



T.C.
NECMETTİN ERBAKAN
ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ



**PATATES (*Solanum tuberosum* L.)
TRANSKRİPSİYON FAKTÖRÜ GENLERİNİN
TANIMLANMASI VE REGÜLATÖR
FONKSİYONEL MARKÖRLERİN
GELİŞTİRİLMESİ**

Şerife Nur TOSUN

YÜKSEK LİSANS TEZİ

Biyoteknoloji Anabilim Dalı

**Haziran-2022
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.

Şerife Nur TOSUN

Tarih:

ÖZET

YÜKSEK LİSANS TEZİ

PATATES (*Solanum tuberosum* L.) TRANSKRİPSİYON FAKTÖRÜ GENLERİNİN TANIMLANMASI VE REGÜLATÖR FONKSİYONEL MARKÖRLERİN GELİŞTİRİLMESİ

Şerife Nur Tosun

Necmettin Erbakan Üniversitesi Fen Bilimleri Enstitüsü
Biyoteknoloji Anabilim Dalı

Danışman: Doç. Dr. Ayşe Özgür UNCU

2022, 100

Jüri

Doç. Dr. Ayşe Özgür UNCU

Doç. Dr. Musa SEYMEN

Dr. Öğr. Üyesi Fatih ERCİ

Patates (*Solanum tuberosum* L.) *Solanaceae* familyasından yumruları yenen bir bitki türüdür. Dünyanın ekonomik açