTATCAA	GCCAGAGAGGAGAATTGCTT	119
>SoSSR863	TTTCAGCTCAGGACTTTTTGG	GAGAATGGGATTTTCGCTATC	204
>SoSSR864	GAGTTGATGAGGGGAAGGAT	CTAGCCGAGCTCAATCTACG	160
>SoSSR865	GCAATCTTCAATGACTCCTCA	CCTCCTATCGAGCCTCTTCT	250
>SoSSR866	CAAGTGAGGTGGATGGAAAG	CGGAGGAGGTATCCAAAACCT	215
>SoSSR867	TGTGGGGATATAGGCAGTGT	CCATCACCAGGAATGAGAAC	211
>SoSSR868	CACTTTTCTGCGATGAGCTT	GAAGGCATGAGCTTGAGAGA	237
>SoSSR869	TTGGGGATGGAACAGTAGAG	AAGCCTCGGTACCAGTCAAT	203
>SoSSR870	TACGACAGCTACGCCTTCTT	CAACGGTATTCCAACACTCC	232
>SoSSR871	TCTCCAGTTTCGGACAAGAG	TCCCTTCATCATCATCATCC	192
>SoSSR872	TCGTCGAGTACCTCCAAGAC	GTCAAAGCCCTAACCACCTT	187
>SoSSR873	GGGATTGTGAGGCAGGAG	GACTAGGAGCTGGTTCACGA	234
>SoSSR874	CGTCGGTTACTGGAAGAAAG	GCCAATAATCATAGCCCGTA	229
>SoSSR875	CGTCGGTTACTGGAAGAAAG	CGCCAATAATCATAGCCCTTA	230
>SoSSR876	TGTCTGGGTTGAAACAGTC	TGTAGGACGACTCCAGCATT	189
>SoSSR877	CCGCTATCGCTTCTGATGA	TGGTGTTAATGTGGCTAACG	193
>SoSSR878	CTATGCGGAGGAGTGCTG	ATAGCGGCTGTCCAATCC	123
>SoSSR879	GCCGCTATCGCTTCTGAT	TTCTCCAGCATATCCCTTCA	223
>SoSSR880	GATCTGTGGCTGCTTAGGG	TTCCTCCCTCTCTCTCACCT	156
>SoSSR881	CTTGTACTGGACGCTCGACT	TTGGGGTTTATGAAAACAGG	201
>SoSSR882	ACTACCCGAGATGCATGGTA	GGAAGGCAACTGTTTTGAGA	194
>SoSSR883	ATGGCTTCCATCTCCACAT	TTTTAAGGCTGGTGAATTGC	138
>SoSSR884	CCACTCCCCTCGATTTCTAC	AGGAAGACGAGTAGCACGAA	171
>SoSSR885	AATCGCCATTTCCTCTTCT	GCATCGCTGCTAGTTTCTTC	248
>SoSSR886	AGGTTTCATGAGCTGAGACG	CATTACAATCCCTTCCGTA	205
>SoSSR887	GATCACCACCACCACCAC	AGGAAGAGGAACGAGAAGGA	233
>SoSSR888	GGCGATTCTCACACTAGGAA	TAGCGTAATGCCTGAAATCG	137
>SoSSR889	GATGTGGAAGGTCACACACA	TGGAATGGCTATGGTTATGG	241
>SoSSR890	CCATAACCATAGCCATTCCA	GTCGAGATGACCACTCCAAC	225
>SoSSR891	CACCTCATGCTTCTTTTGGT	GTTTCTGGACGCTCAAACT	163
>SoSSR892	CGAGTCAAGTTGGAGAGGAA	TCTAATATCCCCTCGCCTT	224
>SoSSR893	TGAACCCTTCGTCTATGAGC	GCATTTGGAATGAGCTTTGA	220
>SoSSR894	CTCTCTCTCGTCTCGCTCTC	CAGGCAGCTCCAAATTTATC	223
>SoSSR895	GCGTCGCAGAGGATAATAAG	GATCATGTGGGACCATGTTT	116
>SoSSR896	AGATTAATGGCGATGAGTGC	TGGGACCACCTCAGAATTTA	176
>SoSSR897	GACGGGCTGATACTGAAC	CCTCCATCACCTTCTCCAC	200
>SoSSR898	GCCACATGAGGGTTCATTTA	TTCCTCACTGTTGGATTGGT	149
>SoSSR899	GGAAGAGGTCGAGGATTAGG	TCTCATCATCAGCAGCAAAA	246
>SoSSR900	TGATGATGGTGGGAAGAAGT	TCAGGATCAGCTCCATTTTC	228
>SoSSR901	TCCTCCTCACCATTTGTGCT	ATTGAGCCAAACAACGAGAG	230
>SoSSR902	CGGCAACATTGATCAGTACA	TAGCAAAACCGAGTGAGGTC	124
>SoSSR903	CCAAATGGATTTCATCACCAT	TTGAACAGCATCATCAATCG	248
>SoSSR904	GCGTGGAGAGAAAGAACTG	GGTGATCAACGTCAAACGA	222
>SoSSR905	AACGAATGCTTGAGCGAAT	CACACTGCAGGTTACAATGG	213

>SoSSR906	GGCCATTCACAATCAAGCTA	TATATCCTCCACCACCTCCA	239
>SoSSR907	CAGATGGACAATGCACAAAA	CAACCGATCTATGCTGTTCC	240
>SoSSR908	AGCAGTAGGCATTCTTGTGG	CACTTCCAGCCAGTTGAGAT	191
>SoSSR909	GCTGGCAGAGCACATACTTT	CCGTCATTATCATCCACCTC	125
>SoSSR910	CTCGACTTTCTGCAGCTCTC	TCCTCAATCAGGTCTTGCTC	107
>SoSSR911	CTTTTCCTCATGAACCCTCA	ATTTTCGCTAGGGTTGCTCTC	212
>SoSSR912	GAGTCGTGGTTGGATGAGTT	GTCTCCAGCAGCTCTCTCTG	233
>SoSSR913	TCTTCTTCGAGCAGCTTCAG	CACCACCCTATCGTTTTTCAT	152
>SoSSR914	AAGCCGAGCAAGACGAGTA	GGTGTAGGAGTAGGGGTGGT	245
>SoSSR915	ATCTTCACTCCACCCTCCTC	GTCGGGATTTTGAGGTTCTT	175
>SoSSR916	TTCTCATTCTCCCCTCAACA	CGGGAACTTTATCATTCGTG	232
>SoSSR917	CGTTCCCCAACAACTACTC	AGCCGCAGTTGATGAAGTAG	108
>SoSSR918	TGAATGGCCTCCTTTGATTA	TGTACTIONCGGAGGCGAAG	172
>SoSSR919	AAGGAGAAAGAACCAGCTGG	GACCTTCAGCTTCCAAGGAT	228
>SoSSR920	GTACGGACAGCGGAAATCTA	GTGAGGCAGAAGGGGTTAAT	112
>SoSSR921	TGGGGTGTATTTCTTGTGCT	TTCCTTCTCCGTCTCTCTT	107
>SoSSR922	AGGGCAGAAGAAGAAAGGAA	AATCTTTTCCGCAGCCTCT	204
>SoSSR923	CCGTGTTTCATCCTTCTCAGT	CCCCTCCAATATTCATGA	133
>SoSSR924	GCCAGCCTGTCATCATCTAC	GCTGGTCTTCTCCATCGAC	188
>SoSSR925	ACACCACCGTCAACGACTAC	GAACCTCCACGTTTCGTGTCC	233
>SoSSR926	GATGCTCTCGTCATCGACTT	CACCTGTACGAGGAGGACTG	104
>SoSSR927	ACTGGGCGTTCCTCCATATT	TCATCTCTGTGGACGATCAA	224
>SoSSR928	CCCCAAGAGAAGTGGTCATA	AGGGTTATTCATCGGGAAAG	190
>SoSSR929	AGTCTCAGCCTCTCCCTACC	CAACGGCATAAACACCTTCT	213
>SoSSR930	GTTACATGGCTGGAATGGAG	GATACTGCTGCTGTTGCTGA	102
>SoSSR931	TATTCATCTTTCACGCCACA	TTTGAGGGAGCTCTTTAGGG	154
>SoSSR932	CGATAGGAGTAGCACGAGGA	GACGACGATGACGAGAAGC	237
>SoSSR933	CGCCATTA AAAAGCTCGATTA	GTGTTGCCGTCTTTCTTGTT	201
>SoSSR934	CTGGTAGCGGCAAGAAATTA	GAGTGGAGTCGAGAAGCAAA	186
>SoSSR935	AGAAGGCGAACAAGGAAGAT	GTTGGTGGTGAACACAATGA	188
>SoSSR936	ATGGAGTGCTGGTTGAGTTG	GTGGAGGTGTAACCAGATGC	197
>SoSSR937	GGTGCAGTGCTATTTTCTCG	CTACGGATGCTTGCAGATTT	224
>SoSSR938	GAATCATCCAACCAAACACC	AGGATCCGAGCTGATTCTTC	157
>SoSSR939	TGTATTCTGGGAATTGAGTGC	AATCTCTTGGCAGTGTGGAG	237
>SoSSR940	ATGGCTGCTTCATCTCTCAT	GAGAATGAGGGGTTTGAGG	102
>SoSSR941	TCTCTCTGACTTCCGTCCTC	TTGATGCCCTTTCTGAACTC	112
>SoSSR942	GAGAGGAAGACCCCAACAAT	TTCCTCCTCGTTATCCTCCT	125
>SoSSR943	CCCTTAGGAGATTCGCAGA	GTTCTCAACAAACGCATTC	249
>SoSSR944	GCTGCTGCTTCTCCAGATAC	AGGATTTGGGAGAGGAATTG	176
>SoSSR945	GGGATCAACGCAGACTTGT	AACACATCCAGTGCAGAAGG	245
>SoSSR946	CGGAAGAGCTGAAGATGCT	GTGAGCTCGTTGGTAGAAGG	215
>SoSSR947	CATCGGCCCTATCAGCTC	TATTCGTGATCGACGTCTCC	230
>SoSSR948	AGCACAGAAAACAGGCAATC	AAGGGCTCCAGTGTAATTC	205
>SoSSR949	GCATTCTCTCCACTGCCTTA	GCCTTTGAGTTGGCAGATAC	229
>SoSSR950	ACAACGCCTTTGACTCTGAC	GGAGATGCTCCCTCTCTCTC	246
>SoSSR951	CAGAGGCATTGGTTTCATCTC	CTCAATCCGACATCCAAATC	133
>SoSSR952	CGACAACA ACTGGCTCTCC	CCACCCAGTGGTGTAGAAGT	130
>SoSSR953	CCACTGAGTTGGTTCGCATAC	ATGCGCTTCATATCTTCCTG	230

>SoSSR954	CATTTTCAGCAACAACAGCAG	TCATTGCCGTGAAGGTATTT	222
>SoSSR955	TCAAGATAAGGCAGGCATTC	CCTTGAGCTATTTGCTTTCG	140
>SoSSR956	GTCGGAGATGATGACGAAAG	ATGTTTCAGATTCACCGCACT	243
>SoSSR957	CATTAGGTCTCTCCGCAACA	GACATTGGGCTGAATGGAG	162
>SoSSR958	TGATGGATCAGTCCACCTCT	GATACAGTGGCGCTATCGTT	195
>SoSSR959	AGGCTACCATTGCTGTGTTT	AAGAGCACGAATTGCTTCAC	101
>SoSSR960	TTGGAGCACAGCAGTATGAA	GCGATTGGAGAATAGCGTAA	210
>SoSSR961	GCAGATTTCTCCTCGATCT	CTCTTTGAGGTTCTCGTCA	168
>SoSSR962	AATAACTCCACCCACGAACC	TGGTGTGTCGCTTTTGTG	247
>SoSSR963	TCCATCTCCATCCAACAATC	GGTTTCTTGGCGACCTTATT	131
>SoSSR964	AAACGCGAATGCTAGGAATA	TCAGAACCTCTCCATCTTCG	140
>SoSSR965	ATCAGCTCGAATTTGGTGAG	CCAAGTCCCAATGCATAAAG	191
>SoSSR966	CGCTTCAGAATGATCGTCTT	ATCTGCCACATCAAAGTGGT	240
>SoSSR967	ACTCGTCTCACTGCAAAAGC	GGAGTTCGTCTTCGTTCTCA	174
>SoSSR968	GCCTCTTTCATGATTGGTGT	AAGAACCCAAGACCCAAGAC	248
>SoSSR969	GTCTTGGGTCTTGGGTTCTT	TCTTTCTTGTCGTCCTCGTC	218
>SoSSR970	AGAGGAAAGCAACCGAATCT	CTCCATTTAGCGTGTCTT	231
>SoSSR971	AGAGATTGATTCCGATGGTG	GGCAGGGAGTCCAGTAATCT	247
>SoSSR972	GCATGCTAACCTCATGGAAA	CTGGGAGGCTGAATATGAGA	237
>SoSSR973	CGCAGAAAGAAGCAGATAGC	ATGTTTGTGGTTGAGCCAGT	161
>SoSSR974	TGCTGTCAATCAGCAAGAAA	GGGCAGGTAGTTGGAAGAGT	149
>SoSSR975	CGTGATTTACCTCCAATTC	CCAACACCAACACATCTTGA	235
>SoSSR976	AGACACCAGCACAAAAGAGG	TAGCTTGGGATTCATCTTCG	149
>SoSSR977	CGCTATTGCTTCTTCTCTG	CAACCTGCCTACTCCTCTCA	200
>SoSSR978	ATTATGTTCGATGGGTTTCAAG	ATAGGACAGCCAGATGATGC	200
>SoSSR979	CCATTTGCACAACCACATAA	TCGAAGGACTTGAATTGAGC	237
>SoSSR980	GACGAAGATGATTCGGAAGA	TAAACCTGGCCTTCTTAGGG	217
>SoSSR981	TCGACTCTGAGGATTCTTGG	TCCCTCCCATTGACTGATAA	175
>SoSSR982	GAAGATCGCTGATCGCAAC	CGCTCCATTTTCCAGAAGT	189
>SoSSR983	CCGCTACTACTCCCACCAGT	TTAGCCAAGAGAGCAGCAGT	222
>SoSSR984	GGTGTGAGGCAACAAAAGAT	CTCATCAACAAAGTCATCCTCA	126
>SoSSR985	GTTCTGATGGGGTTGACTTG	CCTTTCTGCAGTGAATGCTT	140
>SoSSR986	ATCGTAGCTCCCGACTTCTT	TGTTTCGTCACCTCCATATC	241
>SoSSR987	CTATGCAACTGGCAACACAC	CGCTGAAGCTCCACATCTAT	146
>SoSSR988	TGGGGAGTATCTGAGCAAAA	GCTCAC'TTCTTTCCCAACCT	244
>SoSSR989	GTCGGAGAACAAGGAGACAG	TGGAAACTGATGATGACGTG	220
>SoSSR990	TTTGGAGGAGATCTTCAACCT	CACAAAATCCGTTTCAAGTCC	242
>SoSSR991	TGGCTTCTGCTACATTCTCA	AGTTCAAATTGCGGTTGCT	250
>SoSSR992	AAGTCAGTTGAATCCGTCCA	ATCCAGCTTCAGCTTTGATG	199
>SoSSR993	TCAGCTTCCGATGAAGAAAG	GTTTCCCCAAAATCATCCTC	183
>SoSSR994	AGTAAGAAAGCAGGGGAGGA	CCATGTACTGTTGCTGGTTG	226
>SoSSR995	CCTCTCCTCCAATCCCTAC	CGAGAAGAATCCCATGAAGA	166
>SoSSR996	GAGAGGGTTGTGATGAGG	GGCTCCAAACACTTGAAGA	141
>SoSSR997	AGGGTGTGCCTGCTAATATG	GGCTGCAATAGGTTTTGTTG	232
>SoSSR998	TCAGCTTCGTCAACACTTCA	GTTTGCAGTACTTGGCTGCT	222
>SoSSR999	CCCAGCAACTAAAGCAGTTC	TTGGCCAGAAGAGTTATGCT	230
>SoSSR1000	CACTGCGCGGACTAATAAAT	ATGCTCTTCTCAGCTCCTT	198
>SoSSR1001	AGATGATCGAGGAGGAGGAG	CGACTCGTTATTTGCCAGTT	101

>SoSSR1002	TTTGATCTGAACGCTGGAGT	GGTAGGTCCTGTTGATGTCG	222
>SoSSR1003	GCATCGACTGAGAAGCAGAT	TCACCCACCTTCAAATGACT	170
>SoSSR1004	AGTCATTTGAAGGTGGGTGA	ACTGCACCGTCTCTTCTTG	140
>SoSSR1005	GAAATGAATATGGCCAGCAG	CCAAGTCCACAAGGATCAAC	106
>SoSSR1006	GTTTTGTGCGTGTGTGTGAT	ACCACATGCTTAACCAATCC	107
>SoSSR1007	AGCAACCACCCTCCTCCT	CCGGGAAACTTGTGAGATT	178
>SoSSR1008	CAAAAGAGGTGGGATGACAG	ATCTCCGTCTCCCATAAACC	163
>SoSSR1009	TGAATCTTTTGGGACTGAA	ATATGTGATGCTCCCGTGTC	205
>SoSSR1010	CACATGTTCGAGAATGCAGAG	ATTCAATGTCACCACCGATT	240
>SoSSR1011	CGACAAATCGTATCCACTCC	TTTCCGAATGCACTGAAACT	225
>SoSSR1012	CCTTTGAAGCACTCCACATT	TCTCTCCCTGTAGCATTCCA	185
>SoSSR1013	GACTCGATGCTCCGTTATCA	CCGAACTCCTCCTTCCTAAT	186
>SoSSR1014	GTCGTTCCCTTTCGATTGCTA	ACGAAGGTGCTCAGAACAAG	213
>SoSSR1015	TCCACTTGCAATTCATTCTT	CCCACCTCATACTCTTGCAT	235
>SoSSR1016	CTCTCCTAATTTCTTGGAGCTG	CGAAGATGGAGTAGCCTCTG	246
>SoSSR1017	AATCGGAGAAGGAGGATCTG	TCGACACTTTACTCCCATC	246
>SoSSR1018	GGAGAGAGAGTGTGTTGGA	CCCGAACTTCTCTTCATCT	118
>SoSSR1019	ATGGCTCTGGAGGAAGAGTT	TCTCCCTCATGGACATCATT	143
>SoSSR1020	AGGAGTGCAGACAGGAACAG	GATGCAAGCTTTTGTGGTGT	149
>SoSSR1021	ACACCACAAAAGCTTGCATC	AACGCATAGGAGGAGGAGAC	167
>SoSSR1022	CAGAAATTGGAGAAGGCTGA	ATCATTGTTCGGATTCTGTGC	244
>SoSSR1023	GCTCAAAGCAACCAAACACT	TATTCCAGTAGGCACCTTCG	200
>SoSSR1024	GGAATTTGAAAGGGGAGAAA	TCGATATTCTCCTCGTCTGG	240
>SoSSR1025	AGAAACACACCCATCTCTCG	GGGGTAAAAGGGTCATTCTG	133
>SoSSR1026	ATCCAGTGTCTCAGCTTCTG	CTGGTCGTTATTCCATGAGG	161
>SoSSR1027	CGATCTCATCACCAAAGTCC	GTACACCCACTGCATGTCTT	118
>SoSSR1028	AGCCGAGAATCGACCATC	ATGAGGTGCGGAAAATTGA	227
>SoSSR1029	CACATCGATTCCGGTAAAAC	AGCGGCGGATTGTATAGTAA	159
>SoSSR1030	AAACCCATGCTAGAACAACA	TTCGGAATTTCTTGAACCTG	189
>SoSSR1031	TCTCTCCTCCTCATCCTCCT	ATGTCGGTGAAATCGGTAGA	144
>SoSSR1032	GATGCTTGAGGAGTTGCATT	GCCTATCGATTTCTGTCCAA	245
>SoSSR1033	TCCTGTTCCCTGCTCAACTTC	ACTTGATCTGAACCGCACAT	138
>SoSSR1034	AGAAGCACAGATGCAAGCTC	CTCAACACGTAGCTCGACAG	241
>SoSSR1035	GCACTGCCTCATAATGCTCT	GGCACTGGTGTGCTTCTTC	248
>SoSSR1036	ATCTGATCTCGAAGCTGGTG	GAGCTCCCTCTCCTCATCTC	187
>SoSSR1037	ATCCTCCTCAACCCTGTCTT	GTCCTCCATCGCAAATTTCT	247
>SoSSR1038	CAAGGTCAAGGAGCTGTGTG	CGTCATCTCCGTTAGAGGAA	128
>SoSSR1039	ATATCGGCCGGCTATACTTC	CGCTCTCAATTTTGTCCAGT	145
>SoSSR1040	TCTTCCTGGAGCCTCTATCC	ACCTAGAGTTGGGAGGCAAA	187
>SoSSR1041	CTTGGATGAGTTCCTGATGG	CACTCCTTCTCACTGCACCT	204
>SoSSR1042	AGGGATTTTCTGTGTGGTGA	ACCACATTCACGTTTCCTGT	234
>SoSSR1043	CGAAACTGATTGGGGTACAG	CATGAGCTAAGGCCCATAAA	165
>SoSSR1044	CGATTATCTGCATTGCTCCT	AGCTCTCGTTTTTCTCGATT	245
>SoSSR1045	CCAATCTCGGTCCCTCCTATT	GCGAGGAGAAATTAGGGTTT	249
>SoSSR1046	CTTGAATCCGAAGTCGAAGA	AAACCAAGGGTCTCTTTTCC	116
>SoSSR1047	GGATTCTCTGCAAATGCT	CATCTCCACCTTGCTCAAAA	248
>SoSSR1048	CAACTCCTCCGAATCCAGT	CTTGGGCCTCTTGTCTGTGA	174
>SoSSR1049	TAGGTCTGCTGGAGGATTTG	AGTCAACAAGATCGCCATTC	117

>SoSSR1050	CACTTCCAGCAGCTTTTCAT	CTTGCGCCTTTTCATCAC	223
>SoSSR1051	AGTCTGAATTCCACCAATGC	CCGACCTCGTCTTGATACAC	154
>SoSSR1052	TGATGAGGCTTTCCCTAAAA	AGGTTTGGAGGAGGGAGAG	150
>SoSSR1053	GTCTCTTCTCCACCATCCT	GTGAAGACGAATTTGGGGTA	212
>SoSSR1054	CGAGCACTTCCACATTCTTT	ATGGCTTCCTTGTCTATCC	228
>SoSSR1055	CGATCACTGTGCCTGAAATA	TCTCAGCCATTCTCTCCATC	183
>SoSSR1056	GAAAATCGTCACCACAAAGC	TTCTTCGCCTTCTGTTGTTT	246
>SoSSR1057	CGTCTTCCAAATCCTCTCAA	AAGGTGTTGTTCGGAGTGAAA	204
>SoSSR1058	GATGGTGCAGAAGGATCAAG	TCTGCAAAAGCCAATAATCC	167
>SoSSR1059	GGATGTGGATGATGAGGAAG	TCTCAAGTTTGGCTCTACC	164
>SoSSR1060	CTGCGGTTGATAGAGAAGGA	ACCAGAACTCCATCAGTGC	221
>SoSSR1061	AGCCATAAACTGCCCAAGAT	CTTGGAGTGTGGTGTGATGA	232
>SoSSR1062	AACCATTTAAGGCATGGTCA	TTGCTCCAAAACAAGTCTCC	170
>SoSSR1063	ACCCTCTCAACCTCCATCTC	GCTCAAAGTGGTGTCTTTGT	150
>SoSSR1064	GGAACAGCGTTAGCATTAGC	TGATGATATGGGGTTGTGTG	239
>SoSSR1065	GCCGATCTCGACTTACAAAA	TAGCTCGTTATGGACCTTCG	237
>SoSSR1066	GTTGGGCCACAAGAAGAAGT	GTCTTGAGGTCACCCTTTC	226
>SoSSR1067	GTAGCAGCGCATCAACATC	TTGGGTGTTAGTGGCAAACT	146
>SoSSR1068	AGATGGTGCCAATGGTTTTA	GGCTGTACCGTCTATTACG	193
>SoSSR1069	GCCGTTATTTGGTAATGGTG	TGGTGACGAATCGGAACTAT	199
>SoSSR1070	ACTTCGCCATGAGATTGAAG	GCTCATAGCTGATGGTTGCT	206
>SoSSR1071	GAGAAGACAAGGGACGTGAA	GGAGTTTGCTTTCCACACAT	156
>SoSSR1072	CGACAACAAGGATCGTCACT	TCCTCGAGAAGAGCAGATTG	219
>SoSSR1073	CCTGTCGATCCTAGGGTTCT	CCTTCACCGAGATGTTGTTC	208
>SoSSR1074	TTCTTCGAGCAGATTTTAGG	ATGATTGAACCTCCTTTCC	228
>SoSSR1075	TCTTCTCCTGCTGCTACCAC	CACAGGCAAGCTCTGATTCT	149
>SoSSR1076	GCTCTCTCTTTCAGCTCCA	TCAGCTCCAACCTTCAACCTC	195
>SoSSR1077	TCTTCAGCAAGTTCGACACC	TACAGATCAAACGCCTCCTT	229
>SoSSR1078	AACGGCATTGCACATCTC	TGAAGCTGAGGAGCTTGATT	120
>SoSSR1079	TCCTCTTCATCTTCGTCAGG	CCTCGCTTTCATAACCAGAA	202
>SoSSR1080	TTCGGTAAAGCAACGGATAG	GGAGAGGAAGGAGAGATTGG	207
>SoSSR1081	TCTTGCCTTGATCTCGAAAC	CCTCTTCATCAGACGTGAGC	145
>SoSSR1082	TGATGGAGAAAATGGTGGTT	GAGCACCAAGAAGATGCAGT	105
>SoSSR1083	AAATGTTTCGAGGTGGTAAA	TGTGTGCCTCCCTCTTTTAG	112
>SoSSR1084	TGAGGCTTCCACCTCCAC	GGAGAGGGAGAAGGTGTTGT	155
>SoSSR1085	ACCCCTCTTCCTCACACTCT	GAGGTTAGGGCCCAGTATTT	204
>SoSSR1086	TACGATTACTCGCCGTCTTC	TTTCATCACCGCATACTCCT	225
>SoSSR1087	GAAGAAGCTCGAGAACGAGA	ACCGAACATAAAATCGTCCA	199
>SoSSR1088	TCCCTGCAAATAGAAACAGG	GTTGGAATGGCATCAATAGC	166
>SoSSR1089	TGAAGTCATGGATGTTGTGC	AATGGCTTGAATTGAGCTTG	174
>SoSSR1090	CTACTTCCGCTCGCTCAA	GAGATGCAAATGCGATTCA	217
>SoSSR1091	TCACCATCACCATCATCATC	TGTTGGAATTTGGGAGGTAG	249
>SoSSR1092	AACTGAAGAACTGGCGATG	CCACCTCATCATCTTGGAAC	231
>SoSSR1093	CGAAAAGAGGTTCTGATGGA	ATTAGGGTGAGTTGGCTGGA	237
>SoSSR1094	AGCAACTGGAGTGTGTCCTC	TTTTGAGGCGAGCTTCTCTA	187
>SoSSR1095	AAAGGCAAGTGCACAACAAG	ATGTTTGTGTGGGTGGAGTT	179
>SoSSR1096	CAGTGGGAGGAATCGAAAG	ATCCGCATAGTGACGAGCTA	188
>SoSSR1097	CTAGCGCCCAGACGTAAAT	GCAGTGGTGAATTTGTGTGA	242

>SoSSR1098	GATAGCTCTGCAATCCCTCA	CAGAACACTCAGCAGTCACG	220
>SoSSR1099	CTCAAAGCCGGAAACAGAT	CTCCGTTTTTCAGAGACTCCA	225
>SoSSR1100	TGTTGCTGGTGGTGTAGTTG	AATCCTCCCTCAAAAGCCTA	176
>SoSSR1101	TAGGCTTTTGAGGGAGGATT	TCCCACATGGACCGTATATT	204
>SoSSR1102	GCTGTCCTGCAACAGCTACT	GTGATGATCCCCATTGACAT	168
>SoSSR1103	TCAGATCACGCAGAAGATTG	CTAGTGTCCTAACGCTCCA	155
>SoSSR1104	CCCTTACAACAACGGTATGC	TCCGATGTAATTGCTTTCGT	234
>SoSSR1105	CCCATGAAAAGGAGAAGCTA	TAGCCCGAGTTGGATATCAG	113
>SoSSR1106	GGATCGGGAAGATACTACA	GCCTGTTTTGCGATATTCAG	249
>SoSSR1107	CCACCAGAAATTACGCTGAT	TCCGGGATGAAGGAGTTT	214
>SoSSR1108	AGTCTCTCAACATGGGATGG	CGAGGTAAATTTGGAAGCA	218
>SoSSR1109	GATTCTGCGAGTGTGGTTTT	TCACCCCTTCTTCTCTCTCT	148
>SoSSR1110	CAGGAGCAGAGAGATGAGGA	TCCGGCTTCTTTAGACCTTT	167
>SoSSR1111	CTAGCTTATGGTCGGCTTTG	TGGTGGTGGAGGAGAACTTA	185
>SoSSR1112	CTAGCTTATGGTCGGCTTTG	AAGTATGGAATGGAGGTGGAG	219
>SoSSR1113	AGAGCAATGAGCAGAGGAGA	CCGTTTTCTGGTAGTGATCG	147
>SoSSR1114	ACCAATCGGAAGAAGAGGAA	CACCTCACTCTCACCTTGT	204
>SoSSR1115	GCAACAGCAGCAACATCTAA	AAATCCTCGGAAAGCAGAAG	215
>SoSSR1116	ATATTCAGGTTGGGCAGGAT	CTGATATGCAACAGCACCAG	173
>SoSSR1117	TTCCTGGGCCATTAATAATCT	CGTAGCATTGGATCAGGTTC	247
>SoSSR1118	TCACCTCCTGCATTCTCTTC	AGGGTCAGGGCAGTGTAGTT	214
>SoSSR1119	TCATCATCTTTGCACTTCCA	AGGGAGGAGTAGTTGGTGCT	179
>SoSSR1120	AGGCAGCAGGCTACATACAT	TTTGGAGGGACTGGACATAA	195
>SoSSR1121	CACTTTACACAACCGCCTCT	GATGACTTCGAGATCGTTGG	232
>SoSSR1122	GCTCCATTCTCTTCACTCTCC	TTAGTGTCGGTGGAAATTGGT	217
>SoSSR1123	TGTGTCGATGAATGTTCCAG	CTTGGTGAAGCCAGAACCTA	241
>SoSSR1124	CCTGAGAGAGATGCTGATGG	CCTGCAGTTTCTTGAGCTGT	241
>SoSSR1125	CGACTCTAGGTCTCCGATCA	CTTTGACTTTACCGGTGGTG	101
>SoSSR1126	CATTCCTCAGCTCAACAAA	CATGCAGTCCGTACAGCTC	203
>SoSSR1127	AGAAAGATGGTGTATGGCAAG	TCCAGTAGTTCCCCCTTTTCC	215
>SoSSR1128	TATGAAGATGACGACGGTGA	CGATGATCCTCCCAGTACAC	184
>SoSSR1129	CCTTCTCACACCTCTCCAAA	CTGGGGATGTTCTTTTCTT	229
>SoSSR1130	GAGGAAACGGAAACCCTAGTT	TGCACGGTATCAGCTTCTTA	235
>SoSSR1131	TCCATCGATTTGACTTCTG	CTTCTGCTTCTCCTCCGATT	216
>SoSSR1132	TGGAGATTGTGGAGGTTGTT	ACCCTTCACTCTCACACTGC	120
>SoSSR1133	GCAAAACCAAGGAATCTTGTGA	GGCTAAGGTGTTGAGGGATT	242
>SoSSR1134	ACCTCTCCTGCGTTTCCTT	GTCGCAGCTCAGGACTAAAC	209
>SoSSR1135	GAAAAGAGGCTCAACTTCATCA	GACTTGCTCTGCAACATCCT	235
>SoSSR1136	TGACCCATAACTGTTTCTCCA	GGTTTTGAGGGTCTGAAGGT	161
>SoSSR1137	TGATAATGTCTTCTCCACCA	TGCTCCTGCTTGCTTCTTTT	187
>SoSSR1138	AGAGCAGATGATTCGAGACG	TTTCACTGGGCCTTCTATCA	229
>SoSSR1139	CGCACAAGATGATCAGACAG	CTCATCTAACAAATCAGATCCCAGA	110
>SoSSR1140	CCTCTTGGCGCTAATACTGA	GTGTCCACCACCATCTTCTG	228
>SoSSR1141	CCCACCACATCGTAAATCTC	AAACATTTGATGGGAGCAGA	164
>SoSSR1142	ATGTGGGCTTGAATCATTGT	CCATCCTCTTGAACACAACC	184
>SoSSR1143	CTATGAAAGCTCGTCCCTCA	CTTCTTTGCCCTTGTTCCTCA	178
>SoSSR1144	CAGCGAGTTGGTGGTTTATT	TCCATCTATGCCTGGTGACT	194
>SoSSR1145	CCGATGACACCGTTTTCTAC	AGGATGTAAACGACGAACCA	247

>SoSSR1146	GCTCAATTGGGAGATGAGG	CAATACCATCGCCATTCTTC	181
>SoSSR1147	GGGCTTCTTCATCTCCATCT	ATCTTGCCTTTCAGGCCTAT	157
>SoSSR1148	AGTCAGGAGCAACAAATGGA	CAAGCTCAATCCCCAAGTTA	237
>SoSSR1149	GCAAGCACTTCAGCCTTTTAA	AGAACCCAGGTGAAATAGGG	155
>SoSSR1150	CCCTATTTACCTGGGTTCT	GAAAGCTCGAGTAGGGGTTCT	188
>SoSSR1151	GTTAAGCGTCTGGATGAAGC	ACCCAATGACCCTCTGAAGT	225
>SoSSR1152	CAATGTGAGCAACAAAACCA	GGTCCTGCTAGGAGAAAAGG	247
>SoSSR1153	AATTTTCAGCATCACCTCCA	GCTCTTGGGCTAGGACTTTT	226
>SoSSR1154	GAAGCGGATGAGATGTCTA	CCCTCAGGAACCTGTCTGT	137
>SoSSR1155	GCGACGAGATGTCTAATCAAG	CCTCCCTGATCTTCTTCTCC	213
>SoSSR1156	GAGCGCGAGCTAAACAGAT	TCACGAGAGAAGAAAATGGAG	163
>SoSSR1157	AGCGTCTGACTCCATTTTCTT	TCGCGTCACTTATGTTGCTA	205
>SoSSR1158	TGGAAGTGCATTATCTGCTG	TGGCTTTTCTGGCTTACATC	228
>SoSSR1159	ATAACCAGAAATCCCCAAACC	TCTTTCCGTGAAAATGGAGA	210
>SoSSR1160	CCTGGGTTTAGGATTTGGAT	TGGAGCAAACATCAACCTTT	125
>SoSSR1161	TCTCACTCTGACCCTCTTCG	CTACCCCTCCACGAAACTCT	156
>SoSSR1162	GTTTCGGTTGCAGAAGAAGA	AGAGCGACTGCGCTTATG	124
>SoSSR1163	CTGCGGATTGATAAGGAGAA	GGAGACTTGCCTTGAGTTGA	241
>SoSSR1164	CGGAGTTCGACAACGTCAT	GAGATGCAATTCACCCTGAT	241
>SoSSR1165	ATCATTGTGAGTTCGCCATT	CTGATTGAAGGATGGATTCTG	159
>SoSSR1166	CCAGCAGTGTCTACCTAA	CCCACGTTTGCTATGAAGTC	234
>SoSSR1167	GGTTAGGGTTGATGATCTCG	CGCTTTACAGTATCCCTCCA	148
>SoSSR1168	GTGGATGGAGTATGGAATGC	CTAACCTCATCCTCCCCAAT	214
>SoSSR1169	GTGCCATCAATGCCTAAAAG	CTTCTTGACACCACCTTGCT	152
>SoSSR1170	TGAAATCTAGGCCGTCAAAG	CTTCTTGACACCACCTTGCT	226
>SoSSR1171	CTTGCTTCTTCAGGCCAAA	ATCCCCGACTACGAGAAACT	183
>SoSSR1172	CTCTCTCCTCATTCCCATCC	GTCAAGTCGCAGATGAACCT	105
>SoSSR1173	TCTTCAACAGACCCAACAGC	GGAGAAGAGGGATTGATGG	231
>SoSSR1174	AATAACGCCGAAATAGTGGA	GTGGCATTGTTTTGCTTCTT	200
>SoSSR1175	CACCCCAATTACAGCTCAA	CTCTTCCCTCCATTTTCTCC	233
>SoSSR1176	AGATGGTTGTTTGTCTCCA	AACTGCCGCATCACTACTTC	205
>SoSSR1177	TAAC TTCGGAATCGGACTTG	CAATCTTTTCCCTCCAAAGCA	246
>SoSSR1178	TAAAAGGAATGGCGATCAAC	CAGCACAACTTTGACAGCAC	244
>SoSSR1179	GCATCAAGAACTCGAACTGG	GTAGTAGTTGTGCGCGTCTGT	242
>SoSSR1180	CACGCCCTCTCATTGTC	GGTAGTGGTTGAACGTCGAG	124
>SoSSR1181	CTGTGGAAGGTTGGAAAATG	CAATCGACGCAGATCTTGTT	224
>SoSSR1182	GGATCCCATCTCATAGCAAA	CTTGCCCTTGAAGAAGAGGT	177
>SoSSR1183	TTTCTACCACGACCACGAAT	AACGATTTCTTGCAATGAGC	175
>SoSSR1184	ACGAATCGGAGACCACAATA	ACATTCTGAGACGTGCTTCC	232
>SoSSR1185	GTTACAGGGAATCGGGAAGT	TTTAGGGTTTCGCTTGTTTG	241
>SoSSR1186	TCATTCATCCTCACGTCTCC	CAGGTTGTTGTGCCTGAAG	234
>SoSSR1187	CCAATCGCTACGTCTTCG	TAGCCGGGGTACTGGTAAT	118
>SoSSR1188	CGCCGTATTACCAGTACCC	ATCCTGAGCGTCTCCTTCTT	232
>SoSSR1189	CGCGATGATGATGAGTATGA	CAAAGCCTATTTGCTTCCAA	190
>SoSSR1190	CAGAGCGCCTCCTATTACCT	TTCTTGAGGAAGGGTGGAAAT	235
>SoSSR1191	GCTCAGTTAGCCATTTGGAA	TCCTCAGCCTTCTTGTCTATC	217
>SoSSR1192	GATCACAGCCATAGCCATTT	GCCTATGAGGATGATGATGG	239
>SoSSR1193	ATTGGAATCGAGCTTTTGTG	GCCTATGAGGATGATGATGG	151

>SoSSR1194	AGAGGGGCGAAGTTCCTTCTA	TCACCTTCACCTTCACCTTC	209
>SoSSR1195	ATCCACCGCCACATTCTC	GTCGGAGGGCTCTAAGTAGG	202
>SoSSR1196	CGGAGCTTTAAGAAGGTGTG	CTTCTTCACGTCGATTTCTGT	195
>SoSSR1197	TCCTGCATCCCTATACGAAG	ACCCTCATCTCCTTGAATCC	248
>SoSSR1198	AGCTGTTTCCATTTCCCTTC	GTGGGGAGAGAGTGATTGTG	200
>SoSSR1199	GCTTTTGTGGACTTGCTCAT	GAATCTGCTCCTGAGTGGAA	172
>SoSSR1200	AGATTCCGTCGCTTCTCG	GAGAAATCCGACTCATCCAA	124
>SoSSR1201	GAGAATCGATGCAGAAAACC	GTCTTGAGCGTCTTGTCGAT	199
>SoSSR1202	TTTCCAGACGAGTTGAGAGG	GCACAACACTTTGAGGGATT	149
>SoSSR1203	CGATACGCCTCCTTCTCTCT	GAGTCTGAAGCTGGAGTGGA	230
>SoSSR1204	GCCTCTCGATGAAGAAAACA	CAATCGTCTGCCAGTACTCC	191
>SoSSR1205	TTGAATTCCTTGGAGAGTCG	GCAAGTGGTTTGCCAGTATT	214
>SoSSR1206	CTTGTCTTCCCTCACTGCAT	TGGAATAGCCTTTGTTTTGG	164
>SoSSR1207	TCCCGTGAGAATAACAATGAG	TCATCAATCCCATCTTCAGC	215
>SoSSR1208	CAAGTTTTGAGAGGGTGTGG	CATGTCCCTTTGAATCCATC	167
>SoSSR1209	ATGGGGAGAATGAAGGTGAT	TTCTTCTTTGCCTTCTCCAG	193
>SoSSR1210	GAAGTTCGTC AAGGAGTGGA	TCATACGGGGCTGAGGAG	142
>SoSSR1211	TTTACCAACAAGCCAAGTCC	GTTGCTGATGGAGCTTCTGT	239
>SoSSR1212	GATCGAAGTTCTGCAAAATCC	TTCGTCTGTCGATTAGGGTTA	163
>SoSSR1213	ACCCTGCTGGATCAGTAACA	CCACTTCTGCACTCTCCACT	237
>SoSSR1214	CTTAAC TTCCCGACCAAAT	ATAGTCAAGGCCCGATTCTT	213
>SoSSR1215	CAGCAAGATATTCGGCTTCT	CACTCTCGGAATTTAATGG	250
>SoSSR1216	AAGACGACAAATTCGTTCCA	TCCTTCCACTGCTTCTCTTC	188
>SoSSR1217	GCCTCACTTACGTTTCGAGT	CTCCCAAGGCTCTTTCTAC	209
>SoSSR1218	TTTGGAGAGGGTGAAGAGTG	CCAACAATTGGAGAAACGTC	245
>SoSSR1219	GCCCGAAGAACTGAGTGTA	ACCACCATCATAACCATTGC	189
>SoSSR1220	TTCTGACGAGGAGGACTCTG	ATGTGGAGGATTGTTGGAAA	175
>SoSSR1221	AATGAAACACCTCGCCTATG	CTTCAAATGGGGAAACTCCT	247
>SoSSR1222	GTGGTTGAAACTGGAAATGC	GTA CTGGT C GAACCAGATG	229
>SoSSR1223	CCGCTTCTCTAACCCCTCAGT	ACGGTCGAGATCATTCTTCA	197
>SoSSR1224	GTGGAGGAGCACAACAAGAA	TTTTAGCCAGTACGGAGCAG	219
>SoSSR1225	TGTTTTGCGGTGTACCAGAAT	ATCCAAGTCCAAAATCAGCA	171
>SoSSR1226	GTGGCGATTTGTTCAACTTC	TGTCACGCTTCTTGTGTTTC	240
>SoSSR1227	CTTTACTGCGACGGAAATGT	CCAGACGTAGTCGAAAGACG	103
>SoSSR1228	CCACACCTCAATCCTCGTC	GATCCTTTTGTCTTCGCTGA	219
>SoSSR1229	AAACGACGACGTACAAGGAG	TTAGCCTGCTTTTGAATTC	240
>SoSSR1230	CGTCGATCACTGGTACAACA	GCAGTAGTGC GTGACTTCG	184
>SoSSR1231	GAAGAAGCGGAGGAAGAGAG	CTCTCGAACATGTCCACCAT	150
>SoSSR1232	TCTCTGAAAACGCTCCTCAC	GCTTGAGCCGAGTTGAAATA	139
>SoSSR1233	GTCTCCGACTTCCCTCTCTT	AAAAC TTCTTCGCTCCAAT	204
>SoSSR1234	GAACAAGAAGGGGAGGTTGT	TGTATCGTCACGTTACCAG	195
>SoSSR1235	TATGTAAATTCGCGCTGGTT	AAATCGGATATTCCTCACC	156
>SoSSR1236	GAATGGGAAGAGAGTGACGA	ATCAAGCCGTGGTAGAGGAT	198
>SoSSR1237	GATGATGATCCCGAGGAAG	GCATCACTCCATTGCTCCTA	197
>SoSSR1238	CTAGGGATTTGGAGGAGGTG	CCTGAGATCCCTGCCTTTA	182
>SoSSR1239	ACCCGATTCTTACACAGCAG	CGAATAAAGGCAGACTCCAA	217
>SoSSR1240	CATCGACAGTGGAGAAAACC	CCTAACCCAATTTTCATCAGC	216
>SoSSR1241	GATCTTCGTCGCCTCATC	CCAGGACCATCCTCTTTCTT	179

>SoSSR1242	GTCGCGGAAATAGTTGAGAA	GCACTTCAACAGGATTGGTT	250
>SoSSR1243	TCAATCTCTTCTGCAAAGG	CATCTTGTCCTCCACATTC	187
>SoSSR1244	AAACGCCGACCAGAATAAG	GATCAGGAGGAAGAGGAAGC	214
>SoSSR1245	CCGATGTTTGATTCTTTTGG	GGAAGGTCCACCTAACTGGT	231
>SoSSR1246	CCGAGATTCAGACGATGAAG	TCCCTCTCTCTATCCCTTGG	135
>SoSSR1247	AGAGGCCAAGGGATAGAGAG	TTGCCTATCACCATTCTCGT	224
>SoSSR1248	GATAGAAGGGATCGGGAGAG	TCGACAGCAACATTCACATT	203
>SoSSR1249	CAAGATCAAAATCCCGAAGA	TGAAGCTCCTATTGCATTCC	101
>SoSSR1250	AAGCCACAACAAGTTC AAGC	TCTTCTTCCCTCCACCTC	207
>SoSSR1251	AGAGAGCATTGCATCGTTTC	CGGAGAGCTGAGACCTTCTT	113
>SoSSR1252	AGGCTGACAAACAACAGGAG	CAATGCAACAAATTGAGACG	243
>SoSSR1253	TTCTTCATTGCAATTCCTC	TCCTCCTCAGATGACGTTTC	222
>SoSSR1254	TTCTACTGCTCCCTTGTTTCG	TGAGCATCGAAATTCTCCTC	120
>SoSSR1255	GTGGAGAAGTGAGCTTGAGG	TCCTGTCCTTTGGAGATTCA	201
>SoSSR1256	GGCCTCTCGCAATATCTCTC	GGCTGTTTACGTGGATCAGT	143
>SoSSR1257	CCTTTCCTTGCCATATGTTG	CCATTCCTGTTGGAACAATC	189
>SoSSR1258	ATGAACATTTGGCCGTTCT	TCTGCAGGGAGGAATAGTTG	244
>SoSSR1259	AGCAGAGGAGCAACCCTAAT	TCTTCTCACTTTCCCTTC	162
>SoSSR1260	TTCCGAGGAGACTGAGAAGA	TCCGCTAAGTCAAGCAACTC	246
>SoSSR1261	CTGCCCTCCAAATGAGAGAAA	TCATGTCCAAACACATGAGC	230
>SoSSR1262	CCAAAGTTTCTCTGCCTTTG	ATTCTCCTCCTCCTCCTCCT	184
>SoSSR1263	CAAAATGAAGGAGAGGCTGA	TACGGCTAATCACATCACGA	136
>SoSSR1264	CGTCACCACCATCACATCTA	CGAGGGCTCTACCTCTACCT	220
>SoSSR1265	CATCTTCTCCGACTCTCCAA	GGCACTAATTTCTCTGCAA	172
>SoSSR1266	GAAGCCTTGCTTGTTCTACG	GTTAGCTGGCAGTGGAGAGA	235
>SoSSR1267	CTGCCAGCTAACAATGCTT	GTGCTATCAGCAGGGAATGT	182
>SoSSR1268	CCTTCTCTCCACCTTCTTC	GCTGTATGAGCCTCGTGATT	151
>SoSSR1269	GAGCAATCGTCTTTCAGCAT	CCGACAGACCAATCAACTTT	183
>SoSSR1270	AAGTGC GGACAAGAAACAAG	CTGTCTCTAGCAACGGCTTC	242
>SoSSR1271	TATCCCAAGCGAACATTCTG	TGATCGGTCAAATCACCTT	226
>SoSSR1272	CAGTCACATTCCCTCCAAAA	CTGCCACTGCTCCTTGTACT	250
>SoSSR1273	CGAAGAAATCGAACAAGCTG	TTCCTCTCCGGAATATCCTC	198
>SoSSR1274	GGCGAAATTAACGAGGAAAT	TAGATCTGGGTGCTTTCCAA	204
>SoSSR1275	ATGTATTCCACCACCAGCAT	GCAATCCGATAGGTTTAGCA	195
>SoSSR1276	GTCTCCTCCGCTACAGCTACT	AGATCGCAGTCACAATCACA	184
>SoSSR1277	TGTTGGATTGCATCACTTGT	TGTGATATCGCGAACTTTCC	244
>SoSSR1278	GTCTCCATCACCAGGGAGTA	CATGAGGTGTTTGGAGGACT	168
>SoSSR1279	CCCTGTTCTGTTT CAGAATG	TTTAAATCCTCCCCTTGACC	108
>SoSSR1280	GGAGTCACACTCACCACCTC	GGTAGAAGCAGACGGAGACA	110
>SoSSR1281	ACCACGCTCTCCTCCCTAC	GTCGTCAGCGCAGTTATCC	183
>SoSSR1282	CCTCTTCAGAACCTCATTTCG	CATGCCTTCCGTCTCTGTAG	236
>SoSSR1283	TCATTCGTCTCAATGAAGCA	CCAAGTTGAGAAATGGATCG	219
>SoSSR1284	AACCAGTTCATCTCCAACA	TGAGACGTATCCTCCAAGT	129
>SoSSR1285	AGCCTCTCTCACTTCCCATT	GGGAATACGGATTTCTGGAC	105
>SoSSR1286	GGAAAAGAAAGGGTCGAAAA	TGTAGGTGTGGCCTTTGAAT	171
>SoSSR1287	CCAGTCCCAACTGAACAAAG	CTCTCCCAAGAAAACCATT	152
>SoSSR1288	TGTGAAAGAGGAGGTCAAGG	AGGCCATCAGAAACAGTCAG	159
>SoSSR1289	GAGCGTGAAAGGAAACAAAA	TCGAGCTTCAGCACTTACCT	228

>SoSSR1290	TCTCCAAGCTTTCAAAGGA	TTCTAAAAGGATGGGCCTCT	243
>SoSSR1291	GAAGCAAACCCAGTTGAGAA	CACCTCTTGCTCTCGCTTTA	250
>SoSSR1292	ACCACACTCACGAATCCAGT	TTTACCGGAGAGCAGTCGTA	101
>SoSSR1293	AGAATCCGAGATCCCTCAAC	ACGATCTCTCCCTCTCGTCT	201
>SoSSR1294	GAAGCAACCTGATCTTTGGA	CACAGCAGAAGCATCATCAG	208
>SoSSR1295	CTACTGGAGGACCCACATGA	GGAGATGTGAGATTGCTGCT	240
>SoSSR1296	GACCAGATCGTGAGGTTCC	CTCACCGTCTTCTTCAGCTC	194
>SoSSR1297	TACGGCAACTACGACAACAA	TTTCGAGAGCCGATATGAAC	196
>SoSSR1298	ACCCACCAGAAATCCATGT	GAGACCCTTTCTTGGAGCTT	245
>SoSSR1299	GAGAACATTGAGGGCAAGAA	GAAAAGCTCCAACAAACCA	172
>SoSSR1300	ACTCAAAACACACACGCAA	AGTTGACGGACAGTGAGAGG	159
>SoSSR1301	CGCTCATGCTGTAACCTCTT	GTCGAGTCTGTGCTTCACCT	173
>SoSSR1302	CCTAGATTCCGTGAACGATG	TTTTCTAGGATGGCAACAGC	180
>SoSSR1303	AATGGCGATAAGTGCTTCAG	GAGGTAGGAGGAGGATGAGC	229
>SoSSR1304	GACTGGGATGGTTATGGTGA	TTATGTGCTGAACGTTGGTG	236
>SoSSR1305	AGGGCTAGGATGGAGAGAGA	TTCATGACAGGTTGACCAGA	128
>SoSSR1306	CAGCAGCAGCAACTCTGATA	CCAAAATTCCCCTTCATCTT	181
>SoSSR1307	GAAGGTGCCAATAGCAAAGA	ATGACGTTGGAGCAACATTT	241
>SoSSR1308	GCGAGGAGAAGAAAATGGAT	AACACTTGCTGACGTGGAG	250
>SoSSR1309	ATTGGTCGCTGATGTGAGC	TCTCACCTCCAAATTCATCG	243
>SoSSR1310	TTCAAGAGTCATCAGCGACA	AAGAGGGTGGAGAGAGAGGA	225
>SoSSR1311	ATCGCCAAGCTACGTACAAC	GGATCAACTTCGTCATCAGC	196
>SoSSR1312	GAAAATTTCGATGGGAAAGT	GGCAACACTACTTCGGATTG	132
>SoSSR1313	CACACCAACATTAGGCATCA	GCTGCTGCTCTTCATCTTCT	103
>SoSSR1314	AAACTGAGGAGGAGGAGGAA	CAGCAATTTGGAGAACGATT	230
>SoSSR1315	GAGAAGCGGAAGAAGAACAA	GCATCGACTCCACGATCT	154
>SoSSR1316	CGCTCTCTCTCTCCTCATCTT	ACATTGTGCAGGCAATCTG	119
>SoSSR1317	CCTCCCTCTGTGCTCTCTCT	GGAGGTGGCGAGAAGTAGTAG	232
>SoSSR1318	AGGAGCCGTATAAGGACGAC	ATAGCCATAAATCCCCTTGG	194
>SoSSR1319	GCCGAGTAAAGGAGGGTATG	TGTTGTCTTGGTGTGGTG	171
>SoSSR1320	AAAATGTGGAAGCCTTTGG	CCATGTACAGATCCCCGTAG	228
>SoSSR1321	TGAATGAGAGAGGGTGGATG	CGAGGAATTGAAGAGGGTTT	241
>SoSSR1322	TGGGTTATCTCTGGTGGCTA	TCGGCATCGTCATCATTATT	246
>SoSSR1323	AACAGCGACAGCGAAGATAC	GTAGGATGGGAATCGAGAG	198
>SoSSR1324	GGCTCCATTTACTTCCTGGT	TGCTGCTGCTTCTTCTTCTT	113
>SoSSR1325	GGCTCCATTTACTTCCTGGT	AAAATCCTCGGAAATGCTGT	186
>SoSSR1326	TGAAAACAATCCCCTGAATG	GAACCCATGTGAGACAGACC	189
>SoSSR1327	AAGAAGCGAAAGACGGTGTA	CGTCTTCTTCTCCTCGTCGT	219
>SoSSR1328	AATTTCCGCTCCAGTCTTTT	CTGCAAATCCTCGTCACTCT	147
>SoSSR1329	CTTGGCGACTTTGATGAACT	CACGCCCTTTCACGATCC	156
>SoSSR1330	CGTCTGGAAGAACAAAGTCG	CATTTCTGCTTGAAGGTGCT	178
>SoSSR1331	GTCTTCCTCCCAACTCCAAC	GGGGAGATCGATACAGAAGG	133
>SoSSR1332	GGGTGTTGGGGTTCTTGTA	AGGTCACTAGGGTGATGTGC	150
>SoSSR1333	TCGACTCCCAAGACCTTCTA	CCTCTTCTCTCCCCTTCTT	250
>SoSSR1334	TCTCTTCCCTGGGTTTGTGA	TCGAGTTGCTCATCCTTTTC	243
>SoSSR1335	TGAAGATGTGCTAATGTGCG	CCTTATGCCTAACCATGCAA	138
>SoSSR1336	GCGCCACAATCATAAGAGAT	CTGCTCGAGTCTTCTCCTCA	214
>SoSSR1337	TCTGATCTTCCCTCTGTTACG	TTCGATGGTGAAGGAATTGT	148

>SoSSR1338	TTTACCGAACCATTCTCCAA	CCAAATCTTTCGCTTCTTCA	159
>SoSSR1339	GTGGGTGTTGCAGATTTCTC	GAGATGGAAGTGCAGGAAGA	233
>SoSSR1340	TTGAGAGGAATTCTGGTGTG	CTCATTGGGATCAGGTTTTG	194
>SoSSR1341	TACTGGGTATGATGGGCAAC	AGTCCCTTGGTCTTGGACTT	212
>SoSSR1342	GTTGGAGGATCAGGATCAAG	TTTGACTTTTCCGAAGTTGC	223
>SoSSR1343	TGGAGGATCAGGATCAAGAG	TTTGACTTTTCCGAAGTTGC	221
>SoSSR1344	GTCGAGCAACTGGAAGGAT	TCGATCCTCAAGGACATTTT	206
>SoSSR1345	CCGACTTGCTCATCTACGAC	TCTTCTCATGCTCCTCCAAC	186
>SoSSR1346	CGAGTGGCGATCTTGAGTT	GTCGTGCATTCTCATCCTTC	174
>SoSSR1347	AGGGTGTCCAGAAGCTGAAT	GTGGAAGTGGAGCTGAGAGA	204
>SoSSR1348	ATCAGAGTCGCACCGAATAG	GTTTGCCCGAATATGTCAAC	159
>SoSSR1349	AGCTCCACAGATGATGAAG	TTGATTTTGATTCCCACCAC	242
>SoSSR1350	GCCATCATTTTCATGCCTAAT	CCCATTGGTGGTAGATTTCC	148
>SoSSR1351	ATGAAGCTCTCCCTCCTCAC	GTGGTAGGCATTGTTGAAGG	138
>SoSSR1352	AATTCGGACCTTCCGTTATC	CTGTGACCTCCCCTGTTATG	157
>SoSSR1353	TCTCTTGAAATGTGGGGTTC	ACTTCGTCATCGCCAAGAT	204
>SoSSR1354	TTATGCCCATGTTGGAAACT	TTTACGGTCCTAACCAAAA	218
>SoSSR1355	GTATCTATTGTGGGCCGTTG	GCACGCCAACATCTGAAT	203
>SoSSR1356	CTCAAGCCAAACAGCTTCAT	TGAAGAAAGCTACGGAATCG	228
>SoSSR1357	GCAGCTGGTATGGTTTCATC	CACTTGCGATTGGCTTATTT	169
>SoSSR1358	GCGAGATGAACGAGGTAATG	CGGAAATGGTGAATTCGTAG	186
>SoSSR1359	AACCCTTGAGAAATGGAAGC	CCTCAAACCTCTGCCTCAAAA	250
>SoSSR1360	GAGGAGGAGGAGGATGTGAT	CCTCAAACCTCTGCCTCAAAA	150
>SoSSR1361	ATGAATAGTAGCCCTCAAATG	AGTAAGGGAGCTGGCAGAAT	250
>SoSSR1362	GGAATGGTTATCGAACAAGG	GCCTCGTCTTCTTCTCCTCT	244
>SoSSR1363	CAGGTTGAAGGTGCTGAGAT	GCTTCTTCAGGTTTCCCTTC	158
>SoSSR1364	CACAAATCACCCACCTTCTC	CAAGCAGTAGCGTTTCGATT	197
>SoSSR1365	GGGGTTTCTTTTCAACCAGT	TCTTCCCACCTTCTCTCACG	144
>SoSSR1366	TGGTGAACACATTCGACAAG	CAACCTTGGCAGATCAGAAT	239
>SoSSR1367	CTTTGTCGGTTGAGAAGCAC	TCTTCATTTGCATCCCTTTC	217
>SoSSR1368	GGGTGATGAGGAAGTACAGC	ACCGTCAAATCACACTCCAT	190
>SoSSR1369	ATCTGGAAGGGCCTAAAGAA	ACAGCCATTTTCAGCAAACCTC	125
>SoSSR1370	GTTATCCGGACGAGAAGGT	CTAGCAACCCAAAAGCTGAA	244
>SoSSR1371	CGGCCTCTCGATATTCCT	CAGTGACGCCAGAAGACTG	209
>SoSSR1372	AATGGTGGTGTGAGCAGAT	CAATGGGAATTTGATCTTGG	156
>SoSSR1373	ATCATGGAAGATGGACCTCA	CGTTCCACCAACAGAGTAGC	225
>SoSSR1374	GTCCGTGGACAGGATCATAG	TTCCTCCTGACAAATTCTGC	226
>SoSSR1375	AGAATCTGACGAGGATGCTG	TTTACCCCAAAGTCCCTTC	220
>SoSSR1376	TGGATACCTTTTGGAGGACA	ATGGAGACATAACGCCTGAA	246
>SoSSR1377	TTCAGGTTCCGCTTGTGTAT	CCGCTTCCTGAATCTTCTTT	198
>SoSSR1378	CACATCATCGATTCCCTCTC	AGCATTCTTGGAAATTTGCTG	151
>SoSSR1379	TGGGTCATGAGTTCCTGAAT	CATCATCAAACTGCCAGAA	210
>SoSSR1380	GCCTACAAAGTCTCTGCTCCT	GATCCAGGACTCGTACATCG	154
>SoSSR1381	GAAGAGAGTGCAGGAGGTGA	TCTTGACGGTGACGATGAG	197
>SoSSR1382	GGCGTGCTTAAAGATGACAC	GCATCCTATGAGCTCCTTCA	163
>SoSSR1383	AGTTGAACCAGTTGGCTACG	TTGGAATCAACCAAATCTCG	136
>SoSSR1384	TTCAAGAAAGCCTGTGGAAG	TGGGTTTGATTTTGGAGAGA	101
>SoSSR1385	ACCTCCGACTTCATGTCCTC	GAGGTTGAGCTGTTTGGTGT	240

>SoSSR1386	CAGTATCCCTCCATGGTCAG	AGCTTCCAAAAGGTTGGAGT	159
>SoSSR1387	CATCTCTGCAAGCCTCTGAT	AGAGGAACAACCCACCAAAT	138
>SoSSR1388	GCAGCAGCTAGTCATGGAAT	GGTGGGAGAAAGAGAGAGAGA	250
>SoSSR1389	CCAAACCCATCTCCCTCTAT	CGCAAAGGAGTAGTCAACG	192
>SoSSR1390	TCCGGAACCACAACACTACTTC	GCAAAGAGAAAAGGCGAAAA	164
>SoSSR1391	TTCAACAGCAGAATCATCCA	CTGCTACCAATCCCATTGTC	126
>SoSSR1392	CAGGAGAAGGAAAAGGAGGA	CAGCTTCGATCTCTTICTGG	208
>SoSSR1393	GTTGAGCCATAAACACAGG	CCAAGGAACATGAAGATTTCG	170
>SoSSR1394	AGAACCACGAGCAAAGACAG	ATCTTGGAGCGGTAGCTCTT	242
>SoSSR1395	CAGCATTACTGGGCAGTTT	TTCCTTCAAACCTCGAGCATC	181
>SoSSR1396	CATGCTCAACATCCTCGTC	GCCAAGAAAACCCTATCGAA	189
>SoSSR1397	ACGAGCAACCGTAATCGTAG	CTGCCATCACCCAATATTTTC	248
>SoSSR1398	TAGCGTGTCAACGATGAAGA	CTGCCATCACCCAATATTTTC	152
>SoSSR1399	AGAAGAAGACCCGAAAGGAA	AGGGGAACGTTATAGGTTGG	230
>SoSSR1400	ATGATGTGGGAAGCTGGAG	CAGTGCTTTGGCAGATTGTA	215
>SoSSR1401	CTGTCTCCGGTGTGTCTAT	GAGGCCTACTTTCAGTCACG	151
>SoSSR1402	GTCGTCCAAATCGTCCTTC	TCCGAGTCCAAGAAAATCAA	238
>SoSSR1403	TGGCTCCTTTGACTTTCTCTC	TGTACTIONGGCCTTGAAGAGG	189
>SoSSR1404	CTATGTTGGGGTTTCCACAC	TTCCACTGTGAAGGCTAGGA	111
>SoSSR1405	GACGAAGACGAGGAGGAGAT	CACTCTCGGCTAAGTTCTGC	250
>SoSSR1406	GCTCAAATCATCCCTCGTC	GCGGAGTATTTGTGATTTTCG	164
>SoSSR1407	CCTCTCCCTTCCATCTTCTT	GAGAATGTTGGCTAGCAGGA	120
>SoSSR1408	CGCAAGAATCACCTCAAAT	TTGTTGCGCTCTTCATCTTC	244
>SoSSR1409	GGAGGTTGGAGGATGAGATT	ACTGCTCCACAAAGCAAGAC	129
>SoSSR1410	AAGGTTCCGGAGAAGAAGAA	AACCTCCTCATATGGCACAG	245
>SoSSR1411	CTCCGTAGCCGTACCTATGA	GCGGAACCAATATGTCAGTC	139
>SoSSR1412	CGAGTTCGTCTCCGTCTACT	CCACCTCACCTATACTCC	221
>SoSSR1413	CTGAACAGCAACAAGCTCAA	GTCCACACCTCCTAAACCT	192
>SoSSR1414	AACAGCAGCAACAACAACAG	GGGTTTTACAGGGGACCTAA	209
>SoSSR1415	TGTTTTCGCTCTCCTCTTTG	CATCTCTGTCACCACCAACA	212
>SoSSR1416	TGGAGCTGTTGATGAGGATT	TTTGGGAACCAAGTATAGGA	167
>SoSSR1417	ATGAAGATTGAGGCAGGTGA	CTCATCAGCACCATTGGAA	147
>SoSSR1418	TGAAGATGGTGTTCGTTT	AAGCTTCTTCTCCACCAAT	149
>SoSSR1419	CCAATCTACCACCACTACCG	CTTATGGTCCCATCATCAGC	245
>SoSSR1420	AGCTGATCCTCACAGCAAAG	GCTCCTTCTGCCTCTTCTTC	243
>SoSSR1421	AGAGTGATTGCCAGAAAGA	CACAAATCCCCAAACAGAAG	111
>SoSSR1422	GCATGATCTGTTGAGGGATT	CACAGAAAGGGCACCTGATA	220
>SoSSR1423	AACGCTAGAACAATGCAAGG	ACGCTAGGATCTTCATGTCCG	224
>SoSSR1424	CCTTCCACCTCCTCTACTCC	TCTTCTCTTCTCCTCTCC	250
>SoSSR1425	CCACACAACCAAGAATCTCC	TGAGAATAGGCAGTGGAAAGC	216
>SoSSR1426	GTCGGTGTCCAAGAAGAGAA	GCTCCCAAACGAGATGTAAA	142
>SoSSR1427	GTCGCCTAAGTCAATCCAGA	TCGAAGACGAAGAAATCGAC	134
>SoSSR1428	AGAGGTGGATGAAAGACACG	AACAACCTTGAATGCCTCTCG	175
>SoSSR1429	GAAAGAGCGGATGTGGTAGA	GCAATTTCTTCCGACGTTTA	235
>SoSSR1430	GACGAGAAATCGCTGGATAA	ATGTACAGATCGCCGTTAGG	173
>SoSSR1431	GTAAAGCGCGTTGTGAAAT	CCGTAGTTCTTGATGGTTGC	166
>SoSSR1432	TGAAGTTGATCCCTTGAGAA	ATAGAAGTGGACGCAACTGG	202
>SoSSR1433	ATTATCGGTTTACCCGGAAC	GCCGCATGATTTACTTGG	246

>SoSSR1434	CACCATTCAAGACCCTTTCA	GTTGCCGAGATGGTGATTTA	139
>SoSSR1435	GAGGGCAAGAACGTCTTTTA	TACCCATCCATCATCTGCTT	239
>SoSSR1436	GCCTCAAGATCTCCTCAACA	AGAAACACCCGATTCTCAT	148
>SoSSR1437	CAGCCAGCTACGTGATATT	GAGCCGTCTCTCTCCTTCA	231
>SoSSR1438	AAGTATTCCGGACACATGGA	TTGTGCAATTGTAGCATCGT	204
>SoSSR1439	TCTTGCAGGGATTTCTTGAG	TGAGACTTGTCCGTTTCATCA	231
>SoSSR1440	GGCTGAAGATGTGATCGAGT	TCACTGCAATCCTCCCCTCTA	236
>SoSSR1441	AGGCTGAAGATGTGATCGAG	AGAAGTGGATGCAACTGGTC	240
>SoSSR1442	AAATGAAATCCGCCTTCTTC	CCTCTTGAGCTTGTTTTCCA	221
>SoSSR1443	CATCTCTTTGATCGCCTTGT	ATGTACTCCGCGACTTGTTT	125
>SoSSR1444	CCGTGAGCAACATAAGAACC	TCAGGACTTTGGAGAAGGTG	228
>SoSSR1445	AGGAGAACACTGCACAAAGG	TGTACCAGCACGAGAACTGA	216
>SoSSR1446	CAAAATGGCAAAGAGAATGG	TCCCGTCTTTGATTAGGTCA	200
>SoSSR1447	TCTCGAGGAGAATCGTATGG	CAATACCACTGCCGCTAACT	181
>SoSSR1448	GGGATCTTGCCTTTCTTCT	GCTGTAGCTCTTGCTGCTCT	131
>SoSSR1449	GATTCCTTGGACAGCCTCTT	GCGACGATCTTAGACCTCAA	236
>SoSSR1450	CAAAATGGCAAAGAGAATGG	TGATCCCGTCTTTGATTAGG	221
>SoSSR1451	GACGTTGGGTCTATCATTGC	ATCACAAGGTTGGGTGACAA	203
>SoSSR1452	TGGAGAAATTGTCACCCAAC	ATAGCCAGTGTGCATCCATC	213
>SoSSR1453	CTACCAAATCCAAGCAGTGG	GGCGCATTAAAGTGAGGTTA	222
>SoSSR1454	TGGCTTATGCTGCACTGATA	CCTCTTGCTGCTCCAATCTA	164
>SoSSR1455	TTACGCTGCACTGATGTCTC	CGAATGCGAGTTTCAAGATT	104
>SoSSR1456	TTACGCAGCATTGATGTCTC	GCATTCACACTCTTCCTGCT	183
>SoSSR1457	GGAGGAGTTCAAGGTGGAAT	TCCTGGCAAGAGTTGTCTTC	227
>SoSSR1458	CCTCCTCCTCTCTCTCCA	CTAGATTCTTGAGGCGTGGT	244
>SoSSR1459	GGACCATATCGGGGTTGTAT	GAGAGGGGAGGATGAGACTG	209
>SoSSR1460	TCTGAATACGAATCCGGTGA	GTCTGCGTCTGTGCTTTTTG	188
>SoSSR1461	CATGCTGAGGGAGGAAGTTA	ACCTGTAGGACCCATGAACA	250
>SoSSR1462	CAATGGATTCTGCAATAGGG	CTGATCCGGTCTTTGCTTTA	207
>SoSSR1463	AACTTTGGCACAGAGGATTG	CTAGCAACGGATACGCAGAT	197
>SoSSR1464	CTCTCAAATCAGGAGCCCTA	GAGAGATAGCGGTGAGACGA	112
>SoSSR1465	TGGATCGGTTTTGTACGTTT	CAGCACCTCATTCTGGTTG	223
>SoSSR1466	TTCACCAAGAGCAACAACAA	CGGAGAGGTCGTGATTGG	157
>SoSSR1467	CTCCTACCAATTCCAAAAC	TCCTCGTCTCATTGAGCTTC	241
>SoSSR1468	TATGCATTTCTTGCAGTCA	AGTAGATTTGGCCGAGCTG	210
>SoSSR1469	ATGCATTGGTGGCAGTTG	GTCGAATCCGGTCTCTC	117
>SoSSR1470	CAGTTGCCTTTCTCCTCCTC	GTCGAATCCGGTCTCTC	105
>SoSSR1471	AGGGAGCTCATGAAAGTTGA	TTGATGAAGCTTGGTCTTCC	231
>SoSSR1472	CGACAGACGCAAGAGAAAAG	AGAATTTGGTGTGCTGGAG	108
>SoSSR1473	CAGCAACACCAAATCTGCT	GTGGTAGGCGGATGAGAAA	220
>SoSSR1474	AGGAGCTCGCTCTTTAACC	TGAGGTCTAATTCGGTGAGG	114
>SoSSR1475	ATGGGATCGATGTCTACCG	GGCACGTCTTCGAGAGTGTA	248
>SoSSR1476	AATACCGACGTTTGGTTTCA	TTTGTCTGCTTCCAACAAC	171
>SoSSR1477	AATGTGGGTGTTGAGACCTG	TTGGTAGAGGTGCAGGTAGG	226
>SoSSR1478	ATGGCAAACCTCTCCAAAA	CTTTGTGCTTCACGAGGTT	188
>SoSSR1479	AAGAAAACGAGGATGGCTCT	CTCTTCCATTGGCTTGGAC	215
>SoSSR1480	TACGCTGCACTGATCACTCT	TTCCTGCTGTAGCTCTTGCT	170
>SoSSR1481	ATTGGGTTGCTTGTGTGATT	ATGCGAGTTTCAAGATGGAC	138

>SoSSR1482	AGCAATCACTGGAACAAAGC	ACCTCTTGCTTCTCCCATCT	200
>SoSSR1483	GCCGCATTGATCTCTCTACA	TTCCTGCTGTAGCTCTTGCT	167
>SoSSR1484	AGTCATTGCTGGTCGCTATC	TTATCGCCAGGATCATGTTT	195
>SoSSR1485	ATGGATAGTCCCAGCGTACA	CGTGGTCGGCTACATCTTTA	188
>SoSSR1486	CCAAGAGAGCTGTGTGGACT	GCACTACCCTGATCACCATC	195
>SoSSR1487	TGGAAGGTCTGCAAGGTG	TTGAACAACCCCAACTTCTC	250
>SoSSR1488	AATCCGATCCCTTCTTCAAC	CAAAGAGCGAGGAGAGATCA	168
>SoSSR1489	CATGGCCTTCATCTTCGTC	CGACGATTCTCTGAAACAGC	162
>SoSSR1490	CAGATTTATGGGCTGGAGAA	CTGCATTCTGATCTCCGTTT	235
>SoSSR1491	TCCCTGACTACATCGACACC	GAGGTGGTGGAAGTTGAGG	104
>SoSSR1492	CCATGACCTACTCCATGGTC	GGACGAGTGTGTTAGGCTGT	116
>SoSSR1493	TTCCACGTGTTTGTITTTGTG	CATGGACATCATGCTGAGAG	159
>SoSSR1494	TCCAACTCACCTTCTTCGAC	GTCGCTCTCGGTAAACTCAA	224
>SoSSR1495	CCTCCAAAGACGAAGATGAG	GGAATTGGGGATTTGAGATT	175
>SoSSR1496	ACGGACTAAAGCGTCTCAGC	ATGAGCCTTGC GACTCATAC	155
>SoSSR1497	CCCAACCAGAGCTTGAGTAA	CGCAATAGGAATCGAGAATG	118
>SoSSR1498	GCGAGAGGAATGTACGAGAA	AACCTGACTGCTTCATCTGC	244
>SoSSR1499	TGAAGCCTCTCAAATTAGGAAC	CTTCCGCTGAGAGATAGCAG	120
>SoSSR1500	CTCAATTCCGCCACTCTCT	CAGCCTTTTCCTCTGGTAGG	153
>SoSSR1501	CTCACCGGAACTCTTGATGT	TGAAAAGCAGGATTGCTACC	227
>SoSSR1502	GGCCTCTCAAATCAGGAAAC	ATCGGAACAGATCTCAGCAA	103
>SoSSR1503	GTCTCAATTCCGCCACTCT	CGGTGAGAGGAGTCACAGAT	100
>SoSSR1504	CATCTGGATTTCGCTATGAGG	AGGAGTCACAGATCGGAACA	111
>SoSSR1505	CGCTTCTGAGGCTTCTATCA	GGGTC ACTTGGAATGATTTG	189
>SoSSR1506	TGTGCTATTTCTCGCTTTC	CACCGCACCTTG TAGCAC	244
>SoSSR1507	TCTTCAACACCGAACACTCA	CGGAAACTTCTCCACCACTT	218
>SoSSR1508	TCTTTGTTGAGATCGAACACC	AAAATCCGATCCATCTCCAT	241
>SoSSR1509	GAGGCGTCTCAATCTACGC	GCATGGTTCCGATCAAATAG	176
>SoSSR1510	GTTGGTGGGATATGTTTGA	ACCTCAGCCATCATGAGAAG	242
>SoSSR1511	TCCCCATACCAATATGTTT	CGGCATCTATGCTCTCTTTC	173
>SoSSR1512	CAAGGCCAAGACCAATTTT	GCACGTAGGGGTCGATATAA	218
>SoSSR1513	GCGAGTTCATGATGGATAGG	CCTCGCTCTCGTTGTAATGT	201
>SoSSR1514	GGCAGTTTCTTTGGCATCTC	GTGATTCCACCTCACCAAAC	181
>SoSSR1515	TCCCTAAGTGGTGCTATTTCG	AATACCCAGCATCTTTTCC	219
>SoSSR1516	TAAAGCCTCGTCCACAAAAG	AAAATCGGAACATCTGTGCT	117
>SoSSR1517	AGAAAACCTCTCCACACCT	GTCTTGCTCGGATTCATTTG	198
>SoSSR1518	GAATTCTTCGGTGTCTTTTG	TTTCGCCATCTATAGCCATC	216
>SoSSR1519	TGGGTGCCTCATTGTATGT	TTATAAAGCGCCATCTCCAC	206
>SoSSR1520	CGTATCGAGCTCCAGTGAGT	GGATCATCTTGGTCTGCAAT	240
>SoSSR1521	GTTGTGGATGGGCAAATAAG	ATGTCCAACCATCTCCACAC	231
>SoSSR1522	TTCAGAGACTCACGGAAAG	CACGATAAATGCGGCTTAGT	177
>SoSSR1523	TGAGGAAGGAATGACTTTGG	AGAATGAAAACAGCGACAGG	159
>SoSSR1524	TCCTCTCCTCCTCTCTCACC	TGCTCTTCTTTCCATTCTCG	222
>SoSSR1525	GTAAACTCGCCTCCTTCCAT	GAGAGGGAGGTGTTCCAGA	160
>SoSSR1526	ACGGCTAGTGCTGATAGTGC	ATGCTTCTCTCCACCTTCT	228
>SoSSR1527	ATCATTTCTTCCAGCACAA	CCCGACATACGTGAAGAAAC	219
>SoSSR1528	AGGAGCTGAATCAAGGGTTT	GTACAGCCAACCATCTCCAC	193
>SoSSR1529	GGACACAGAGACCTTCAGGA	CTCAGTAGAAAGCCCAACCA	183

>SoSSR1530	TGCTCAGAGTTGCACAGAAG	GTTTGATGGGAACATTGAGG	229
>SoSSR1531	CTTCGCCATCACTTCCTCTA	ACACCTCCACCATCATCATC	178
>SoSSR1532	AGCCATGGAGAAGAAGAGGT	GCCATAATCAATCGTGAAGG	198
>SoSSR1533	CGACGAAGATGGAGAAAACAC	GCTCCGAAGAAGGAGAAAAC	208
>SoSSR1534	AGCAGGGGATTACATCCTTC	TCATCATCATCATCGTCGTC	121
>SoSSR1535	AACTATCCCGTCCACTCACA	ATACGAATCAATTGGCCTCA	158
>SoSSR1536	AATCTGGGAGGAAGTCGAAG	TGCTTTTCCTGTTCAGGTTT	112
>SoSSR1537	AAACAACCCGCAGCAGAG	GTCCGCCTTCATGATCTTC	214
>SoSSR1538	TTCATCATCCAAAAGGTCGT	TCAGCTGTCATTTCTTTCC	146
>SoSSR1539	TGCAAATTGAGAGACGACAA	TATGATCCGCTTCTCCACAT	230
>SoSSR1540	TTCTCACTCCAATCACTGCTC	TGCCTCAATCCAGCTAGAAC	236
>SoSSR1541	TCTTCTGGATTGGTGTGGT	TCCACTGTTGAAGGAAGGAG	248
>SoSSR1542	GCCAATTTCTCCACTCTCAA	CTCGGAAGCGTTTCTTCAT	208
>SoSSR1543	CAGAATTGATTTGGGGATGA	TTTCATCGCTGCTATTGCTA	107
>SoSSR1544	GCGTTCATCATCATCTTCAC	CTTGGAACCTTGAAGCACAG	179
>SoSSR1545	AGAAGGCAAAGATGCTGATG	CCCGTTTGAGTTTGAGTTTG	187
>SoSSR1546	CTGCCCTCTAACACGGTTGAT	CTGCTTCTTAGGCAACTCTG	236
>SoSSR1547	ATGGACCCTCTCCGTTTTAC	CTTCTCATCGAGCTCTTCCA	187
>SoSSR1548	ACCTCGACCTCTCCAACAAC	CATCTGGCGAAGAAGAAGAG	250
>SoSSR1549	TCCTTCCTCATCCAGATCAA	ACCCCAACTCCTCCTCAAT	220
>SoSSR1550	TGGAAAACGATTCTGAAGAAG	AACGGAACAGCAATCAAGAG	158
>SoSSR1551	ACAATCGCAACTTCTTCAGC	GTCAGATTCATGGGGTTCTG	243
>SoSSR1552	AGCAACAACAACAACAGCAA	TTGAAGGTGATCCCCTTGT	231
>SoSSR1553	TAGCTTCTTCGCAGTTGCAT	GATTCTTGGTGACTGCTGCT	212
>SoSSR1554	TATGATCCCCAAATGGCTTA	AATGGGAGACATGAAGTTGG	250
>SoSSR1555	TGAGTCGTCATTGGTACACG	TAGGATCAGCCCAACCATAA	197
>SoSSR1556	GACGGCGAATTCAACTAAAG	TCGTGGAGAAGAGAATCCAC	102
>SoSSR1557	TTCTCTTCTCCACGAAGCTC	GTTCATTTTCACGGGTTTTG	250
>SoSSR1558	AAAGACAGCTTCGACTTGGA	ACTTCGCCTCTAGCTCATCA	118
>SoSSR1559	TCTGGTCGACTCTGACGATT	AAGTTGAACGACGAGGTGAG	163
>SoSSR1560	GAGCAGTTCTGCAAGGACAC	CACTCTAGCTGCTTCCCCTA	234
>SoSSR1561	GGAGGTTTTGTTGATTTTGG	TTTGGTGGTGGTTTTGAAGT	112
>SoSSR1562	AATTCCTTGACGGGTTAGG	GTGTAAATGGCGAGATCAGG	233
>SoSSR1563	GTATGGAGCATCGGGAAAG	GCCATCCCAAATTTAAAGG	228
>SoSSR1564	CACCGAGAAGGAGCAGAATA	AATCTTGACAGCAGCCAATC	172
>SoSSR1565	AGCCGTTGATGCTGTTTTAG	GTGAGATGCAAGCCGAAC	226
>SoSSR1566	GTGCATCAAGGGCAACTC	CTCGGCAACTCAATGATCTC	246
>SoSSR1567	AAACATCGGTTGATTGTGGT	GGCAATGTCACAAAGAGTCC	230
>SoSSR1568	GGAAAGCGGAGGAACAAG	GAGATACCTCATGCGTCCAG	177
>SoSSR1569	AGGCATCGTCATCACTGTTT	ATTGTCGTTGTCGTTTTGGT	187
>SoSSR1570	ATAAACTCGAACAGGGTGGA	GATCGTCGGGGAAGAATATC	175
>SoSSR1571	TGATGAGGATGATAACATGTGC	CCTATCCTCTGCACACCATC	174
>SoSSR1572	GGAGGAAGAAAACGAAGAGG	CCGAAGAAATAATCCCAAGAC	246
>SoSSR1573	CTTCACCAACCAATCCAGAG	CCTGACCTCGACTCGTACAT	171
>SoSSR1574	CTCGACGGTGGGCTATTT	GCATAATTGGCAAGCAGAAG	243
>SoSSR1575	GGGGCAGAAATTGTCCATA	AGTGAAACAGAGGCAGAATCA	216
>SoSSR1576	TCCAAGTTGAGGAAATGAGG	GACATCAATACCGCTCCATC	242
>SoSSR1577	GAGGATTAGGCAAGCTGAGA	GTCTCCTCCTCGGAATCAGT	237

>SoSSR1578	TGAAGAAGCTTTTGGCAGTT	CGTAGTCCTCTGCAATCACA	192
>SoSSR1579	TTTCGGGGAATTACGATATG	CCTCGTTTCTCCGCTTTTACT	137
>SoSSR1580	GTAAAGCGGACAAAGACGAA	CAGTTCTCACTGCCTCCATC	130
>SoSSR1581	TCACGATGTTGTGAATCTGG	TTTTCTTGCGGATACTGAGG	156
>SoSSR1582	GCAAACCTCCCACATTTGTTT	TTCCCTGGCTTGAGATGTAG	178
>SoSSR1583	CAAAGAATGCCTACGAAACC	CGCCGTATTCTATCATCTCC	197
>SoSSR1584	CATCCATAACGGAAACCCTA	GAGATCCAATCAATCCGAAA	242
>SoSSR1585	CTTCTGCTTCTCAGCCTCCT	GATTTTGGAGTTGGGATCG	228
>SoSSR1586	CATTTTCTGGTGGATCAAGG	TGCTTGAACATGCCAACTAA	216
>SoSSR1587	GGACTCTAAGGCTGCTACCC	ACGGATTCAACATCAGCATT	222
>SoSSR1588	GAGAATCAGCGGAAATCAA	AGGGGCGGTAGACTTTGTAG	100
>SoSSR1589	TCTCTACCGTATTCATCGTC	GAGGCAGAGCTTCGAGGT	249
>SoSSR1590	GACACCTCGAAGCTCTGC	ACATGCACCCGTAGTGGTAG	142
>SoSSR1591	TGGAAGACAAAGAGCACCAT	CTGACTTCCCATGCTTCAGT	171
>SoSSR1592	AGCTAGTTTTGGTGGTGGTG	TTGGGAAGCATCATAAGGTC	213
>SoSSR1593	CGGTAGTGGTGGTAGATTGG	GTGGTTGAGAAACGGAGATG	138
>SoSSR1594	GATGCTTCAAATTCCTGGTG	GCCCAATTTTCTGGTTTCTT	237
>SoSSR1595	CCTCGGATGTCGTGTCTAAT	ACAAATCCCACCTCACAGAA	232
>SoSSR1596	ACTAAATTGGAGCCACACGA	GCCGGATCTTCTAACATGAG	122
>SoSSR1597	GATTGCCAGATAGGTGTTGC	GCATCATCAGGAATCCAAAA	104
>SoSSR1598	GGATTGCCAGATAGGTGTTG	GACTTGGCATCATCAGGAAT	111
>SoSSR1599	TATGCCCTAGCTTCAACAGC	GCGGAAACCTCCTTATCTTC	142
>SoSSR1600	CAACCGCGATTTGATAGAAA	CGGCGTCAGAAATACAAAAC	247
>SoSSR1601	AGGTCCTTGATGTGTGGAGA	GGGATGGGCTAAAGAGAGAG	201
>SoSSR1602	TGGTCTGTTCCTTTTGGAGC	GCAACGGGATTTATCATTTG	201
>SoSSR1603	GTTAGGGGATTGGCTGAGTT	ACCTTAGCCCACTCAGGTTC	176
>SoSSR1604	TGAAGATCCGAAAGAATCCA	CTACTCCTACGCCCTGACAA	123
>SoSSR1605	TCCACCAAGCTACAACCAAT	AACCCCATCATCGAGTCAG	163
>SoSSR1606	CTCAATCCACAGGCACTCTC	AGATGGTGGAGATGGAGACA	155
>SoSSR1607	CTCCTGCTGCAAAGAGTGAT	AGGTGCACTAGCACAAAGGAG	185
>SoSSR1608	ATCTCACCACCACCACCAC	GGAGTAGGGAGGAAATGGAG	230
>SoSSR1609	TTTGTCCATCGCAGTTATGA	AGATGTGATCTCACGCCATT	239
>SoSSR1610	TTTCTATGGTTTCGCAATGGT	AAAGCAAGAATTCCCCACAC	181
>SoSSR1611	GGTTGAGTTTCTGGTTGCTG	GCCAAGGATGTCATGTATT	130
>SoSSR1612	AAGGTCGGAAGAGAAAAGGA	GCCAAGTGCTCTAGGTTTTG	140
>SoSSR1613	TAACAGCGTTGATTGCGTTC	ACGTGGCAAATATCAACACC	141
>SoSSR1614	GGGAGACGAATGATGGTGT	CTGTAAAGGTTCTGGCTCA	210
>SoSSR1615	GAGACGGATGATGGTGTGAT	TAAAGTTTCTGCATGCTGCT	109
>SoSSR1616	GTGTCGTTGTCTTTCCCATC	CGTGCTCCTTAATTCTCCAA	134
>SoSSR1617	TGCGAGGGATTGATCTAAAG	CAGAGCTTCCTCAACGACAT	199
>SoSSR1618	CATGTCTGTCGAGCTGAGG	GATTCACCCACGTTTTCAAG	244
>SoSSR1619	GCTGAAGATGCAATCGAAGT	ATAGCTGTCAAATCCCACCA	237
>SoSSR1620	CACTGCTAAGGTGTGGGATT	ACGACCCTTCTCGAACTTTT	234
>SoSSR1621	TCGGAAAGAGGAGGAACAA	ATGGCTTTCACACTCCAGTT	139
>SoSSR1622	CAGGAAGAGGTGAAGATGGA	TGAGATCCATGTACCCGAAA	115
>SoSSR1623	GATCTTCTTCGAGTGGCTGA	GTTTCTTCGAAAGGGAAAGG	231
>SoSSR1624	AGTCGCCCTTCTCTTTTCAT	GCGGAGAGACTGGATATTGA	178
>SoSSR1625	AAGCTGAGGATGATGATGGA	ATTGCTCAATGGAACACGAT	215

>SoSSR1626	GAGTGTCCGGTGTGATTTTC	CGCTTCCAACATTCTCTTGT	137
>SoSSR1627	CCTCAGGCATCAGTTTCAGT	CCGGTGTAATATGTAGCAGACG	170
>SoSSR1628	CATGAGAAGGGTGATCCAAG	TGGTCCAACATTATGCTCCT	204
>SoSSR1629	TTCTCAAAGAAATGGCATGG	TTCGTTGAAGACGAAAGAGC	132
>SoSSR1630	CTGAGAGAATCGCTGGAGAA	TCCAATTCCTCCTCTCCTCT	167
>SoSSR1631	GGCGTCCAAGTTTCTCCT	CCTCCATCTCATCCCACAG	197
>SoSSR1632	ACCTCAATTTACGCAGCAAG	TGCAACACTAGCATCTACGC	226
>SoSSR1633	GCGGAGGAGAAAGATGAGGT	GAATACCATGTGGCGAGAGT	224
>SoSSR1634	AAGATGAGGTAGCAGCAGCA	CCATCTCCAAAGAGAAGCAA	245
>SoSSR1635	CTCTCTAAGGCCCTCACTCC	ATTCAAAGGCCCGCTACTAT	240
>SoSSR1636	GGGTTCCCTTTCATCTTCAT	AAAGCTGCCATAGATTGTCTG	159
>SoSSR1637	ACCTCAACTCTCGACTCCAA	GGAATTTGGACCTTCAGCTT	242
>SoSSR1638	GCTTCCCCTTCTCGTTTACT	CCAGCTACCTCTGTTGGAGA	194
>SoSSR1639	TATGGGAAATTGGAATGTGG	TGTGATGAGAGACCTGCTGA	110
>SoSSR1640	TTCCCCTTCTCGTTTACTCA	GGAGAGCAAGTGATCGAGTG	186
>SoSSR1641	TCAACTCCTCACAGCATTCA	TGGGCCATTACCAAACTTA	225
>SoSSR1642	GCAACTTCAACCTCTTCCTCT	TGTTGGTGTTGACAATGAGC	250
>SoSSR1643	GTCACAGGCTTGGAACA ACT	AACGTCTAACCGCCTCTTCT	235
>SoSSR1644	AAATGCAGCTCAATCGGTAA	AATCCACGTGTCGAGTTTTTC	203
>SoSSR1645	GCAACCTACCTGAAAACCTCG	AGAACGATTTCCGGCTCTACC	222
>SoSSR1646	CGGAGGAGGAGAAGAAAGAC	AACACGCATTAGGATTCTCG	198
>SoSSR1647	TAACCTTATTCCGGCAGTCA	TGTGGCTTGAAGTATGCTA	187
>SoSSR1648	GCACGCTTTTCTACTTCTCCA	ATTCCCCTCCACCACCAC	190
>SoSSR1649	CCAGTCCAAGCTTAAACAG	GTCGAGAGTCTGAGGCACAT	233
>SoSSR1650	TGGCTGAAACTCATCTCCTC	CCATTGCAACCTTCCACTAT	203
>SoSSR1651	AATCTACTGGCCGATCCTTC	TCCAATTCAAACACCTCGAT	210
>SoSSR1652	CACGAAAGTTATGGGGCTAA	AAACACAGCCTCCAAGTGAG	159
>SoSSR1653	GTGCATTCAAGGGTCTGAAG	CAGAGATGGAGAGCGAGAGA	209
>SoSSR1654	CCAGAAAAGTGGTGTTTGCT	TGCACGATTA AACCTTTGT	179
>SoSSR1655	GAACGGGAATATCGTGATTG	CTTAAGGTAGCCCATCCACA	185
>SoSSR1656	GACAGATCTACGGCATCGTC	AGCCTGTGTCCACTGAGTTC	196
>SoSSR1657	GGTAATCGGATGATGAGCAC	TGGGGTTTGAGAGAAACAGA	241
>SoSSR1658	CATTATGAATCAGCCGGAAG	TCGACGGAAGGTTTATCTTG	247
>SoSSR1659	GATCTCCGTTCTCCTCTTC	TAGTCGACGTTAGGGTTTGG	139
>SoSSR1660	GCACGATGCGGTTAATATCT	CGAGACA ACTCTCTCCCTCA	166
>SoSSR1661	GCTCTTCTCGGAAGGCTCT	TTCAGCTGGTCCATTGATTT	131
>SoSSR1662	TTGGGATCAATCGGTAAAGA	TTAGATCGGAGAACTTACTACTGGA	139
>SoSSR1663	TCCCACTCACAGACAGAACA	AGGATTACGAGCAGCAACAG	218
>SoSSR1664	CTACTGCAACCGCAAGTTCT	TGCCGAAAACAGCCAGAG	248
>SoSSR1665	CTGATGCCACCATTCTCTTT	GAACCTCCATGCACATATCC	186
>SoSSR1666	AACCTCCTCCTCCTGCTTTA	GCACTCTTCTTCAGCTCTGG	137
>SoSSR1667	TCAACTTGCCATAACCGAAT	TCAGCTTGCTCAAAC TCAA	217
>SoSSR1668	ATTAGGTGCATGACCACCAC	TGGGATCTCATCGGAATCTA	168
>SoSSR1669	CAAACCTCACGTC ACTACCC	GCGACAGGAGTAGTCGTTCT	169
>SoSSR1670	CCGATTAATTTCAAGCACCA	GGGATGATTTGATCTTTTGC	126
>SoSSR1671	CGGCGATATGCTCTACAAAT	GATGACCCCACTAATTTAGC	120
>SoSSR1672	TGCTGCCACATCTGACTTTA	ACTATCGTCGGGGCTAAAAC	222
>SoSSR1673	GAGCTACCACGAGCCAGTTA	TAATACCACCAGACCCCTCA	211

>SoSSR1674	CAGCTGAAGAGGTTGTGGAT	CAGTTTGCTGAGGTGGTTCT	120
>SoSSR1675	TCCAAGACAGAATCAGCACA	TGCAATGCTTTCAGAGACAA	193
>SoSSR1676	GGGGAAGTTTCGTTAGAGGA	TAAATGCGCACGAGAATACA	220
>SoSSR1677	TGAAGAAACCGAAGATCACG	AATGCTAACAATGCCCTGTC	169
>SoSSR1678	ACAGCATCGAACTCCAGAAC	GCTCCGTTACATCATCCTTG	126
>SoSSR1679	AGGAAGATCACATCCATCCA	ATCTCTTCCCGAAAGCAACT	223
>SoSSR1680	CATGTTTCAAGCGATCAACA	AGCCTTTCATGACGGTTTCT	227
>SoSSR1681	CATCAACCTCGTCATCTTCC	ATCTGAGCTCTGCGACTGAC	202
>SoSSR1682	TGTGAATGTGGGCAACTATG	AAGCCCATCTGTTCCTTCTCC	211
>SoSSR1683	GCCAGTCGCTAAACAGTGAG	CATCCACATCCTCATCTTCC	117
>SoSSR1684	CGGAAATTGAGCCCTACAT	ACCAGTTGCCTTCCAGAGTT	216
>SoSSR1685	ATAACCATCAACCCCTCCAC	GATTATGAGTGGCAGCTTGG	102
>SoSSR1686	GACTTATCCCGCTGCTACCT	AAGTATGGCTGGTGGAGGAT	131
>SoSSR1687	TCAACTGCATCATCATACC	CAGACTCCGATCCTTCTCTCC	216
>SoSSR1688	TTCAGAAACGGCTTTAATGG	GAAAAGCTCCAATACCCACA	204
>SoSSR1689	AGATGCGAGCTCAAATCAC	GCGGAGCTCTTAAACGAAG	152
>SoSSR1690	TGCAGCTTCCTAATCACCTC	GGAGGAGTTGGCGTAGTAGA	234
>SoSSR1691	GTGGAGATGGATGTGGAGAG	TGGGAGGGTCAATCAGTAAA	247
>SoSSR1692	TCAGCAGTGATGGATCAAGA	TTGATCACCAACAGGAAGGT	126
>SoSSR1693	CCTTCCTACTGGTGATGCTG	TCTTCATCTCCAACGCTTC	243
>SoSSR1694	GCTACAACCACAACCCTCTC	CTCGGAAAGTAAGCACCGTA	109
>SoSSR1695	AGTTTGCCAGCTCTCTCCTT	CAGCATCGGCTTCTCTTTT	190
>SoSSR1696	GGTGGTCTATTCTGATCC	TACGTCTCCTCGCATCTCTC	162
>SoSSR1697	AACCGGAATTAGGGCTAGAA	GATGAACTGGAAGTGTGCT	211
>SoSSR1698	TTGTTTTCCTTCACTCCCTCT	ATTGAAATGGGTTGCTTCTG	101
>SoSSR1699	CATCGGTTCTCCATCATTTT	TAAGTGGCCCAATACACGTT	200
>SoSSR1700	TGGCAACTTCTCTCAACTTC	TGAAGGAGACCTGGATCAAA	196
>SoSSR1701	GAGACTGTTGCTGCTGTTGA	AGCCGGTCTACATTCTTGTG	111
>SoSSR1702	AGTCCTGCATGAGCTTTGAC	AAGTGGAACGGCACCAAC	179
>SoSSR1703	TAGGGATCTTGCTGAAGTCG	CCTGCCAGACAAGTAGAAA	120
>SoSSR1704	GTTTGCGAACGAGGAGTTTA	ATTCTCCTGCCATACCTCT	205
>SoSSR1705	CGGTGAAGCAGCAGTTTTAG	GGAAGGAAAGAAGTCTTGG	229
>SoSSR1706	AGAGCTCCCAAAGAATTGGT	ATGTCTCCCTCTTCCACGAT	193
>SoSSR1707	GGAAAAGGGTAGTCGAGCTG	CTTACGTTCTCTTCTCTGA	230
>SoSSR1708	GAGAAGGATCAGGTGGTGGT	TCGATTCTGATCGGGTAG	232
>SoSSR1709	CGGTGGTATTGGCTTAAATG	ATCCTTGGACTGAGATGCTG	249
>SoSSR1710	ATCAAAGGTGACTCCCACA	TGGATGGAGGGAGTATGAAA	214
>SoSSR1711	TGCAGAAGAGGGAGCTAGAG	GAGCAAGTGTAGGCGTCATT	217
>SoSSR1712	ATGAGACGAGCATGGTTGAT	GATCAAAGCTCGAACTCCAA	228
>SoSSR1713	CACTGGCCCTACATCAGAAC	GCTGCAATGTCAAATCCTT	216
>SoSSR1714	TAACAACGCGAAAGTGGTG	CTTGGATGGGAACCAGTATT	148
>SoSSR1715	TAGGGTTAGGGTCCAAGGTC	CAAGCTTCCAAGTAGCCTCA	191
>SoSSR1716	CACTGTTTGTTTTGGGGAAG	GTTGATTTGGCTGCAAGAGT	240
>SoSSR1717	AGGACTCACGGGAGAGAGAC	TTCCCGAATTCCCGTCAT	172
>SoSSR1718	GGTTCCTGTGTCCCTATCT	AGTTGAGCGTCTGTTTCTG	212
>SoSSR1719	AGAGCAAGGAAAAGGGTGTT	CAATCCCTCCATCTTCATCA	168
>SoSSR1720	TCCTTCAAAGAGCAGCATAGA	CCACACCCTAATAGCTCGTC	117
>SoSSR1721	TGTTCCAGACTGTGCGACTA	TTTCGGGAACGACATAAGAC	230

>SoSSR1722	TCTGCATGCCTTTCTTTCTC	ACTTGCTCACTCCCTTCACA	194
>SoSSR1723	ATTCTTGCAGAAGACGTTGG	GGGTCGAGACAATTCATCAG	192
>SoSSR1724	GCCACATGGATCGCTATTAC	CAGCGAGTCAGCCTTGTC	248
>SoSSR1725	GACGGCATCTTGGTCAATAC	ATCATCCAGCCATTTCTTCA	165
>SoSSR1726	TGTTGATGCACGGTAAAGAA	AACTCCCCATAGTCCCAAAG	111
>SoSSR1727	GAGATTGACGGAGGTGAAAG	AGGAAAACCACCGAATTCTC	204
>SoSSR1728	CCACCTCGTCATGCTTAAAT	AGGAAAACCACCGAATTCTC	243
>SoSSR1729	TCCGTCTACGAGAAGAGCAT	CTCAGCTTCCGAATGGATAA	242
>SoSSR1730	TTGGATGATGGTTTCCTTGT	CACGATACCCAACAAACTCC	202
>SoSSR1731	CTTCTTCGCCTTTGTCATGT	CGTGGTCTTCTTCACGAGTT	177
>SoSSR1732	AGAAGGAGGCAGAGAAGGAG	TCAACAAAGCGTCTGTGAAA	171
>SoSSR1733	TTCGGGCATAATAGATTTCCG	GGTGAAAACCTCCATCTCCAG	168
>SoSSR1734	CTGACAATAGCCACGCAGTA	TCATCCAGCACGATCATGTA	173
>SoSSR1735	TCCGCCTGCTGAATATTTTA	CCGTGGGACTATGTGAAGAT	220
>SoSSR1736	GAACCAATGGATGTTGAAGG	ACTGCATCGTTTCCAGTAGC	209
>SoSSR1737	GGGCTGTTGGAGGATTTATT	TTCCATCTTCTCCATTCCAA	250
>SoSSR1738	ATGCTCAACGAGAAGCAATC	TGCCACATAACCTCTTCTCC	187
>SoSSR1739	CAAGAAGATGAAGCCGAAAA	AGGAACAAACTCCCCATTCT	196
>SoSSR1740	CCGGTAGCAGGTACTTCATC	CGAGGGAGCCGTAGTTTT	206
>SoSSR1741	CAACAAGACGGCGTACATTC	AGGATTATGAGGGGCAGAAG	204
>SoSSR1742	CCTTCTTCGACATCAAATGG	GGTATCGGAAGGGTAGAGGA	179
>SoSSR1743	CATCGTTCCAAACCTCAAAC	GGCAGGAAATCGTAGAACTG	222
>SoSSR1744	TGAATTACTTCGGGAACCTGC	TCAGCAAATCGAATTTAGGC	234
>SoSSR1745	GAGGAGGTTGGAGGTTTTGT	CAATTTAAGGCCATCTGCAA	154
>SoSSR1746	ACCATTGTTCCAAACCTCAA	GTAACGGAAAAGGGGCATAG	123
>SoSSR1747	ACCATCGTTCCAAATCTCAA	GCAGGAGGTCGTAGAACTGA	229
>SoSSR1748	CCCAATTCCTCAACACAATC	ATCGACGGCATAAAATCGTA	247
>SoSSR1749	GCCTCACTTCACTCAACCAC	AGTACATGACTTGCCGGAGA	168
>SoSSR1750	CAAGGCCAAGACCAATTTTC	CCGGAAGAACTGGTAGCC	217
>SoSSR1751	AAAGCGAAGGAAGTGGAAAT	CTATCCTCAAAGCTGCCAAA	233
>SoSSR1752	GCGATTCTTCTCCAAGAC	ACGGGAGCCTCAGCAGTT	195
>SoSSR1753	AAACAAACGTCAAGCCACAT	ACGATGCAGATTCTTGAGAG	240
>SoSSR1754	ACCCACCACTCTCCTTATC	ACTGGCTGTGTCGGAAGAC	164
>SoSSR1755	GAGGGAAGAAGAAAGGAGGA	ATAGGCTTGGAGTTGGCTTC	176
>SoSSR1756	GTGCCAGCCTTCTACTTGAG	TGGACATCATTGCTTTGAGA	220
>SoSSR1757	TCGGTTGGTAAAGTGAGGAG	CACAACACACACCGACAAAT	162
>SoSSR1758	TGAAATACCTAGCGGAATCG	GTGGTTATGAGCCACAGTCC	176
>SoSSR1759	CCCAGTCTACCTACCAGAA	AGCTTCTCCTTGGTGGTCAT	180
>SoSSR1760	CATGGAAGAAGCAGATCCAG	TGTAGACTCTTCGCCACCTC	128
>SoSSR1761	GCGCTTGCTTTTACTACTGC	GCTCCTCTTTGTGCATCATT	241
>SoSSR1762	GTCTGCCGAAAGGAAGAAA	AAACCCTCCTTCAAGTTCGT	178
>SoSSR1763	TCAGCTATGGCTTCTGGAAT	TTTGCTCCAGTCATCTCTC	118
>SoSSR1764	CGGTGATGATGAGAGGAAAG	CGCCTCGATCATTAAGAAG	242
>SoSSR1765	CTGGCTGAGTAGGGCTACAA	CATACAATGCTCCACCTTCC	164
>SoSSR1766	GAAGTTCTCTCGGGGAGTGT	CGGACTCTGCTGTCACGTAT	229
>SoSSR1767	GAAACACCCCTCTCCATCTT	TTATCCCTGCCCTTGTTGTA	190
>SoSSR1768	GAGACACCCCTCTCCATCTT	TCAGGCTTATCAGTCCCATC	196
>SoSSR1769	CCGCCTTTATTATGGTTTCC	GTAAAACGACTCAGGCGAGA	175

>SoSSR1770	GTTCCCAATAGTGGATGCTG	CGTAATCTGGCAATCCATTC	119
>SoSSR1771	GTTCCCAATAGTGGATGCTG	AATTGGCACCATAAATCACG	137
>SoSSR1772	ATCAAATCCGCCATCACTAA	ATGGAAGGTGTTGTTTTGGA	219
>SoSSR1773	GGTGGTGATGAACAAAGAGC	AGAGTCGAATCCATCCACA	243
>SoSSR1774	ATGGCCAACAAATCTTTTCC	GCTCCTGCACTTCCTTGTTA	142
>SoSSR1775	CGGCAGTACAACCTCAATCT	GCCAAACAGCAGCATTATTT	229
>SoSSR1776	GACCGAGAAGCAGAAGAAGA	GATCGCTCATCCGCTAAGTA	182
>SoSSR1777	TATCTTCCACCACCACCAAC	TGTTCTTCGGAAAAGAGCTG	139
>SoSSR1778	TTAGCCGACCTAGGCTCTTT	AACCCACGTCTGAGCTAGTG	174
>SoSSR1779	TACACCGGAAAATTTGGAGA	ACTTCCAAGATCAGCACGAG	226
>SoSSR1780	ATGTGTGCGGATTTGTAGA	GCTCCCGTACTTGAACCTCT	222
>SoSSR1781	CCAAACTGCTCACCAGAAAT	CTCGTCCAAGTAGCTTTCCA	137
>SoSSR1782	TCCTTCGTTTCTGTCTTTGC	CTTGACTCGGTGGTGGTGT	234
>SoSSR1783	TGCTGTTTCTCTGAGCTTT	TCACCGAAGAATGGGATTTA	185
>SoSSR1784	GAGTATGGGAGCAACACCAC	TGTCCAGAAACGTCTTCCAT	248
>SoSSR1785	GAGAATGGAATCGCAACAAG	CCCGATATTGGACTGAACAT	245
>SoSSR1786	TGGGTAGGAATTCATCAAGC	CACTCTCCTGCTTCCATTGT	151
>SoSSR1787	ACGGCGGATTGTTACAC	ATTCTACCGCTAGGGCATT	130
>SoSSR1788	GAGCAATCAGCAGAGTGGAT	AAAAGGTTGGCTATGTGACG	178
>SoSSR1789	GGAGTTTTCAACGCAGACAT	CGACGTCCAGCAGTTATCTT	147
>SoSSR1790	CCATCCTCTGGTCCCTCT	GGTCGAGGAAGGCGTAGTAG	237
>SoSSR1791	GTGGATTTACAGTCCAACAG	GGAATCGATCCGTTGATGTA	185
>SoSSR1792	CGACAGAGGAGAAGAAGTCG	GCTCTGTATCTTCCGTGGTG	165
>SoSSR1793	AGGCCATTTCTTCATCTTCC	ATCCTCGCCCTCTTCTCTTA	198
>SoSSR1794	AGGAGAGAGAGGCATTGGTT	GCGTGAAACACGAGAGAGTT	202
>SoSSR1795	TTTGGCTGATTTTGAAGAGG	TTCTGTGTTGGTGATGATGC	215
>SoSSR1796	ACTTTGGTTGCAAGGATGAA	ACCTCTTTCCTAGCAGCAT	159
>SoSSR1797	GAGAACGACGAAGGATTCAG	GAGGAGGGCATTGTTGATTC	186
>SoSSR1798	CCGGAGGGTTTACTTTCAAT	TCAGTCACACCGTAGGGAAC	190
>SoSSR1799	GTGGAAGAGGAGCTGAGACA	GAGTTCGAGCCATCGACAT	221
>SoSSR1800	TTCTGGCGAAGAAGAAAATG	CATCACTCTGTTCCCAGTCC	146
>SoSSR1801	ATCGCCACTCACCAATAAAA	GGTTGGTCTTTGATCACCAC	125
>SoSSR1802	GTTAGTTCGAGGAAGGCTGA	GATCTTTGGCTGGCTTTTCT	134
>SoSSR1803	ATCCTCAACCTCCACTCCTC	CCATGACCGACAGTTTCTTC	100
>SoSSR1804	GCTTTCATTTTCATCCCCTTT	GGAGGGAGACCAGGTACTGA	247
>SoSSR1805	TAACAATGTCAGCCCTCCAT	AAGTAGCGGAGCGTCTTCTC	241
>SoSSR1806	CCCCATCTCAGACTCAAG	TCTCATCCTCATTTGGCTTC	227
>SoSSR1807	TGGTGTGGTGTGGTGTGTT	CTTGAATCGGTCATTGTGCT	249
>SoSSR1808	GTGTTGGTGTGGTGTGTT	CTTGAATCGGTCATTGTGCT	247
>SoSSR1809	TCAAGTCGTACCAGGCTCTT	CATGGCCTTTGCTCTTCTTA	228
>SoSSR1810	TCAGTTGGATTTGCTCTCAAG	GCACGCACAAGATTAGGTTT	129
>SoSSR1811	CAAAATGCACTCCCAAAACT	TGATTTTGAGGGAAAGGTGA	220
>SoSSR1812	ATCCACCGAAACCAGAGAAT	TTCTTCTCCCCTTCTTCCAC	163
>SoSSR1813	CAAAGCCTGAGGTAGTTGA	AGAACAAGGGCTCTCCAAAT	237
>SoSSR1814	GAGTGCTCAAGGTTTGGATG	ACTCCATTTCCATTCCCATT	178
>SoSSR1815	CAGAAGGGGAGGAGGATAAA	ACCAAACCAACTGATCTCCA	175
>SoSSR1816	GGCATATACGCATTTTCTGG	TAGCCAAGGCAAAGACAAC	151
>SoSSR1817	TCGTCTTCTCGGATTCTTG	GAAACAGAACCCTGCTCTCA	195

>SoSSR1818	GTTGTGGTGCAAGAAAATCC	ACATCACCAACATTCCCATC	191
>SoSSR1819	GTATCCATCGTGGCTTGTGT	CTCTAGCGGGAGGGAATCT	229
>SoSSR1820	GATTCCTCCCGCTAGAGT	GAATCTCTTGTCAAAGCCGAAT	221
>SoSSR1821	GGCCTCTTGTCTCATTGCTA	GTTTCAGCTCTGCCTTCTCA	162
>SoSSR1822	GTGGAACCAATGAAAGTCG	AAAACCGGGTATCTAAAGC	152
>SoSSR1823	GCACCAACTTCGTCTACCC	AAAGGGGAGGAAGAGAGAAA	124
>SoSSR1824	CCACCAACTCCTCATTCACT	TTCAGTTGGGAAGAAATCCA	194
>SoSSR1825	ATGATAACGCCCTCAACAAG	AAGGCTTCTCCTCGTGGTAT	237
>SoSSR1826	AGCTCAACGACCAGAAGAAA	GCTTCTGTCCCATGAAATTG	200
>SoSSR1827	CTGTCTCTGCTCTCCCTGAA	AGCCGGAAATGTTCATAACAC	124
>SoSSR1828	GCAATTGAGAGTCCGAAGAA	TATCCTTGCAGGAAGAGTCG	199
>SoSSR1829	CCTAGATGCAGCACCAGAAT	CCACAGGTCTGAGGAAGAGA	181
>SoSSR1830	TGAATTAGGAGGAGGAGGAAC	AGTTGTGATGCTGATGGTGA	149
>SoSSR1831	TCACCATCAGCATCACAAC	CTGGTTAGGAGGAGGAGGAG	142
>SoSSR1832	TCACCATCAGCATCACAAC	AGTGATCCCAGGTGAAAGTG	188
>SoSSR1833	CAGGATACGAAAAGCAAAT	AAACAAC TAGGGGCAGAGGT	205
>SoSSR1834	GCTGACACTGGACCAACAAT	ACAATATCCCAAGCAAGTGG	240
>SoSSR1835	ACTCGTGTGCCTACAAGGTC	ACAATATCCCAAGCAAGTGG	190
>SoSSR1836	CACCTCAACCGCTCTCTG	CACAATTCCGTCTGGTGACT	170
>SoSSR1837	GAGACGCCTTCAATCACTTC	CAATCTCGCTTGTGTTTCT	182
>SoSSR1838	TCTAACCAAAGGAGCGTCTG	CAAGGGACAAGAAAGACACG	206
>SoSSR1839	GAGCCAGAGGAAAAGACAACA	AGGTAATGGGTGAGCAGATG	241
>SoSSR1840	GAGAAGAAGGCATTGAACGA	ATTTCAAGAGCACCACGTTT	226
>SoSSR1841	TCGGACAAGAATCCAACAAT	AGTTCAGCAGCCATGATCTC	240
>SoSSR1842	CCTCCTCTCCACGAGCTAC	CTTGATCTCGACCTCACCAC	193
>SoSSR1843	GAGACACGGACAAGCAAGAT	ATTCAGCACCCCTTTTCTCT	196
>SoSSR1844	CTTCCACGTCATCAAAATCC	CAAATGCTCCAAAGAGGAGA	190
>SoSSR1845	ATCTGGGGCACATAAAGACA	GATTCGGTGGTTCATCACTT	152
>SoSSR1846	TTCAAACGTCCTGCTCTTTC	TGATCCTGTCCGTTGATTTT	184
>SoSSR1847	AAAACCAGCCTCTCTCCCTA	GCTTGGCAATATCCATCATC	176
>SoSSR1848	ATATCGTGGCCCTCTTCTTC	GCTGTTAGCAGTTGCGTTTC	239
>SoSSR1849	ATTTTCAGCGCACACTCAAT	AAATTC TTGTGGGGCTTCTC	241
>SoSSR1850	CGATGATGCCTTGAGAGAAC	CTGCCAAATCTGAGACTGCT	235
>SoSSR1851	TAAGAATCACCACGGAGAGG	TTCGACAGTGGAGAAGGTGT	214
>SoSSR1852	GAGCCACAAGTTCCATGTCT	ACCCTAGGTTTTGGCTGATT	184
>SoSSR1853	TCACGAACGATAACAGCTCA	AATCCGCGTAGTCCCTTTAC	206
>SoSSR1854	CAGATATCAAATGGCAACCAC	ACAGTGAGAGAGCAGCATCC	119
>SoSSR1855	AGACACATGCAGGCTGAGTT	GCGCATTTTCTCAAACACTT	161
>SoSSR1856	GGAATCTGATATGGGGAAGG	CACGTT CAGGAAGTACCACA	218
>SoSSR1857	TGCCATGAAAATCTCCTTGT	GAAGTTTGGCTCTTGATTGG	154
>SoSSR1858	GACTCCACCGTCACACAATC	GGTGGTGAGGAAGAAAAGAAG	136
>SoSSR1859	AGTTGGGAGACATGGTGGTA	GCTCCGTTGCAGCATATTTA	247
>SoSSR1860	TACCAAGGGAAACACAATGG	AGGGTATTTGGAGAGGATGG	135
>SoSSR1861	AGCAGGACTTCTTGGGACTT	CCCAGTCCTACCTCCTCAGT	140
>SoSSR1862	GATGCTGGAACAAGGATTTG	TACTGTTTGCCAGCCTCTTT	243
>SoSSR1863	CAAGTGCTGTTGGGCATAG	GCGTCTCCTTTTCGTAATCA	160
>SoSSR1864	CGACGCCAAATACCATAAAC	TCAAGACAATGGAAAGCACA	222
>SoSSR1865	AAGTGCAAGGGATGTGATGT	CCCAAGATTCCCAACAAGTA	105

>SoSSR1866	GTTGATCCAAACCAGTGGAG	TCACTGCTGCTACATCCAGA	223
>SoSSR1867	CATGGAGGTTCAATCCAGAC	AGCTATCCTCGTGTTCATCG	169
>SoSSR1868	CGGGATGAGGAATTTAGGT	CTTCTCCCGTCTCTTCTTCC	168
>SoSSR1869	CGCACAGCTCAATTACTCT	ACCTCTGTTCTGCATCTTGG	231
>SoSSR1870	CCACTTCGTATTTGCTTGCT	TGCTGAGAAAGGAGATGGTC	107
>SoSSR1871	CAGCCATAGAGCACCAGAGT	AAACCACAACCTCCTTCTCC	192
>SoSSR1872	GTAGAATGTACGGCGTCGAG	GAGGACTTCTTGCTGATTGG	172
>SoSSR1873	CCCAATCAGCAAGAAGTCC	CCTCGAAACCTAAACCTA	229
>SoSSR1874	TTTTGAATGAGCCGAAGAAG	AAACGCTCCATATCCTTTCC	201
>SoSSR1875	GAAGGAGAACAGCTGATGGA	TCCCCTATCTGAAACATTGC	179
>SoSSR1876	TCCATGTGTACCTGTGCTTG	TACGCGTTACCATGAGGTTT	159
>SoSSR1877	GCTCCGATACAGTGGAGAAA	ATTGGTGGGTGACAGAGAGA	223
>SoSSR1878	CAAGCCAGAAGAAGAGACCA	GTTTTGGTGGGTCCGATT	236
>SoSSR1879	CACAGGAAAAGCCCAAATC	CCCATAGTAACCGTCGTAGC	209
>SoSSR1880	GGAGGAGGAGAAGAAGGATG	CTGAACAATTGAGGGGCATA	165
>SoSSR1881	CGAATAAGGCCGACATAGAA	CCTTTGCAGCTTCACTTTGT	231
>SoSSR1882	CTCTTCTCCATCTGCAAAA	AGCCATCCGAGTACCTGTAA	104
>SoSSR1883	TGAAGCTCCAAGTTCACGAC	TCGACGCTTTTATCTCATCC	186
>SoSSR1884	GAAAGAGTGGGGAAGAGGAG	GATATCTTCCTTGCCGTTT	233
>SoSSR1885	CAAAGAAAGAGGCATTCAGC	AGACAAGCAGGGTGAACAAG	248
>SoSSR1886	GGTTGTAGTTTGGCAGCAGT	CCTAGTCCTCACCACCAGAA	147
>SoSSR1887	GTACCAGAGTTTCGCGTGT	TAAATCCCTGACCACAGCTC	235
>SoSSR1888	TTGGGAGCTGAAATCTTTTG	TGGTGATGCTGCTAAAACCTG	106
>SoSSR1889	ATCCACTGAAAACCTCCTC	AAACGGCCAGTTATTCTTGA	229
>SoSSR1890	AAATATGGGCCTTCATCCTC	ATGCTCCGGAATCATTAAACA	224
>SoSSR1891	CAGTTGATACGTCCGGTCTC	ATCTCTTGGTCAGCATGGAA	229
>SoSSR1892	TGTTGGGTTTCATTTGGATCA	TGGAGAGGAGAATGGAGATG	189
>SoSSR1893	CATCTCCATTCTCCTCTCCA	ATGAAAGGGGTGATCTGGTT	192
>SoSSR1894	TCTCCTCGAATAACCACAA	GTGGCCAGAAGAAGACGAT	186
>SoSSR1895	TGCGAACAATCTGCTTATGA	ATGCTCCAAACGTCGTCTC	216
>SoSSR1896	TGGATTCTCCTTCAACATCC	GCAGCAAACCATAACCATAGG	193
>SoSSR1897	TGAAGCAATTCGCAGATACA	GAGACGGCACTAAGACAGGA	183
>SoSSR1898	GCCACATTCGGCTAGAAATA	CCGATCAGGATAGTCACTGC	226
>SoSSR1899	AGATGCTGCAACCAGTAGGA	AGTCACGGATGGAGTTGAAG	182
>SoSSR1900	CTTCAACTCCATCCGTGACT	CTGAAGGTTGATGAGGGAAC	210
>SoSSR1901	ATGGCTTCCTCTGCTTTGA	GTTGTAGGTCTGGGGTTCTT	135
>SoSSR1902	TAGGGTTCGAGGACTACGTG	CATACCGTACATCCCCTCAT	173
>SoSSR1903	GAGGGGATGTACGGTATGAA	GACCCGGATCCTATATGGTAA	125
>SoSSR1904	AAGAGCGCCAAAGAAGTCTC	TGTTTCAGCTCCTCCGTACTC	235
>SoSSR1905	GAGAGGTTGGCTGTTGCTTA	TCACCAATAGTCCCCACAA	204
>SoSSR1906	GGTGGAAACCTCACTGTCAAC	CCGGAGAAACATAACAAAAC	171
>SoSSR1907	CAGCGATTGTGATTCTTCT	GCTAGCTCGTCCAAATCG	150
>SoSSR1908	TGAGCAGACCTACACGTTCA	CTTCCCCTTCTCCTTACAT	100
>SoSSR1909	CTTGAATGCGAGGTTGAGTT	GCCAAGGATGCTCATGTATT	129
>SoSSR1910	TTTGAACGTACACCGAGTTG	TTCCTCCCCAAAATTTCTTC	200
>SoSSR1911	AAGATCCTGCCTGGGCTA	GGAGGATTTGGAGCTTCTTC	144
>SoSSR1912	GACGTTGAAATTGGAAATGC	GGACGATCATCTGGATTTCAT	149
>SoSSR1913	ATGAAGAAGCGAAGAAAGCA	CTTCAGGTTTGCATGAGGTT	164

>SoSSR1914	CGTCTCTCTCCAGCTCATTC	TCCAATTCATCTTCCTCAGC	240
>SoSSR1915	GTTTCCTCCGTTTCTGCTTC	ACCTGCTTCCTCTCATTGC	186
>SoSSR1916	CGAGTTTGAGGAAGACGAAA	TTCAGCCACAACAGTGAAAG	148
>SoSSR1917	AGAACCAAACCTTCCAAGA	CCTGTGAAGCTGCCACTATT	248
>SoSSR1918	GACTACGCCAAGTTCGTGAT	AACTGGGAGTGCCTTAGCTT	153
>SoSSR1919	CAAGAAAGAGAGCAGGCAAG	CACCAGACAACCTACCGTCT	115
>SoSSR1920	CAAGAAAGAGAGCAGGCAAG	ACCATCTGCGGTTGTTATTC	223
>SoSSR1921	AATTTTCGAGTGCTGGTGTGT	TGCCTTCTCTGCTTCTCAAC	225
>SoSSR1922	GAGTGGCTCTCTTGGCTCA	TGGTGGTGCACATGAAATAG	125
>SoSSR1923	GCAGCTTTCAAACTCGTAGC	GAAACACAAACAGCCCTCTG	244
>SoSSR1924	GTTCTCCGACGACGACATAA	CTCCTCGTCCACAACTCAT	136
>SoSSR1925	GCAGGATGTCGAGAGCAG	GAACATGACGACGACAAACA	139
>SoSSR1926	GATGCTAGTCGAGCTCTTGC	CTTCTGCCATCTTTGCAACT	226
>SoSSR1927	GGCAACAAATAGCGAGAAAA	CTTCATACCGGCGACTCTAA	191
>SoSSR1928	ACATGACAGAGAGCCCATGT	CGTCTTCGCGATTATCTCTC	244
>SoSSR1929	TCCAATTTTCGTAGCACAACA	CATGCTGCAGGTAATTAGGG	223
>SoSSR1930	GTGGTGGTGAAGATTTTGAG	CCATCAACCTCAAGAGGAAA	165
>SoSSR1931	CAACGCCACAACCACTACT	ATCGCTAATCGTGCAGTCTT	115
>SoSSR1932	ACGAGCAACAATCACCAGAT	TAAAGACTGCGGTTTGGGAA	156
>SoSSR1933	AGCAATCATTAGGCGTTTTG	TCTGTTTCTTCGCCATATCC	245
>SoSSR1934	TGGACTCTGACCCTGAAAAC	AAGTTCCCAAGTCCCATAG	231
>SoSSR1935	ATGAGCAAAGTGTTGGCAGT	GCGAGGAGCATTAGAACGTA	155
>SoSSR1936	TTCGGGGATAATGAAGTCTG	GATTTGCAGTTCTCGTGCTT	221
>SoSSR1937	TGACTAATGGTTCCTCGAA	CTCGATGCAAATACCCAATC	120
>SoSSR1938	CCCCTCTGGTTCATCTCTCT	GTGAACACCTTCACCGTAGC	194
>SoSSR1939	CACTGCAATCATCCAGTACG	AGCTTCTTCGTGACATCCAG	171
>SoSSR1940	GCAATTCTTCCCTCTTCAACA	ATGAGATCAAGACCGCACAT	172
>SoSSR1941	CTTTCTTGAAGGACTGAAGG	GAATCGCGAAAACAATCATC	122
>SoSSR1942	GAATGAAGGCTTGCGAAATA	TCGCCACACTATATCACCT	221
>SoSSR1943	CCTACAAAATCCAGGCTCAA	GAACGAGTTGACGCAGAGAG	213
>SoSSR1944	CCGAGATCAGCAATCTCAAC	CGATCCGTCATTTGATTCTT	247
>SoSSR1945	CACCAAACCTACCCGATTCC	CGTGGTTGATGGTGGATTT	224
>SoSSR1946	CTCATCAGAGCTCGCTTCC	CAGATGAAAAGGAGCGGATA	178
>SoSSR1947	CAGCAGTTGGGTATTGCTCT	CATTGCAACGGGAAGTTAGT	146
>SoSSR1948	ATGGTTTCTTCCCTCCAAAG	GGAAACCTAAATCGCCTCTC	230
>SoSSR1949	CCCTCTCCCTGAAAGAACTC	TGCAAGTCCAACCTCCAATCT	121
>SoSSR1950	CCATTTGTTGGACGTAAAGG	AGGTTC TTGTTTGTGCTG	186
>SoSSR1951	CTGACAGTCGGGAAAGAGAA	GCTAATAGGCACCGTCTTCC	141
>SoSSR1952	GGTGACAGGGCTTACAGAGA	TCATAGGCAGCTCTGGTTTC	176
>SoSSR1953	CAAAGAAGAAAACCTCTCTCC	GGTGACTGGGAATTGATTTG	108
>SoSSR1954	TACCCCTCTGCATTTAGCAA	GGATCCGAAAGGCTAGGATA	245
>SoSSR1955	TCTCTCGCTCATCTCCATTC	TGTTGTTCGGAATCTCGACT	137
>SoSSR1956	GGATTCCCCACAATTCTCTT	CACAGCTCATCCTTCACGTC	224
>SoSSR1957	TCTACCAGACCTTGCAGACC	GATGACTCGGAGATTGGATG	238
>SoSSR1958	GATGTTGGGAGAAACATTGC	AACACGGCCTTAAATGATGA	214
>SoSSR1959	ATCGTCGAAAGGGAAGAAAT	GAATCATCTGCACCACCTTC	201
>SoSSR1960	TCGAATGAACATCGAAAAGC	GCTTACCTCCACAGCAGAAA	191
>SoSSR1961	TTAAGCTATCAAGCGCCAAT	ATCGTCGTCTAATCCAACCA	179

>SoSSR1962	GCGACCTGCTTACTCACCTA	TTTTACCGATTCCACCCATA	100
>SoSSR1963	CATCGAATTCCTCCGAAGAAAT	TTGGAAACACACCTGAACCT	213
>SoSSR1964	AGCCGAATTCCTTTTCAGAG	ATGCTTCCTCATAACCCACA	186
>SoSSR1965	TCTGAAGCAAGAGGATAGGG	TGCTGATTGTATTGCACAGG	207
>SoSSR1966	CGTTGTGCTCAAGAACTGTG	CAGAGCCTTCAGCCTGTGT	161
>SoSSR1967	AGAGGCCGACTTCACTTTCT	ACTTGAATACAGAGCCTTCAGC	233
>SoSSR1968	CTAACGCAGACGCCTTTC	CTCCATCAACTCCTCCATGA	243
>SoSSR1969	GAGGTGGATGAAGAGTTGGA	CACCAACATCAGAATCAGCA	208
>SoSSR1970	TCGCGAATTAGGTTTGTTC	AGACGGGCTAGATGAGGAAT	159
>SoSSR1971	ACTACCTCAATTGCCCTTCC	TTTCGAGAGAGACGATGAGC	141
>SoSSR1972	TCACATCAAAGACGGAGTGA	AAGCCCTGTCTCTTTTGCTT	172
>SoSSR1973	GACTATTGGGAGCATGAGGA	CTACTCCGGTGTGTTTTGCT	116
>SoSSR1974	GAGGAGGATGAAGAGGAAGC	ATCGACACCTCCAGCGACT	153
>SoSSR1975	TGGAGAGCATGTACAGAAA	TTCCTCCTTCTCTGCCTCTT	203
>SoSSR1976	GCCAGCTGAAGAGCTATCTG	GGAAATAATCCGAGGCAAGT	220
>SoSSR1977	CAAGAAACGTAAAGCCTTGG	GATCTTCGGTGACTCCTTGA	225
>SoSSR1978	TGTCACAACCTTCCCTCCTA	TGTCAGCCTCTGATCTCTCA	233
>SoSSR1979	TCTTCACCTTCCACTTCACC	GAGGACTGGATGACGAGTTG	250
>SoSSR1980	CCTCGCTTCTCTACAACGTC	TCTTCCACGTCGGTTAATGT	172
>SoSSR1981	TTCTGCCTCTTCTCTTGICC	GGGAGAGGTGTTGGGATG	135
>SoSSR1982	TGCTGAAATGGACTGTTGTG	CATCCCTGAACCCAATTACA	179
>SoSSR1983	AAACGGAGAGCCTTCCAC	ACATGTCCTTGGCCTGTAAA	249
>SoSSR1984	CTGAGAAAGCCGTCCATTAG	ACCTGAGCAGTGAGTCCAAG	218
>SoSSR1985	CCCCAAGAAGATCTCCAAAT	CCCTCCATTTAGCACAGAGA	242
>SoSSR1986	TAGACTGGTTCCGATCAGC	CCTTGCCATTCTTGAATTTG	177
>SoSSR1987	GACACAACCGATTTACCTC	ACCTCGTCGTCTTCTCTCT	119
>SoSSR1988	AGAGGTGGATCTTTCCCAAG	GCTGGTGAGATGGGCAAC	197
>SoSSR1989	ATATCGTCGGCAATCAAGAG	TCCCAACTCCTGTGGACTAA	163
>SoSSR1990	GAGGTGGCTCTCGTTTGTA	ATCTGCATCTCCAGCTCATC	194
>SoSSR1991	CAGCAGATATCGCCTAGCAT	CTCGATTCTCTGTCTTTCCA	173
>SoSSR1992	CGAACCTCTGAGCACATCTT	CCTGAGGGAAAATGCAACTA	222
>SoSSR1993	TCCCAATTTCTTAGGCACAA	GGATCTAGGGAGGTGGTGAT	245
>SoSSR1994	ATCACCACTCCCTAGATCC	CTCGAGCGAGTTTTGTCC	248
>SoSSR1995	ATAGCCCACCATACCATGA	CATCCATGAGCATGATTTGA	229
>SoSSR1996	AATGTCACGTCACCGAAATC	CCAAATCGCTCCAGAAACTA	193
>SoSSR1997	AGGAAGGGGAGATATTGGTG	AGCTGCCTTATTGCCTCTCT	139
>SoSSR1998	CAGAAGAAATCCGAAGATGG	AGCAGTTTTGCCCCGATAGTT	207
>SoSSR1999	CCCTGTCACTGCATATTTCC	GAAAGATGGGACCGCTTTAT	184
>SoSSR2000	CACGACACTTCTCCATTCC	TCTTCCACAAGCCTTACACC	191
>SoSSR2001	ACATTGTCCCAAATGCAGAT	ACTCCCTCTCCACATCCTTC	157
>SoSSR2002	GAAGGTGGTGAGTTTGATGG	CCTTCTTCTTCTTCCCTTG	223
>SoSSR2003	AAACCAAGCTGCACAAGAAG	GACGCCGCTACTTTCTATCA	242
>SoSSR2004	CTTCTGATCTCCACCACCAC	CGAACATTCTCGAACCTTGT	187
>SoSSR2005	AGCACCTAAGCCTGGAAACT	CATTTTGGCACTGAAGAAGG	206
>SoSSR2006	CGGTTCTTCTGCTCAACAAT	AGGATGCCCTAGGAATTGAC	248
>SoSSR2007	TCAGGCAACTACCGTGCTAT	GACTAGGAGGTGGTGGAGGT	215
>SoSSR2008	ACCACCTTCCAACAAATGAA	GGTCGGTCTTGGTCTCTCTT	241
>SoSSR2009	AGAAGATCCGCGGTCCAG	TCGGTGCCGATAGAATAGAG	244

>SoSSR2010	GCAGCTTTCAGTTTCCTCTG	ATTACCCGCATCGAGAGTTT	241
>SoSSR2011	CACAACCTCTCGACCACCAA	AATTGGAGGGGATTTCGATAG	249
>SoSSR2012	GAGAAGTGGGACGTGGAAT	GAATCCTCCTCGATGTCGTA	203
>SoSSR2013	GAAAGAGAAAAGAGGGGATGC	CTCCCTATCTCCATGCCTTT	201
>SoSSR2014	GATCAAGGCCGATGAAATAC	GCACTGCTTTGTAGGACGAT	177
>SoSSR2015	CTGCAAGAGGAGAATGATGG	CAGTATGGTAACGCCAAGGT	214
>SoSSR2016	ACATGATGGCGAAGAAGAAG	GCAAATCTCGATGCCAGTAT	147
>SoSSR2017	GTAAGAGAAAAGCGCGACAAG	AGCTCTGCATCTCATCTTCG	237
>SoSSR2018	CATCGAGCGAACTGTCTTTT	CGAATTTCTCGACACTCCAT	170
>SoSSR2019	TTCGCCAAGAGCAATATAGG	ATATGAGCATGCCAAAGAGC	160
>SoSSR2020	TTGGAGTGAACCCAAAAGAG	TCCTCTGTTCCAGCAGTTTC	232
>SoSSR2021	TGGTGTTTAAAGCCAGCAAC	CCGATAAATCCAAGTCAAGG	234
>SoSSR2022	GCTAACAGCCACCACCATAA	CTCCTTCTCCTTGTTTCGTCA	195
>SoSSR2023	AATTGGAGATGAACGAGCAG	CGAGATCATCCTTTTCTCTGA	198
>SoSSR2024	AGTTGAAGTCCCATCCATCA	TTTTATCATGGCGAGAGAGG	214
>SoSSR2025	CTCTCCATGCACATCATTCA	GCATCCCACTTATCATCCAT	185
>SoSSR2026	TTGATCTTCGTGGAGGAGTT	CCTTCATCAAGCAGAAGTGC	184
>SoSSR2027	GATCACTTCCCTTGTGCAAC	CAACCCAGGAAATCTCACAC	204
>SoSSR2028	ATCCGAGAAGCCGTCAGT	TTGGGGTAAATTAGCAAACG	240
>SoSSR2029	ACCTCCTCCTCCTCCTCAT	TTGGGGTAAATTAGCAAACG	191
>SoSSR2030	ATATGCGCCAAGAAACAGAG	AACCTGCGAGAAATCTAGGG	114
>SoSSR2031	AAACCATCCTCGGATTTCTC	TCAGAACATCCCCATTCCTA	154
>SoSSR2032	TCTCCTACTTTCGCGTCAAC	ACGAGAATCCCCAAAATCTC	199
>SoSSR2033	AAGCCCRACTACGAAGAAAT	TCGTCTTCTTGTTCCTCTTA	211
>SoSSR2034	AAGGCTCCCCAAATTTACAC	TGGAGGGGTTAATGAGACAA	137
>SoSSR2035	TGCAATTCAAAGCCCTAGAC	CACCGACGAGAAAGAGGAA	203
>SoSSR2036	CGTTTATGGGTTGTTTCTCG	CTCAAAACCTCATGCCTCAC	202
>SoSSR2037	GATGCGGATAGTGGTGAGAT	TCGGACTCCACATAGATCGT	228
>SoSSR2038	AACCCCTCAAAGCTCAGAAT	TGGCTGCCCTAATATCAGAC	226
>SoSSR2039	GAGGAAAGGGCTAGAGATGG	CTATCCCGGGTAATCTTGCT	164
>SoSSR2040	TCCATCCATTTTCGTTTCTA	CGGAGCTGTTGTGAGTGAG	138
>SoSSR2041	TTGGTCAATAGTGGGTCCAG	CTCTCCAGAGTCGAGTTCCA	215
>SoSSR2042	AGTGATGACTCGGTGAGTGG	TTGAACCCTGCTTCTTCAAC	232
>SoSSR2043	AATCTCTCCCCTCTTCCAAA	TTGTTCCGATTTTGAGTTCC	161
>SoSSR2044	GACAGACCAATGGGAATCAG	TGTACAGCTGCCAATCTTCA	162
>SoSSR2045	CAATCTCAACAAATCGCACA	TAAACCTCAACCGTTTTCCA	166
>SoSSR2046	TCTCCACTGACGAGCACTCT	CAAGTAAGGGTCGTCGGATT	233
>SoSSR2047	AGAAATTGGTCAACCCCTTCC	CCATTGTCGACTTGGTTCTC	139
>SoSSR2048	ATCTACATCCGCCACTGCTA	GTCTCGTCGTCCTCGAACT	227
>SoSSR2049	CAATCACGTCAAGATCCACA	ATTAGGGCCGTACATCAACA	238
>SoSSR2050	CCATGCACCTCTTCTTTGTC	CTTTTGTGGGTTTGTGATCG	174
>SoSSR2051	GGCGAATAGATCTTCTCTCT	ATGAAGTCGTCGGATTTGG	204
>SoSSR2052	CGGGAAGGAGTATACGTTCA	GGGGTTTTGGCGTTTATATC	219
>SoSSR2053	CGCAATCCTAGCAAGAAATC	CGAGCACGTTCTTCTTCATC	130
>SoSSR2054	GCTTGATGACCCAGAGGATA	AGCACGAATTTCCATCACAT	181
>SoSSR2055	TGAGATTTCCGGAGTGTGTGA	GCTTCCGTTACTTCTTCGTG	249
>SoSSR2056	CTCCTGGGCAACATCTACAG	CCGTTATCGTCCCGAACCC	241
>SoSSR2057	GGGTCATAGCCTTCTCTG	GGGGAAGGATCAGATTCAGT	213

>SoSSR2058	AGGATTATTAGGGGCTGGTG	TGGCAACAATCAGTAATAGCAG	213
>SoSSR2059	AGACCACTTCCGGTTGAGAT	GAATTTGGCTTGTCCATCAG	228
>SoSSR2060	GCATAAGCTTGGATGACGAT	AGATGGAGTAGATGCGCTTG	156
>SoSSR2061	AGTGCTCAAGATTCGACGTT	TTCTTGGACGGAGATTGAAA	101
>SoSSR2062	GGTGAGGGTAACGAGAAGGT	GAGCTCAAATTTTCCAAGCA	240
>SoSSR2063	GGATGTGTCGTCGTATGTGA	TCCAAC TACTCCCAGCAATC	125
>SoSSR2064	ATGGCGATACACGAAAAGAG	TTCTAGGGCGATTGTTGAAG	234
>SoSSR2065	CCAGAGACGTGGTTCAATTC	ATCAT TCTCCACCAGCAAGA	208
>SoSSR2066	CACCAACACGTTGCAGAATA	ATGCCGAGAAAATCCCTAGT	192
>SoSSR2067	ACCTACCCTCCATCAACACA	GCGATCCTTACTGGAGGACT	133
>SoSSR2068	CACAATCACCTTCGTCAACA	GGAGAAGACGTGGATGAGC	205
>SoSSR2069	CGCTTCTTGTTTCCCTTAC	CTAATGCTCCCATCTCTCCTC	246
>SoSSR2070	CACGTAATTGAACGGCCTAC	GGCGTATCGTCAATAACCAC	248
>SoSSR2071	ATCAGATCGCAACGAAGAAG	GGCGTATCGTCAATAACCAC	228
>SoSSR2072	CATTCGCTGCTAGTCCCATT	GGATCTCGATGAGGTTGATG	184
>SoSSR2073	GTTCTCTCCTGCTCCTACCC	GAAA ACTGCGTGCTTCACTT	102
>SoSSR2074	TGCTATGTCTGGTGGGATTT	CAGTGTGAGGCTCTGGTTCT	211
>SoSSR2075	GCTCATCATGGTGGTCTTCT	TTTCCACTTCCCCTTCTTCT	133
>SoSSR2076	GACCAAGCTCGAGAATGAGA	TTATT CGATGAGTGTGCTG	102
>SoSSR2077	GAAGGCGAGGTTAAGGAGTT	TCTAGACGAGGAAGCAATGG	210
>SoSSR2078	AAGGAATACGCATGAATGA	AACCCCATAGGATTCGAAAG	177
>SoSSR2079	CCCTCTCCTTCTTCTTTTGC	ATCGGAAATCTCCTGCTTCT	110
>SoSSR2080	AGGAGAATGAGAGGCTGAGG	TGGATTTCTTGGTAGCCTTG	165
>SoSSR2081	AAA ACTACCTTTCCTGCTGCT	GATCTTTGCAGGTGCTGTTT	116
>SoSSR2082	TGGTGTCCCAAGTTTCTGAT	TCCTCGATTTTCCATTCTTG	109
>SoSSR2083	GTGAGAACGACCACCATCTC	ACTGGAACCTGTCCAGCTC	214
>SoSSR2084	CTTGAATCCGAGAAGCTGAA	AAATGCGTGTTGCATCTCTT	212
>SoSSR2085	AGCTCCCTCTGATCTCCAAT	AAGTTTCGCCATCTCCTTCT	208
>SoSSR2086	CATTTGAATGGGAAGGACAG	AAAATGGATACCCCAAAGC	197
>SoSSR2087	AGCTGCTCACGTGTTACCTC	GGCGATGTGAAGGAGTAATG	175
>SoSSR2088	CAACAGGAGACTGCATCACA	ATGCTGCTCCAGA ACTTGAC	195
>SoSSR2089	TGTTATTGCATCGATCCTCA	AATATTAGACGGCGGTGGA	211
>SoSSR2090	GATTAAGAACGCGATTGTGG	TCACTGGAGCTTTTCTGCTC	183
>SoSSR2091	ATGCCAACCAACAATTCTCT	GGGCTCTGCTCTCTCTTTT	164
>SoSSR2092	CCCTAAACATTTCCCATTTC	CCTTCTTCGGATCTTTTGGT	134
>SoSSR2093	CTCAATCCACCTCAGCTTTG	GGGATACCCATTTGCATCTA	218
>SoSSR2094	GCAGACCTTCTCCTTCATCA	GTCGAATTATTGTCGCATCC	201
>SoSSR2095	CCTTCAGGTCTCTCGTGGT	CACAAACACTCTCCCGAAAT	181
>SoSSR2096	CCTCGTTGATTTCTCGATTG	GGAAGCAAAGAGCTGTACCA	172
>SoSSR2097	GTCTCTGGTTTCTCCCGTTT	GAGCCTCACCATGCTCTTTA	113
>SoSSR2098	GCCTCAAAGCTATCAGTCCA	CTCTGCGTTTGGCATAAACT	152
>SoSSR2099	GAGGGAGTGGAGGATCAGAG	AGATGAAGGGCTGGAAAATC	214
>SoSSR2100	GACGAAGAGGCAATGTTTAC	ACACGCACTCAACACACATT	168
>SoSSR2101	CTGCATTCCAAGAACGAACT	CATTGAGACGAAAGAGATGGA	207
>SoSSR2102	CCTCTACTACCCCAACCAC	CAAGACGCCATTGATAGACC	235
>SoSSR2103	GTCTATCAATGGCGTCTTGG	CGAATCTTCTCGGAAGGAAC	180
>SoSSR2104	AGGTTGTACCTCACGCTTTG	GATTGTCGATGTTGGGAGAG	243
>SoSSR2105	CGCCAAA ACTCTCTCAAAAC	ACGAGCTCAAGAATGTCGTC	249

>SoSSR2106	GGTCTTCCTCCTCCACCTC	CCGTTTAGGGTTAGGTTTGG	120
>SoSSR2107	CTCTCTCCTCCACCCTTCTC	AAACCTTCTTCGAGTGCAAA	151
>SoSSR2108	ATGACGATGACGATGACGAT	TCCAGTAGAACCACCTGCAT	183
>SoSSR2109	TGCTGTTGATAGCTGTTGGA	CAAGCACTCTGAACACGTTG	216
>SoSSR2110	TTATAAATGCAACCGGAAGG	TCCCGCACTATCAGTTATCC	241
>SoSSR2111	CTGGTTCAGTTCGACCTAA	TCAACGCCTCCCACACAT	142
>SoSSR2112	GCAAATTCCTCTTCCTCACC	AAGCAGTTGGCGTAGGAGTT	130
>SoSSR2113	GAGTTTGGTTGGGTTTTGTG	CAGAAGCAGCAAGAAGATCG	105
>SoSSR2114	GCAGAGTCAAGCAGAGGAAG	AACGAGTATGGAAGCGTGAG	219
>SoSSR2115	TACAAGCTACGAACCCGAAT	AAGTACACCTCCAACGACGA	232
>SoSSR2116	GCCCTTATTCCTCTTTTGCT	TTGCCATAACCAATTCCACT	176
>SoSSR2117	AACATGCTTCCACTCCTCCT	GGCGGGTTTTAAGATGTTGT	117
>SoSSR2118	ACCAGGTGAATGGCAGAATA	GAAACCACCAAATGACTGC	175
>SoSSR2119	CCTACGTCTACCACGCTCTC	CCCTTCTCAACGTGATCAA	202
>SoSSR2120	TGGCCACTGCTCTTCTCTAT	TGTACAACGGTCTTGAAACG	177
>SoSSR2121	TATGGGCCTCTTTGACTCTG	GAGACGGTTTCTCGTGCTAA	240
>SoSSR2122	AATCTTCGGAGACGACGAG	AGTACCACGGCACGTTTATG	242
>SoSSR2123	AATGAAGACTGCGCGTGTAT	TTCACCCCTCTCCTAACCTT	182
>SoSSR2124	GCGAACTCCATAACACGTCTA	CCTTGTACAGCGACTTCACC	190
>SoSSR2125	CAGAGGCTAGTTTTGGGACA	CAGTAGTCAAGCTGGCAACA	168

---

## 6. KAYNAKLAR

- Akhondzadeh, S., Noroozian, M., Mohammadi, M., Ohadinia, S., Jamshidi, A. H., & Khani, M. (2003). Salvia officinalis extract in the treatment of patients with mild to moderate Alzheimer's disease: a double blind, randomized and placebo-controlled trial. *Journal of clinical pharmacy and therapeutics*, 28(1), 53-59. <https://doi.org/10.1046/J.1365-2710.2003.00463.X>
- Akkawi, R., Valente, A. L., & Badawy, S. Z. A. (2012). Large Mesonephric Cyst with Acute Adnexal Torsion in a Teenage Girl. *Journal of Pediatric and Adolescent Gynecology*, 25(6), e143-e145. <https://doi.org/10.1016/J.JPAG.2012.06.009>
- Ali, M., Li, P., She, G., Chen, D., Wan, X., & Zhao, J. (2017). Transcriptome and metabolite analyses reveal the complex metabolic genes involved in volatile terpenoid biosynthesis in garden sage (*Salvia officinalis*). *Scientific Reports 2017 7:1*, 7(1), 1-21. <https://doi.org/10.1038/s41598-017-15478-3>
- Alic, A. S., Ruzafa, D., Dopazo, J., & Blanquer, I. (2016). Objective review of de novo stand-alone error correction methods for NGS data. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 6(2), 111-146. <https://doi.org/10.1002/WCMS.1239>
- Allam, A., Kalnis, P., & Solovyev, V. (2015). Karect: accurate correction of substitution, insertion and deletion errors for next-generation sequencing data. *Bioinformatics (Oxford, England)*, 31(21), 3421-3428. <https://doi.org/10.1093/BIOINFORMATICS/BTV415>
- Alonge, M., Lebeigle, L., Kirsche, M., Jenike, K., Ou, S., Aganezov, S., Wang, X., Lippman, Z. B., Schatz, M. C., & Soyk, S. (2022). Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biology*, 23(1), 1-19. <https://doi.org/10.1186/S13059-022-02823-7/FIGURES/2>
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403-410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Altshuler, D. L., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., Collins, F. S., De La Vega, F. M., Donnelly, P., Egholm, M., Flicek, P., Gabriel, S. B., Gibbs, R. A., Knoppers, B. M., Lander, E. S., Lehrach, H., Mardis, E. R., McVean, G. A., Nickerson, D. A., ... Peterson, J. L. (2010). A map of human genome variation from population-scale sequencing. *Nature 2010 467:7319*, 467(7319), 1061-1073. <https://doi.org/10.1038/nature09534>
- Azevedo, M. I., Pereira, A. F., Nogueira, R. B., Rolim, F. E., Brito, G. A. C., Wong, D. V. T., Lima-Júnior, R. C. P., de Albuquerque Ribeiro, R., & Vale, M. L. (2013). The antioxidant effects of the flavonoids rutin and quercetin inhibit oxaliplatin-induced chronic painful peripheral neuropathy. *Molecular pain*, 9(1). <https://doi.org/10.1186/1744-8069-9-53>
- Badiee, P., Reza Nasirzadeh, A., & Motaffaf, M. (2012). Comparison of Salvia officinalis L. essential oil and antifungal agents against candida species. *journal of Pharmaceutical Technology and Drug Research*, 1(1), 7. <https://doi.org/10.7243/2050-120X-1-7>
- Bao, Z., & Eddy, S. R. (2002). Automated De Novo Identification of Repeat Sequence Families in Sequenced Genomes. *Genome Research*, 12(8), 1269-1276. <https://doi.org/10.1101/GR.88502>

- Barra, V., & Fachinetti, D. (2018). The dark side of centromeres: types, causes and consequences of structural abnormalities implicating centromeric DNA. *Nature Communications* 2018 9:1, 9(1), 1-17. <https://doi.org/10.1038/s41467-018-06545-y>
- Bateman, A. (2019). UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*, 47(D1), D506-D515. <https://doi.org/10.1093/NAR/GKY1049>
- Behradmanesh, S., Derees, F., & Rafieian-kopaei, M. (2013). Effect of *Salvia officinalis* on diabetic patients. *Journal of Renal Injury Prevention*, 2(2), 51. <https://doi.org/10.12861/JRIP.2013.18>
- Berbers, B., Saltykova, A., Garcia-Graells, C., Philipp, P., Arella, F., Marchal, K., Winand, R., Vanneste, K., Roosens, N. H. C., & De Keersmaecker, S. C. J. (2020). Combining short and long read sequencing to characterize antimicrobial resistance genes on plasmids applied to an unauthorized genetically modified *Bacillus*. *Scientific Reports*, 10(1). <https://doi.org/10.1038/S41598-020-61158-0>
- Berlin, K., Koren, S., Chin, C. S., Drake, J. P., Landolin, J. M., & Phillippy, A. M. (2015). Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. *Nature Biotechnology* 2015 33:6, 33(6), 623-630. <https://doi.org/10.1038/nbt.3238>
- Birol, I., Raymond, A., Jackman, S. D., Pleasance, S., Coope, R., Taylor, G. A., Yuen, M. M. Saint, Keeling, C. I., Brand, D., Vandervalk, B. P., Kirk, H., Pandoh, P., Moore, R. A., Zhao, Y., Mungall, A. J., Jaquish, B., Yanchuk, A., Ritland, C., Boyle, B., ... Jones, S. J. M. (2013). Assembling the 20 Gb white spruce (*Picea glauca*) genome from whole-genome shotgun sequencing data. *Bioinformatics (Oxford, England)*, 29(12), 1492-1497. <https://doi.org/10.1093/BIOINFORMATICS/BTT178>
- Blake, J. A., Christie, K. R., Dolan, M. E., Drabkin, H. J., Hill, D. P., Ni, L., Sitnikov, D., Burgess, S., Buza, T., Gresham, C., McCarthy, F., Pillai, L., Wang, H., Carbon, S., Dietze, H., Lewis, S. E., Mungall, C. J., Munoz-Torres, M. C., Feuermann, M., ... Westerfeld, M. (2015). Gene Ontology Consortium: going forward. *Nucleic Acids Research*, 43(D1), D1049-D1056. <https://doi.org/10.1093/NAR/GKU1179>
- Bleidorn, C. (2015). Third generation sequencing: technology and its potential impact on evolutionary biodiversity research. <http://dx.doi.org/10.1080/14772000.2015.1099575>, 14(1), 1-8. <https://doi.org/10.1080/14772000.2015.1099575>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-2120. <https://doi.org/10.1093/BIOINFORMATICS/BTU170>
- Bourgeois, Y., & Boissinot, S. (2019). On the Population Dynamics of Junk: A Review on the Population Genomics of Transposable Elements. *Genes* 2019, Vol. 10, Page 419, 10(6), 419. <https://doi.org/10.3390/GENES10060419>
- Bourque, G., Burns, K. H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvák, Z., Levin, H. L., Macfarlan, T. S., Mager, D. L., & Feschotte, C. (2018). Ten things you should know about transposable elements. *Genome Biology* 2018 19:1, 19(1), 1-12. <https://doi.org/10.1186/S13059-018-1577-Z>
- Bozin, B., Mimica-Dukic, N., Samojlik, I., & Jovin, E. (2007). Antimicrobial and antioxidant properties of Rosemary and Sage (*Rosmarinus officinalis* L. and *Salvia officinalis* L., Lamiaceae) essential oils. *Journal of Agricultural and Food Chemistry*, 55(19), 7879-7885. [https://doi.org/10.1021/JF0715323/ASSET/IMAGES/MEDIUM/JF-2007-015323\\_0004.GIF](https://doi.org/10.1021/JF0715323/ASSET/IMAGES/MEDIUM/JF-2007-015323_0004.GIF)

- Brent, M. R. (2005). Genome annotation past, present, and future: How to define an ORF at each locus. *Genome Research*, *15*(12), 1777-1786. <https://doi.org/10.1101/GR.3866105>
- Brent, M. R., & Guigó, R. (2004). Recent advances in gene structure prediction. *Current Opinion in Structural Biology*, *14*(3), 264-272. <https://doi.org/10.1016/J.SBI.2004.05.007>
- Brooksbank, C., Bergman, M. T., Apweiler, R., Birney, E., & Thornton, J. (2014). The European Bioinformatics Institute's data resources 2014. *Nucleic Acids Research*, *42*(D1), D18-D25. <https://doi.org/10.1093/NAR/GKT1206>
- Butkiewicz, M., & Bush, W. S. (2016). In Silico Functional Annotation of Genomic Variation. *Current Protocols in Human Genetics*, *88*(1), 6.15.1-6.15.17. <https://doi.org/10.1002/0471142905.HG0615S88>
- Cantarel, B. L., Korf, I., Robb, S. M. C., Parra, G., Ross, E., Moore, B., Holt, C., Alvarado, A. S., & Yandell, M. (2008). MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Research*, *18*(1), 188-196. <https://doi.org/10.1101/GR.6743907>
- Cao, H., Hastie, A. R., Cao, D., Lam, E. T., Sun, Y., Huang, H., Liu, X., Lin, L., Andrews, W., Chan, S., Huang, S., Tong, X., Requa, M., Anantharaman, T., Krogh, A., Yang, H., Cao, H., & Xu, X. (2014). Rapid detection of structural variation in a human genome using nanochannel-based genome mapping technology. *GigaScience*, *3*(1), 1-11. <https://doi.org/10.1186/2047-217X-3-34/FIGURES/4>
- Capek, P., & Hřibalová, V. (2004). Water-soluble polysaccharides from *Salvia officinalis* L. possessing immunomodulatory activity. *Phytochemistry*, *65*(13), 1983-1992. <https://doi.org/10.1016/J.PHYTOCHEM.2004.05.020>
- Carta, C., Moretti, M. D. L., & Peana, A. T. (2011). Activity of the Oil of *Salvia officinalis* L. Against *Botrytis cinerea*. <http://dx.doi.org/10.1080/10412905.1996.9700647>, *8*(4), 399-404. <https://doi.org/10.1080/10412905.1996.9700647>
- Chaisson, M. J. P., Huddleston, J., Dennis, M. Y., Sudmant, P. H., Malig, M., Hormozdiari, F., Antonacci, F., Surti, U., Sandstrom, R., Boitano, M., Landolin, J. M., Stamatoyannopoulos, J. A., Hunkapiller, M. W., Korf, J., & Eichler, E. E. (2014). Resolving the complexity of the human genome using single-molecule sequencing. *Nature* *2014 517:7536*, *517*(7536), 608-611. <https://doi.org/10.1038/nature13907>
- Chaisson, M. J. P., Wilson, R. K., & Eichler, E. E. (2015). Genetic variation and the de novo assembly of human genomes. *Nature Reviews Genetics* *2015 16:11*, *16*(11), 627-640. <https://doi.org/10.1038/nrg3933>
- Chikhi, R., Limasset, A., & Medvedev, P. (2016). Compacting de Bruijn graphs from sequencing data quickly and in low memory. *Bioinformatics*, *32*(12), i201-i208. <https://doi.org/10.1093/BIOINFORMATICS/BTW279>
- Chikhi, R., & Medvedev, P. (2014). Informed and automated k-mer size selection for genome assembly. *Bioinformatics*, *30*(1), 31-37. <https://doi.org/10.1093/BIOINFORMATICS/BTT310>
- Chikhi, R., & Rizk, G. (2013a). *Minia-Short manual*.
- Chikhi, R., & Rizk, G. (2013b). Space-efficient and exact de Bruijn graph representation based on a Bloom filter. *Algorithms for Molecular Biology*, *8*(1), 1-9. <https://doi.org/10.1186/1748-7188-8-22/TABLES/3>

- Clark, D. P., Pazdernik, N. J., & McGehee, M. R. (2019). Molecular Evolution. *Molecular Biology*, 925-969. <https://doi.org/10.1016/B978-0-12-813288-3.00029-X>
- Clarke, J., Wu, H. C., Jayasinghe, L., Patel, A., Reid, S., & Bayley, H. (2009). Continuous base identification for single-molecule nanopore DNA sequencing. *Nature Nanotechnology* 2009 4:4, 4(4), 265-270. <https://doi.org/10.1038/nnano.2009.12>
- Cole, R., Ost, K., & Schirra, S. (2001). Edge-coloring bipartite multigraphs in  $O(E \log D)$  time. *Combinatorica*, 21(1), 5-12. <https://doi.org/10.1007/S004930170002/METRICS>
- Compeau, P. E. C., Pevzner, P. A., & Tesler, G. (2011). How to apply de Bruijn graphs to genome assembly. *Nature Biotechnology* 2011 29:11, 29(11), 987-991. <https://doi.org/10.1038/nbt.2023>
- Cordaux, R., & Batzer, M. A. (2009). The impact of retrotransposons on human genome evolution. *Nature Reviews Genetics* 2009 10:10, 10(10), 691-703. <https://doi.org/10.1038/nrg2640>
- Cutting, G. R. (2014). Annotating DNA variants is the next major goal for human genetics. *American Journal of Human Genetics*, 94(1), 5-10. <https://doi.org/10.1016/j.ajhg.2013.12.008>
- Cuvelier, M. E., Richard, H., & Berset, C. (1996). Antioxidative activity and phenolic composition of pilot-plant and commercial extracts of sage and rosemary. *JAOCS, Journal of the American Oil Chemists' Society*, 73(5), 645-652. <https://doi.org/10.1007/BF02518121/METRICS>
- de Koning, A. P. J., Gu, W., Castoe, T. A., Batzer, M. A., & Pollock, D. D. (2011). Repetitive Elements May Comprise Over Two-Thirds of the Human Genome. *PLOS Genetics*, 7(12), e1002384. <https://doi.org/10.1371/JOURNAL.PGEN.1002384>
- Detter, J. C., Johnson, S. L., Bishop-Lilly, K. A., Chain, P. S., Gibbons, H. S., Minogue, T. D., Sozhamannan, S., Van Gieson, E. J., & Resnick, I. G. (2014). Nucleic acid sequencing for characterizing infectious and/or novel agents in complex samples. *Biological Identification: DNA Amplification and Sequencing, Optical Sensing, Lab-On-Chip and Portable Systems*, 3-53. <https://doi.org/10.1533/9780857099167.1.3>
- Dubey, V. S., Bhalla, R., & Luthra, R. (2003). An overview of the non-mevalonate pathway for terpenoid biosynthesis in plants. *Journal of Biosciences*, 28(5), 637-646. <https://doi.org/10.1007/BF02703339/METRICS>
- Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Peluso, P., Rank, D., Baybayan, P., Bettman, B., Bibillo, A., Bjornson, K., Chaudhuri, B., Christians, F., Cicero, R., Clark, S., Dalal, R., DeWinter, A., Dixon, J., ... Turner, S. (2009). Real-time DNA sequencing from single polymerase molecules. *Science*, 323(5910), 133-138. [https://doi.org/10.1126/SCIENCE.1162986/SUPPL\\_FILE/EID.SOM.PDF](https://doi.org/10.1126/SCIENCE.1162986/SUPPL_FILE/EID.SOM.PDF)
- Eidi, A., & Eidi, M. (2009). Antidiabetic effects of sage (*Salvia officinalis* L.) leaves in normal and streptozotocin-induced diabetic rats. *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*, 3(1), 40-44. <https://doi.org/10.1016/J.DSX.2008.10.007>
- Eidi, M., Eidi, A., & Bahar, M. (2006). Effects of *Salvia officinalis* L. (sage) leaves on memory retention and its interaction with the cholinergic system in rats. *Nutrition*, 22(3), 321-326. <https://doi.org/10.1016/J.NUT.2005.06.010>

- Ejigu, G. F., & Jung, J. (2020). Review on the Computational Genome Annotation of Sequences Obtained by Next-Generation Sequencing. *Biology* 2020, Vol. 9, Page 295, 9(9), 295. <https://doi.org/10.3390/BIOLOGY9090295>
- El-Metwally, S., Hamza, T., Zakaria, M., & Helmy, M. (2013). Next-generation sequence assembly: four stages of data processing and computational challenges. *PLoS computational biology*, 9(12). <https://doi.org/10.1371/JOURNAL.PCBI.1003345>
- Eriksson, C. E., Ruprecht, J., & Levi, T. (2020). More affordable and effective noninvasive single nucleotide polymorphism genotyping using high-throughput amplicon sequencing. *Molecular Ecology Resources*, 20(6), 1505-1516. <https://doi.org/10.1111/1755-0998.13208>
- Fang, S., Zhang, L., Guo, J., Niu, Y., Wu, Y., Li, H., Zhao, L., Li, X., Teng, X., Sun, X., Sun, L., Zhang, M. Q., Chen, R., & Zhao, Y. (2018). NONCODEV5: a comprehensive annotation database for long non-coding RNAs. *Nucleic Acids Research*, 46(D1), D308-D314. <https://doi.org/10.1093/NAR/GKX1107>
- Ferragina, P., & Manzini, G. (2000). Opportunistic data structures with applications. *Annual Symposium on Foundations of Computer Science - Proceedings*, 390-398. <https://doi.org/10.1109/SFCS.2000.892127>
- Ferrarini, M., Moretto, M., Ward, J. A., Šurbanovski, N., Stevanović, V., Giongo, L., Viola, R., Cavalieri, D., Velasco, R., Cestaro, A., & Sargent, D. J. (2013). An evaluation of the PacBio RS platform for sequencing and de novo assembly of a chloroplast genome. *BMC Genomics*, 14(1), 1-12. <https://doi.org/10.1186/1471-2164-14-670/FIGURES/4>
- Garcia, C. S. C., Menti, C., Lambert, A. P. F., Barcellos, T., Moura, S., Calloni, C., Branco, C. S., Salvador, M., Roesch-Ely, M., & Henriques, J. A. P. (2016). Pharmacological perspectives from Brazilian *Salvia officinalis* (Lamiaceae): antioxidant, and antitumor in mammalian cells. *Anais da Academia Brasileira de Ciências*, 88(1), 281-292. <https://doi.org/10.1590/0001-3765201520150344>
- Geer, L. Y., Marchler-Bauer, A., Geer, R. C., Han, L., He, J., He, S., Liu, C., Shi, W., & Bryant, S. H. (2010). The NCBI BioSystems database. *Nucleic Acids Research*, 38(suppl\_1), D492-D496. <https://doi.org/10.1093/NAR/GKP858>
- Ghorbani, A., & Esmailizadeh, M. (2017). Pharmacological properties of *Salvia officinalis* and its components. *Journal of Traditional and Complementary Medicine*, 7(4), 433-440. <https://doi.org/10.1016/J.JTCME.2016.12.014>
- Girgis, H. Z. (2015). Red: An intelligent, rapid, accurate tool for detecting repeats de-novo on the genomic scale. *BMC Bioinformatics*, 16(1), 1-19. <https://doi.org/10.1186/S12859-015-0654-5/TABLES/4>
- Gnerre, S., MacCallum, I., Przybylski, D., Ribeiro, F. J., Burton, J. N., Walker, B. J., Sharpe, T., Hall, G., Shea, T. P., Sykes, S., Berlin, A. M., Aird, D., Costello, M., Daza, R., Williams, L., Nicol, R., Gnirke, A., Nusbaum, C., Lander, E. S., & Jaffe, D. B. (2011). High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proceedings of the National Academy of Sciences of the United States of America*, 108(4), 1513-1518. <https://doi.org/10.1073/PNAS.1017351108>
- Gomar, A., Hosseini, A., & Mirazi, N. (2014). Evaluation of *Salvia officinalis* L. (sage) leaves on morphine-induced memory impairment in adult male rats. *Focus on Alternative and Complementary Therapies*, 19(3), 156-162. <https://doi.org/10.1111/FCT.12132>

- Goodwin, S., McPherson, J. D., & Richard McCombie, W. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Publishing Group*.  
<https://doi.org/10.1038/nrg.2016.49>
- Gordon, D., Huddleston, J., Chaisson, M. J. P., Hill, C. M., Kronenberg, Z. N., Munson, K. M., Malig, M., Raja, A., Fiddes, I., Hillier, L. D. W., Dunn, C., Baker, C., Armstrong, J., Diekhans, M., Paten, B., Shendure, J., Wilson, R. K., Haussler, D., Chin, C. S., & Eichler, E. E. (2016). Long-read sequence assembly of the gorilla genome. *Science (New York, N.Y.)*, 352(6281).  
<https://doi.org/10.1126/SCIENCE.AAE0344>
- Guo, J., Xu, N., Li, Z., Zhang, S., Wu, J., Hyun Kim, D., Sano Marma, M., Meng, Q., Cao, H., Li, X., Shi, S., Yu, L., Kalachikov, S., Russo, J. J., Turro, N. J., & Ju, J. (2008). *Four-color DNA sequencing with 3-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides*.  
[www.pnas.org/cgi/doi/10.1073/pnas.0804023105](http://www.pnas.org/cgi/doi/10.1073/pnas.0804023105)
- Halicioglu, O., Astarcioglu, G., Yaprak, I., & Aydinlioglu, H. (2011). Toxicity of *Salvia officinalis* in a newborn and a child: an alarming report. *Pediatric neurology*, 45(4), 259-260. <https://doi.org/10.1016/J.PEDIATRNEUROL.2011.05.012>
- Hanada, K., Vallejo, V., Nobuta, K., Slotkin, R. K., Lisch, D., Meyers, B. C., Shiu, S. H., & Jiang, N. (2009). The Functional Role of Pack-MULEs in Rice Inferred from Purifying Selection and Expression Profile. *The Plant Cell*, 21(1), 25-38.  
<https://doi.org/10.1105/TPC.108.063206>
- Hayouni, E. A., Chraief, I., Abedrabba, M., Bouix, M., Leveau, J. Y., Mohammed, H., & Hamdi, M. (2008). Tunisian *Salvia officinalis* L. and *Schinus molle* L. essential oils: Their chemical compositions and their preservative effects against *Salmonella* inoculated in minced beef meat. *International Journal of Food Microbiology*, 125(3), 242-251. <https://doi.org/10.1016/J.IJFOODMICRO.2008.04.005>
- Held, M., & Karp, R. M. (1962). A Dynamic Programming Approach to Sequencing Problems. *Journal of the Society for Industrial and Applied Mathematics*, 10(1), 196-210. <https://doi.org/10.1137/0110015>
- Hernández-Saavedra, D., Pérez-Ramírez, I. F., Ramos-Gómez, M., Mendoza-Díaz, S., Loarca-Piña, G., & Reynoso-Camacho, R. (2016). Phytochemical characterization and effect of *Calendula officinalis*, *Hypericum perforatum*, and *Salvia officinalis* infusions on obesity-associated cardiovascular risk. *Medicinal Chemistry Research*, 25(1), 163-172. <https://doi.org/10.1007/S00044-015-1454-1/TABLES/4>
- Horiuchi, K., Shiota, S., Hatano, T., Yoshida, T., Kuroda, T., & Tsuchiya, T. (2007). Antimicrobial Activity of Oleanolic Acid from *Salvia officinalis* and Related Compounds on Vancomycin-Resistant Enterococci (VRE). *Biological and Pharmaceutical Bulletin*, 30(6), 1147-1149. <https://doi.org/10.1248/BPB.30.1147>
- Horiuchi, K., Shiota, S., Kuroda, T., Hatano, T., Yoshida, T., & Tsuchiya, T. (2007). Potentiation of Antimicrobial Activity of Aminoglycosides by Carnosol from *Salvia officinalis*. *Biological and Pharmaceutical Bulletin*, 30(2), 287-290.  
<https://doi.org/10.1248/BPB.30.287>
- Horváthová, E., Srančíková, A., Regendová-Sedláčková, E., Melušová, M., Meluš, V., Netriová, J., Krajčovičová, Z., Slameňová, D., Pastorek, M., & Kozics, K. (2016). Enriching the drinking water of rats with extracts of *Salvia officinalis* and *Thymus vulgaris* increases their resistance to oxidative stress. *Mutagenesis*, 31(1), 51-59.  
<https://doi.org/10.1093/MUTAGE/GEV056>
- Hu, T., Chitnis, N., Monos, D., & Dinh, A. (2021). Next-generation sequencing technologies: An overview. *Human Immunology*, 82(11), 801-811.  
<https://doi.org/10.1016/J.HUMIMM.2021.02.012>

- Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., Zhu, C., Lu, T., Zhang, Z., Li, M., Fan, D., Guo, Y., Wang, A., Wang, L., Deng, L., Li, W., Lu, Y., Weng, Q., ... Han, B. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics* 2010 42:11, 42(11), 961-967. <https://doi.org/10.1038/ng.695>
- Hubley, R., Finn, R. D., Clements, J., Eddy, S. R., Jones, T. A., Bao, W., Smit, A. F. A., & Wheeler, T. J. (2016). The Dfam database of repetitive DNA families. *Nucleic Acids Research*, 44(D1), D81-D89. <https://doi.org/10.1093/NAR/GKV1272>
- Huerta-Cepas, J., Forslund, K. ;, Coelho, L. P., Szklarczyk, D. ;, Jensen, L. J., Von Mering, C. ;, & Bork, P. (2017). Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Molecular Biology and Evolution*, 34(8), 2115-2122. <https://doi.org/10.1093/molbev/msx148>
- Hunter, S., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L., Finn, R. D., Gough, J., Haft, D., Hulo, N., Kahn, D., Kelly, E., Laugraud, A., Letunic, I., Lonsdale, D., ... Yeats, C. (2009). InterPro: the integrative protein signature database. *Nucleic Acids Research*, 37(suppl\_1), D211-D215. <https://doi.org/10.1093/NAR/GKN785>
- Idury, R. M., & Waterman, M. S. (2009). A New Algorithm for DNA Sequence Assembly. <https://home.liebertpub.com/cmb>, 2(2), 291-306. <https://doi.org/10.1089/CMB.1995.2.291>
- Ip, C. L. C., Loose, M., Tyson, J. R., de Cesare, M., Brown, B. L., Jain, M., Leggett, R. M., Eccles, D. A., Zalunin, V., Urban, J. M., Piazza, P., Bowden, R. J., Paten, B., Mwaigwisya, S., Batty, E. M., Simpson, J. T., Snutch, T. P., Birney, E., Buck, D., ... Olsen, H. E. (2015). MinION Analysis and Reference Consortium: Phase 1 data release and analysis. *F1000Research*, 4. <https://doi.org/10.12688/F1000RESEARCH.7201.1/DOI>
- Jia, J., Zhao, S., Kong, X., Li, Y., Zhao, G., He, W., Appels, R., Pfeifer, M., Tao, Y., Zhang, X., Jing, R., Zhang, C., Ma, Y., Gao, L., Gao, C., Spannagl, M., Mayer, K. F. X., Li, D., Pan, S., ... Mao, L. (2013). *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* 2013 496:7443, 496(7443), 91-95. <https://doi.org/10.1038/nature12028>
- Jiang, N. (2013). Overview of repeat annotation and De Novo repeat identification. *Methods in Molecular Biology*, 1057, 275-287. [https://doi.org/10.1007/978-1-62703-568-2\\_20/COVER](https://doi.org/10.1007/978-1-62703-568-2_20/COVER)
- Ju, J., Hyun Kim, D., Bi, L., Meng, Q., Bai, X., Li, Z., Li, X., Sano Marma, M., Shi, S., Wu, J., Edwards, J. R., Romu, A., & Turro, N. J. (2006). *Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators*. [www.pnas.org/cgi/content/full/](http://www.pnas.org/cgi/content/full/)
- Jug-Dujaković, M., Ristić, M., Pljevljakušić, D., Dajić-Stevanović, Z., Liber, Z., Hančević, K., Radić, T., & Šatović, Z. (2012). High Diversity of Indigenous Populations of Dalmatian Sage (*Salvia officinalis* L.) in Essential-Oil Composition. *Chemistry & Biodiversity*, 9(10), 2309-2323. <https://doi.org/10.1002/CBDV.201200131>
- Jurka, J. (2000). Repbase Update: A database and an electronic journal of repetitive elements. *Trends in Genetics*, 16(9), 418-420. [https://doi.org/10.1016/S0168-9525\(00\)02093-X](https://doi.org/10.1016/S0168-9525(00)02093-X)
- Jurka, J., Kapitonov, V. V., Pavlicek, A., Klonowski, P., Kohany, O., & Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. *Cytogenetic and Genome Research*, 110(1-4), 462-467. <https://doi.org/10.1159/000084979>

- Kanehisa, M., & Goto, S. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28(1), 27-30. <https://doi.org/10.1093/NAR/28.1.27>
- Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, 44(D1), D457-D462. <https://doi.org/10.1093/NAR/GKV1070>
- Karro, J. E., Yan, Y., Zheng, D., Zhang, Z., Carriero, N., Cayting, P., Harrision, P., & Gerstein, M. (2007). Pseudogene.org: a comprehensive database and comparison platform for pseudogene annotation. *Nucleic Acids Research*, 35(suppl\_1), D55-D60. <https://doi.org/10.1093/NAR/GKL851>
- Keller, O., Kollmar, M., Stanke, M., & Waack, S. (2011). A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics*, 27(6), 757-763. <https://doi.org/10.1093/BIOINFORMATICS/BTR010>
- Kelley, D. R., & Salzberg, S. L. (2010). Detection and correction of false segmental duplications caused by genome mis-assembly. *Genome Biology*, 11(3), 1-11. <https://doi.org/10.1186/GB-2010-11-3-R28/FIGURES/5>
- Kelley, D. R., Schatz, M. C., & Salzberg, S. L. (2010). Quake: Quality-aware detection and correction of sequencing errors. *Genome Biology*, 11(11), 1-13. <https://doi.org/10.1186/GB-2010-11-11-R116/FIGURES/6>
- Kennedy, D. O., Pace, S., Haskell, C., Okello, E. J., Milne, A., & Scholey, A. B. (2005). Effects of Cholinesterase Inhibiting Sage (*Salvia officinalis*) on Mood, Anxiety and Performance on a Psychological Stressor Battery. *Neuropsychopharmacology* 2006 31:4, 31(4), 845-852. <https://doi.org/10.1038/sj.npp.1300907>
- Kent, W. J. (2002). BLAT—The BLAST-Like Alignment Tool. *Genome Research*, 12(4), 656-664. <https://doi.org/10.1101/GR.229202>
- Khattab, H., & Mohamed, R. A. (2012). *Evaluation of Hypoglycemic Activity of Salvia officinalis L. (Sage) Infusion on Streptozotocin-Induced Diabetic Rats.*
- Kianbakht, S., Abasi, B., Perham, M., & Hashem Dabaghian, F. (2011). Antihyperlipidemic effects of *Salvia officinalis* L. leaf extract in patients with hyperlipidemia: a randomized double-blind placebo-controlled clinical trial. *Phytotherapy research : PTR*, 25(12), 1849-1853. <https://doi.org/10.1002/PTR.3506>
- Kim, D., Langmead, B., & Salzberg, S. L. (2015). HISAT: a fast spliced aligner with low memory requirements. *Nature Methods* 2015 12:4, 12(4), 357-360. <https://doi.org/10.1038/nmeth.3317>
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology*, 14(4), 1-13. <https://doi.org/10.1186/GB-2013-14-4-R36/FIGURES/6>
- Koito, A., & Ikeda, T. (2013). Intrinsic immunity against retrotransposons by APOBEC cytidine deaminases. *Frontiers in Microbiology*, 4(FEB), 31381. <https://doi.org/10.3389/FMICB.2013.00028/BIBTEX>
- Koonin, E. V. (2005). Orthologs, Paralogs, and Evolutionary Genomics1. <https://doi.org/10.1146/annurev.genet.39.073003.114725>, 39, 309-338. <https://doi.org/10.1146/ANNUREV.GENET.39.073003.114725>
- Korlach, J. (t.y.). *Perspective - Understanding Accuracy in SMRT Sequencing.* Geliş tarihi 08 Mayıs 2023, gönderen [www.pacb.com](http://www.pacb.com)
- Korlach, J., Bjornson, K. P., Chaudhuri, B. P., Cicero, R. L., Flusberg, B. A., Gray, J. J., Holden, D., Saxena, R., Wegener, J., & Turner, S. W. (2010). Real-time DNA

- sequencing from single polymerase molecules. *Methods in enzymology*, 472, 431-455. [https://doi.org/10.1016/S0076-6879\(10\)72001-2](https://doi.org/10.1016/S0076-6879(10)72001-2)
- Kozomara, A., Birgaoanu, M., & Griffiths-Jones, S. (2019). miRBase: from microRNA sequences to function. *Nucleic Acids Research*, 47(D1), D155-D162. <https://doi.org/10.1093/NAR/GKY1141>
- Laehnemann, D., Borkhardt, A., & McHardy, A. C. (2016). Denoising DNA deep sequencing data-high-throughput sequencing errors and their correction. *Briefings in bioinformatics*, 17(1), 154-179. <https://doi.org/10.1093/BIB/BBV029>
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., Fitzhugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., Levine, R., McEwan, P., ... Morgan, M. J. (2001). Initial sequencing and analysis of the human genome. *Nature 2001* 409:6822, 409(6822), 860-921. <https://doi.org/10.1038/35057062>
- Lee, H., Gurtowski, J., Yoo, S., Nattestad, M., Marcus, S., Goodwin, S., McCombie, W. R., & Schatz, M. C. (2016). *Third-generation sequencing and the future of genomics*. <https://doi.org/10.1101/048603>
- Levy, S., Sutton, G., Ng, P. C., Feuk, L., Halpern, A. L., Walenz, B. P., Axelrod, N., Huang, J., Kirkness, E. F., Denisov, G., Lin, Y., MacDonald, J. R., Pang, A. W. C., Shago, M., Stockwell, T. B., Tsiamouri, A., Bafna, V., Bansal, V., Kravitz, S. A., ... Venter, J. C. (2007). The diploid genome sequence of an individual human. *PLoS biology*, 5(10), 2113-2144. <https://doi.org/10.1371/JOURNAL.PBIO.0050254>
- Li, C. Y., Yang, L., Liu, Y., Xu, Z. G., Gao, J., Huang, Y. B., Xu, J. J., Fan, H., Kong, Y., Wei, Y. K., Hu, W. L., Wang, L. J., Zhao, Q., Hu, Y. H., Zhang, Y. J., Martin, C., & Chen, X. Y. (2022). The sage genome provides insight into the evolutionary dynamics of diterpene biosynthesis gene cluster in plants. *Cell Reports*, 40(7). <https://doi.org/10.1016/J.CELREP.2022.111236/ATTACHMENT/1412F8B8-7C97-4824-B72C-C365983DF3D0/MMC1.PDF>
- Li, J., Wang, Y., Dong, Y., Zhang, W., Wang, D., Bai, H., Li, K., Li, H., & Shi, L. (2021). The chromosome-based lavender genome provides new insights into Lamiaceae evolution and terpenoid biosynthesis. *Horticulture Research*, 8(1), 53. [https://doi.org/10.1038/S41438-021-00490-6/44170926/41438\\_2021\\_ARTICLE\\_490.PDF](https://doi.org/10.1038/S41438-021-00490-6/44170926/41438_2021_ARTICLE_490.PDF)
- Li, R., Fan, W., Tian, G., Zhu, H., He, L., Cai, J., Huang, Q., Cai, Q., Li, B., Bai, Y., Zhang, Z., Zhang, Y., Wang, W., Li, J., Wei, F., Li, H., Jian, M., Li, J., Zhang, Z., ... Wang, J. (2009). The sequence and de novo assembly of the giant panda genome. *Nature 2009* 463:7279, 463(7279), 311-317. <https://doi.org/10.1038/nature08696>
- Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., Li, Y., Li, S., Shan, G., Kristiansen, K., Li, S., Yang, H., Wang, J., & Wang, J. (2010). De novo assembly of human genomes with massively parallel short read sequencing. *Genome research*, 20(2), 265-272. <https://doi.org/10.1101/GR.097261.109>
- Li, X., & Waterman, M. S. (2003). Estimating the repeat structure and length of DNA sequences using L-tuples. *Genome research*, 13(8), 1916-1922. <https://doi.org/10.1101/GR.1251803>
- Lima, C. F., Valentao, P. C. R., Andrade, P. B., Seabra, R. M., Fernandes-Ferreira, M., & Pereira-Wilson, C. (2007). Water and methanolic extracts of *Salvia officinalis* protect HepG2 cells from t-BHP induced oxidative damage. *Chemico-Biological Interactions*, 167(2), 107-115. <https://doi.org/10.1016/J.CBI.2007.01.020>

- Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., & Law, M. (2012). Comparison of next-generation sequencing systems. *Journal of biomedicine & biotechnology*, 2012. <https://doi.org/10.1155/2012/251364>
- Logsdon, G. A., Vollger, M. R., & Eichler, E. E. (2020). Long-read human genome sequencing and its applications. *Nature Reviews Genetics* 2020 21:10, 21(10), 597-614. <https://doi.org/10.1038/s41576-020-0236-x>
- Longaray Delamare, A. P., Moschen-Pistorello, I. T., Artico, L., Atti-Serafini, L., & Echeverrigaray, S. (2007). Antibacterial activity of the essential oils of *Salvia officinalis* L. and *Salvia triloba* L. cultivated in South Brazil. *Food Chemistry*, 100(2), 603-608. <https://doi.org/10.1016/J.FOODCHEM.2005.09.078>
- Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., ... Wang, J. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience*, 1(1). <https://doi.org/10.1186/2047-217X-1-18>
- Maksimović, M., Vidic, D., Miloš, M., Edita Šolić, M., Abadžić, S., & Siljak-Yakovlev, S. (2007). Effect of the environmental conditions on essential oil profile in two Dinaric *Salvia* species: *S. brachyodon* Vandas and *S. officinalis* L. *Biochemical Systematics and Ecology*, 35(8), 473-478. <https://doi.org/10.1016/J.BSE.2007.02.005>
- Marçais, G., Delcher, A. L., Phillippy, A. M., Coston, R., Salzberg, S. L., & Zimin, A. (2018). MUMmer4: A fast and versatile genome alignment system. *PLOS Computational Biology*, 14(1), e1005944. <https://doi.org/10.1371/JOURNAL.PCBI.1005944>
- Mardis, E. R. (2008a). Next-generation DNA sequencing methods. *Annual review of genomics and human genetics*, 9, 387-402. <https://doi.org/10.1146/ANNUREV.GENOM.9.081307.164359>
- Mardis, E. R. (2008b). The impact of next-generation sequencing technology on genetics. *Trends in genetics : TIG*, 24(3), 133-141. <https://doi.org/10.1016/J.TIG.2007.12.007>
- Mardis, E. R. (2013). Next-Generation Sequencing Platforms. <https://doi.org/10.1146/annurev-anchem-062012-092628>, 6, 287-303. <https://doi.org/10.1146/ANNUREV-ANCHEM-062012-092628>
- Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bembien, L. A., Berka, J., Braverman, M. S., Chen, Y. J., Chen, Z., Dewell, S. B., Du, L., Fierro, J. M., Gomes, X. V., Godwin, B. C., He, W., Helgesen, S., Ho, C. H., Irzyk, G. P., ... Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005 437:7057, 437(7057), 376-380. <https://doi.org/10.1038/nature03959>
- Mathé, C., Sagot, M. F., Schiex, T., & Rouzé, P. (2002). Current methods of gene prediction, their strengths and weaknesses. *Nucleic Acids Research*, 30(19), 4103-4117. <https://doi.org/10.1093/NAR/GKF543>
- Maxam, A. M., & Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods in Enzymology*, 65(C), 499-560. [https://doi.org/10.1016/S0076-6879\(80\)65059-9](https://doi.org/10.1016/S0076-6879(80)65059-9)
- McCLINTOCK, B. (1950). The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences*, 36(6), 344-355. <https://doi.org/10.1073/PNAS.36.6.344>
- Mcgarvey, D. J., & Croteau, R. (1995). Terpenoid metabolism. *The Plant Cell*, 7(7), 1015-1026. <https://doi.org/10.1105/TPC.7.7.1015>

- Media resource center - PacBio.* (t.y.). Geliş tarihi 07 Mayıs 2023, gönderen <https://www.pacb.com/media-resource-center/>
- Melsted, P., & Pritchard, J. K. (2011). Efficient counting of k-mers in DNA sequences using a bloom filter. *BMC Bioinformatics*, *12*(1), 1-7. <https://doi.org/10.1186/1471-2105-12-333/TABLES/1>
- Metzker, M. L. (2009). Sequencing technologies — the next generation. *Nature Reviews Genetics* *2010 11:1*, *11*(1), 31-46. <https://doi.org/10.1038/nrg2626>
- Miga, K. H., Koren, S., Rhie, A., Vollger, M. R., Gershman, A., Bzikadze, A., Brooks, S., Howe, E., Porubsky, D., Logsdon, G. A., Schneider, V. A., Potapova, T., Wood, J., Chow, W., Armstrong, J., Fredrickson, J., Pak, E., Tigyi, K., Kremitzki, M., ... Phillippy, A. M. (2020). Telomere-to-telomere assembly of a complete human X chromosome. *Nature* |, *585*, 79. <https://doi.org/10.1038/s41586-020-2547-7>
- Mills, S., & Bone, K. (2005). St. Mary's Thistle. *The Essential Guide to Herbal Safety*, 594-596. <https://doi.org/10.1016/B978-0-7295-4172-5.00054-5>
- Miroddi, M., Navarra, M., Quattropiani, M. C., Calapai, F., Gangemi, S., & Calapai, G. (2014). Systematic Review of Clinical Trials Assessing Pharmacological Properties of Salvia Species on Memory, Cognitive Impairment and Alzheimer's Disease. *CNS Neuroscience & Therapeutics*, *20*(6), 485-495. <https://doi.org/10.1111/CNS.12270>
- Mishra, S., Rastogi, Y. P., Jabin, S., Kaur, P., Amir, M., & Khatoon, S. (2020). A bacterial phyla dataset for protein function prediction. *Data in Brief*, *28*, 105002. <https://doi.org/10.1016/J.DIB.2019.105002>
- Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., Brown, S. D., Chang, H. Y., El-Gebali, S., Fraser, M. I., Gough, J., Haft, D. R., Huang, H., Letunic, I., Lopez, R., Luciani, A., Madeira, F., Marchler-Bauer, A., Mi, H., ... Finn, R. D. (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Research*, *47*(D1), D351-D360. <https://doi.org/10.1093/NAR/GKY1100>
- Mitic-Culafic, D., Vukovic-Gacic, B., Knezevic-Vukcevic, J., Stankovic, S., & Simic, D. (2005). Comparative study on the antibacterial activity of volatiles from sage (*Salvia officinalis* L.). *Archives of Biological Sciences*, *57*(3), 173-178. <https://doi.org/10.2298/ABS0503173M>
- Modrek, B., & Lee, C. (2002). A genomic view of alternative splicing. *Nature Genetics* *2002 30:1*, *30*(1), 13-19. <https://doi.org/10.1038/ng0102-13>
- Moss, L., Rouse, M., Wesnes, K. A., & Moss, M. (2010). Differential effects of the aromas of Salvia species on memory and mood. *Human Psychopharmacology: Clinical and Experimental*, *25*(5), 388-396. <https://doi.org/10.1002/HUP.1129>
- Moss, M., Rouse, M., Moss, L., Moss, M., Rouse, M., & Moss, L. (2014). Aromas of Salvia Species Enhance Everyday Prospective Memory Performance in Healthy Young Adults. *Advances in Chemical Engineering and Science*, *4*(3), 339-346. <https://doi.org/10.4236/ACES.2014.43037>
- Mudge, J. M., & Harrow, J. (2016). The state of play in higher eukaryote gene annotation. *Nature Reviews Genetics* *2016 17:12*, *17*(12), 758-772. <https://doi.org/10.1038/nrg.2016.119>
- Neves Carvalho, A., Firuzi, O., Joao Gama, M., van Horssen, J., & Saso, L. (2017). Oxidative Stress and Antioxidants in Neurological Diseases: Is There Still Hope? *Current Drug Targets*, *18*.
- Pavlopoulos, G. A., Oulas, A., Iacucci, E., Sifrim, A., Moreau, Y., Schneider, R., Aerts, J., & Iliopoulos, I. (2013). Unraveling genomic variation from next generation

- sequencing data. *BioData Mining*, 6(1), 1-25. <https://doi.org/10.1186/1756-0381-6-13/TABLES/5>
- Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B. L., Salazar, G. A., Bileschi, M. L., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D. H., Letunić, I., Marchler-Bauer, A., Mi, H., Natale, D. A., Orengo, C. A., Pandurangan, A. P., Rivoire, C., ... Bateman, A. (2023). InterPro in 2022. *Nucleic Acids Research*, 51(D1), D418-D427. <https://doi.org/10.1093/NAR/GKAC993>
- Pedro, D. F. N., Ramos, A. A., Lima, C. F., Baltazar, F., & Pereira-Wilson, C. (2016). Colon Cancer Chemoprevention by Sage Tea Drinking: Decreased DNA Damage and Cell Proliferation. *Phytotherapy Research*, 30(2), 298-305. <https://doi.org/10.1002/PTR.5531>
- Pell, J., Hintze, A., Canino-Koning, R., Howe, A., Tiedje, J. M., & Brown, C. T. (2012). Scaling metagenome sequence assembly with probabilistic de Bruijn graphs. *Proceedings of the National Academy of Sciences of the United States of America*, 109(33), 13272-13277. [https://doi.org/10.1073/PNAS.1121464109/SUPPL\\_FILE/PNAS.1121464109\\_SLPDF](https://doi.org/10.1073/PNAS.1121464109/SUPPL_FILE/PNAS.1121464109_SLPDF)
- Pendleton, M., Sebra, R., Pang, A. W. C., Ummat, A., Franzen, O., Rausch, T., Stütz, A. M., Stedman, W., Anantharaman, T., Hastie, A., Dai, H., Fritz, M. H. Y., Cao, H., Cohain, A., Deikus, G., Durrett, R. E., Blanchard, S. C., Altman, R., Chin, C. S., ... Bashir, A. (2015). Assembly and diploid architecture of an individual human genome via single-molecule technologies. *Nature Methods* 2015 12:8, 12(8), 780-786. <https://doi.org/10.1038/nmeth.3454>
- Peng, Y., Leung, H. C. M., Yiu, S. M., & Chin, F. Y. L. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics*, 28(11), 1420-1428. <https://doi.org/10.1093/BIOINFORMATICS/BTS174>
- Perry, E. K., Pickering, A. T., Wang, W. W., Houghton, P. J., & Perry, N. S. L. (2010). Medicinal Plants and Alzheimer's Disease: from Ethnobotany to Phytotherapy. *Journal of Pharmacy and Pharmacology*, 51(5), 527-534. <https://doi.org/10.1211/0022357991772808>
- Petersen, M., Armisen, D., Gibbs, R. A., Hering, L., Khila, A., Mayer, G., Richards, S., Niehuis, O., & Misof, B. (2019). Diversity and evolution of the transposable element repertoire in arthropods with particular reference to insects. *BMC Evolutionary Biology*, 19(1), 1-15. <https://doi.org/10.1186/S12862-018-1324-9/METRICS>
- Pevzner, P. A., Borodovsky, M. Y., & Mironov, A. A. (2012). Linguistics of Nucleotide Sequences II: Stationary Words in Genetic Texts and the Zonal Structure of DNA. <http://dx.doi.org/10.1080/07391102.1989.10506529>, 6(5), 1027-1038. <https://doi.org/10.1080/07391102.1989.10506529>
- Pevzner, P. A., Tang, H., & Waterman, M. S. (2001). An Eulerian path approach to DNA fragment assembly. *Proceedings of the National Academy of Sciences of the United States of America*, 98(17), 9748-9753. <https://doi.org/10.1073/PNAS.171285098>
- Pichersky, E., & Raguso, R. A. (2018). Why do plants produce so many terpenoid compounds? *New Phytologist*, 220(3), 692-702. <https://doi.org/10.1111/NPH.14178>
- Price, A. L., Jones, N. C., & Pevzner, P. A. (2005). De novo identification of repeat families in large genomes. *Bioinformatics*, 21(suppl\_1), i351-i358. <https://doi.org/10.1093/BIOINFORMATICS/BTI1018>

- Quick, J., Loman, N. J., Duraffour, S., Simpson, J. T., Severi, E., Cowley, L., Bore, J. A., Koundouno, R., Dudas, G., Mikhail, A., Ouédraogo, N., Afrough, B., Bah, A., Baum, J. H. J., Becker-Ziaja, B., Boettcher, J. P., Cabeza-Cabrerizo, M., Camino-Sánchez, Á., Carter, L. L., ... Carroll, M. W. (2016). Real-time, portable genome sequencing for Ebola surveillance. *Nature* 2016 530:7589, 530(7589), 228-232. <https://doi.org/10.1038/nature16996>
- Ramakrishnan, M., Satish, L., Sharma, A., Kurungara Vinod, K., Emamverdian, A., Zhou, M., & Wei, Q. (2022). Transposable elements in plants: Recent advancements, tools and prospects. *Plant Molecular Biology Reporter* 2022 40:4, 40(4), 628-645. <https://doi.org/10.1007/S11105-022-01342-W>
- Rhoads, A., & Au, K. F. (2015). PacBio Sequencing and Its Applications. *Genomics, proteomics & bioinformatics*, 13(5), 278-289. <https://doi.org/10.1016/J.GPB.2015.08.002>
- Roberts, R. J., Carneiro, M. O., & Schatz, M. C. (2013). The advantages of SMRT sequencing. *Genome Biology* 2013 14:7, 14(7), 1-4. <https://doi.org/10.1186/GB-2013-14-7-405>
- Russo, A., Formisano, C., Rigano, D., Senatore, F., Delfino, S., Cardile, V., Rosselli, S., & Bruno, M. (2013). Chemical composition and anticancer activity of essential oils of Mediterranean sage (*Salvia officinalis* L.) grown in different environmental conditions. *Food and Chemical Toxicology*, 55, 42-47. <https://doi.org/10.1016/J.FCT.2012.12.036>
- Russo, P., Frustaci, A., Del Bufalo, A., Fini, M., & Cesario, A. (2013). From Traditional European Medicine to Discovery of New Drug Candidates for the Treatment of Dementia and Alzheimer's Disease: Acetylcholinesterase Inhibitors. *Current Medicinal Chemistry*, 20(8), 976-983. <https://doi.org/10.2174/0929867311320080002>
- Sá, C. M., Ramos, A. A., Azevedo, M. F., Lima, C. F., Fernandes-Ferreira, M., & Pereira-Wilson, C. (2009). Sage tea drinking improves lipid profile and antioxidant defences in humans. *International journal of molecular sciences*, 10(9), 3937-3950. <https://doi.org/10.3390/IJMS10093937>
- Sacchettini, J. C., & Poulter, C. D. (1997). Creating Isoprenoid Diversity. *Science*, 277(5333), 1788-1789. <https://doi.org/10.1126/SCIENCE.277.5333.1788>
- Salzberg, S. L., Phillippy, A. M., Zimin, A., Puiu, D., Magoc, T., Koren, S., Treangen, T. J., Schatz, M. C., Delcher, A. L., Roberts, M., Marcxais, G., Pop, M., & Yorke, J. A. (2012). GAGE: A critical evaluation of genome assemblies and assembly algorithms. *Genome Research*, 22(3), 557-567. <https://doi.org/10.1101/GR.131383.111>
- Sanger, F., & Coulson, A. R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology*, 94(3), 441-448. [https://doi.org/10.1016/0022-2836\(75\)90213-2](https://doi.org/10.1016/0022-2836(75)90213-2)
- Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), 5463-5467. <https://doi.org/10.1073/PNAS.74.12.5463>
- Sayers, E. W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K. D., & Karsch-Mizrachi, I. (2019). GenBank. *Nucleic Acids Research*, 47(D1), D94-D99. <https://doi.org/10.1093/NAR/GKY989>
- Schatz, M. C., Delcher, A. L., & Salzberg, S. L. (2010). Assembly of large genomes using second-generation sequencing. *Genome research*, 20(9), 1165-1173. <https://doi.org/10.1101/GR.101360.109>

- Schneider, G. F., & Dekker, C. (2012). DNA sequencing with nanopores. *Nature Biotechnology* 2012 30:4, 30(4), 326-328. <https://doi.org/10.1038/nbt.2181>
- Scholey, A. B., Tildesley, N. T. J., Ballard, C. G., Wesnes, K. A., Tasker, A., Perry, E. K., & Kennedy, D. O. (2008). An extract of *Salvia* (sage) with anticholinesterase properties improves memory and attention in healthy older volunteers. *Psychopharmacology*, 198(1), 127-139. <https://doi.org/10.1007/S00213-008-1101-3/FIGURES/2>
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210-3212. <https://doi.org/10.1093/BIOINFORMATICS/BTV351>
- Simić, D., Vuković-Gačić, B., & Knežević-Vukčević, J. (1998). Detection of natural bioantimutagens and their mechanisms of action with bacterial assay-system. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 402(1-2), 51-57. [https://doi.org/10.1016/S0027-5107\(97\)00281-9](https://doi.org/10.1016/S0027-5107(97)00281-9)
- Simpson, J. T. (2014). Exploring genome characteristics and sequence quality without a reference. *Bioinformatics*, 30(9), 1228-1235. <https://doi.org/10.1093/BIOINFORMATICS/BTU023>
- Simpson, J. T., & Durbin, R. (2012). Efficient de novo assembly of large genomes using compressed data structures. *Genome Research*, 22(3), 549-556. <https://doi.org/10.1101/GR.126953.111>
- Simpson, J. T., Wong, K., Jackman, S. D., Schein, J. E., Jones, S. J. M., & Birol, I. (2009). ABySS: A parallel assembler for short read sequence data. *Genome Research*, 19(6), 1117-1123. <https://doi.org/10.1101/GR.089532.108>
- Sims, D., Sudbery, I., Ilott, N. E., Heger, A., & Ponting, C. P. (2014). Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics* 2014 15:2, 15(2), 121-132. <https://doi.org/10.1038/nrg3642>
- Slatko, B. E., Gardner, A. F., & Ausubel, F. M. (2018). Overview of Next Generation Sequencing Technologies. *Current protocols in molecular biology*, 122(1), e59. <https://doi.org/10.1002/CPMB.59>
- Smit, A., Hubley, R., & Green, P. (2015). *RepeatMasker Home Page*. <https://repeatmasker.org/>
- Sohn, J. II, & Nam, J. W. (2018). The present and future of de novo whole-genome assembly. *Briefings in bioinformatics*, 19(1), 23-40. <https://doi.org/10.1093/BIB/BBW096>
- Spieth, J., & Lawson, D. (2006). *Overview of gene structure*. <https://doi.org/10.1895/wormbook.1.65.1>
- Staden, R. (1979). A strategy of DNA sequencing employing computer programs. *Nucleic Acids Research*, 6(7), 2601-2610. <https://doi.org/10.1093/NAR/6.7.2601>
- Stanke, M., Diekhans, M., Baertsch, R., & Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*, 24(5), 637-644. <https://doi.org/10.1093/BIOINFORMATICS/BTN013>
- Stanke, M., Tzvetkova, A., & Morgenstern, B. (2006). AUGUSTUS at EGASP: using EST, protein and genomic alignments for improved gene prediction in the human genome. *Genome Biology* 2006 7:1, 7(1), 1-8. <https://doi.org/10.1186/GB-2006-7-S1-S11>
- Stanke, M., & Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics*, 19(suppl\_2), ii215-ii225. <https://doi.org/10.1093/BIOINFORMATICS/BTG1080>

- Tarailo-Graovac, M., & Chen, N. (2009). Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences. *Current Protocols in Bioinformatics*, 25(1), 4.10.1-4.10.14. <https://doi.org/10.1002/0471250953.BI0410S25>
- Thomason, A. (1989). A SIMPLE LINEAR EXPECTED TIME ALGORITHM FOR FINDING A HAMILTON PATH. *Discrete Mathematics*.
- Toyokuni, S. (2016). Oxidative stress as an iceberg in carcinogenesis and cancer biology. *Archives of Biochemistry and Biophysics*, 595, 46-49. <https://doi.org/10.1016/J.ABB.2015.11.025>
- Travers, K. J., Chin, C. S., Rank, D. R., Eid, J. S., & Turner, S. W. (2010). A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic acids research*, 38(15). <https://doi.org/10.1093/NAR/GKQ543>
- Treangen, T. J., & Salzberg, S. L. (2011). Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nature Reviews Genetics* 2011 13:1, 13(1), 36-46. <https://doi.org/10.1038/nrg3117>
- Trindade, H., Pedro, L. G., Figueiredo, A. C., & Barroso, J. G. (2018). Chemotypes and terpene synthase genes in Thymus genus: State of the art. *Industrial Crops and Products*, 124, 530-547. <https://doi.org/10.1016/J.INDCROP.2018.08.021>
- Veličković, D. T., Randjelović, N. V., Ristić, M. S., Veličković, A. S., & Šmelcerović, A. A. (2003). Chemical constituents and antimicrobial activity of the ethanol extracts obtained from the flower, leaf and stem of *Salvia officinalis* L. *Journal of the Serbian Chemical Society*, 68(1), 17-24. <https://doi.org/10.2298/JSC0301017V>
- Voorrips, R. E. (2002). MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs. *Journal of Heredity*, 93(1), 77-78. <https://doi.org/10.1093/JHERED/93.1.77>
- Vujošević, M., & Blagojević, J. (2004). Antimutagenic effects of extracts from sage (*Salvia officinalis*) in mammalian system in vivo. *Acta Veterinaria Hungarica*, 52(4), 439-443. <https://doi.org/10.1556/AVET.52.2004.4.6>
- Vuković-Gačić, B., Nikčević, S., Berić-Bjedov, T., Knežević-Vukčević, J., & Simić, D. (2006). Antimutagenic effect of essential oil of sage (*Salvia officinalis* L.) and its monoterpenes against UV-induced mutations in *Escherichia coli* and *Saccharomyces cerevisiae*. *Food and Chemical Toxicology*, 44(10), 1730-1738. <https://doi.org/10.1016/J.FCT.2006.05.011>
- Wang, M., Kikuzaki, H., Zhu, N., Sang, S., Nakatani, N., & Ho, C. T. (1999). Isolation and Structural Elucidation of Two New Glycosides from Sage (*Salvia officinalis* L.). *Journal of Agricultural and Food Chemistry*, 48(2), 235-238. <https://doi.org/10.1021/JF990761P>
- Wang, X., & Wang, L. (2016). GMATA: An integrated software package for genome-scale SSR mining, marker development and viewing. *Frontiers in Plant Science*, 7(September), 215951. <https://doi.org/10.3389/FPLS.2016.01350/BIBTEX>
- Wang, Y., Chen, L., Song, N., & Lei, X. (2015). GASS: Genome structural annotation for Eukaryotes based on species similarity. *BMC Genomics*, 16(1), 1-14. <https://doi.org/10.1186/S12864-015-1353-3/FIGURES/7>
- Wicker, T., Zimmermann, W., Perovic, D., Paterson, A. H., Ganal, M., Graner, A., & Stein, N. (2005). A detailed look at 7 million years of genome evolution in a 439 kb contiguous sequence at the barley Hv-eIF4E locus: recombination, rearrangements and repeats. *The Plant Journal*, 41(2), 184-194. <https://doi.org/10.1111/J.1365-313X.2004.02285.X>
- Xiao, J., Kumarsekhwal, M., Li, P., Raja, R., Sylvie, C., Wang, X., & You, F. M. (2016). Pseudogenes and Their Genome-Wide Prediction in Plants. *International*

- Journal of Molecular Sciences* 2016, Vol. 17, Page 1991, 17(12), 1991.  
<https://doi.org/10.3390/IJMS17121991>
- Xie, C., & Tammi, M. T. (2009). CNV-seq, a new method to detect copy number variation using high-throughput sequencing. *BMC Bioinformatics*, 10(1), 1-9.  
<https://doi.org/10.1186/1471-2105-10-80/FIGURES/6>
- Yandell, M., & Ence, D. (2012). *A beginner's guide to eukaryotic genome annotation*.  
<https://doi.org/10.1038/nrg3174>
- Ye, C., Ma, Z. S., Cannon, C. H., Pop, M., & Yu, D. W. (2012). Exploiting sparseness in de novo genome assembly. *BMC bioinformatics*, 13 Suppl 6(6), 1-8.  
<https://doi.org/10.1186/1471-2105-13-S6-S1/TABLES/6>
- Zerbino, D. R., & Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*, 18(5), 821-829.  
<https://doi.org/10.1101/GR.074492.107>
- Zimin, A., Stevens, K. A., Crepeau, M. W., Holtz-Morris, A., Koriabine, M., Marçais, G., Puiu, D., Roberts, M., Wegrzyn, J. L., de Jong, P. J., Neale, D. B., Salzberg, S. L., Yorke, J. A., & Langley, C. H. (2014). Sequencing and Assembly of the 22-Gb Loblolly Pine Genome. *Genetics*, 196(3), 875-890.  
<https://doi.org/10.1534/GENETICS.113.159715>
- Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S. L., & Yorke, J. A. (2013). The MaSuRCA genome assembler. *Bioinformatics*, 29(21), 2669-2677.  
<https://doi.org/10.1093/BIOINFORMATICS/BTT476>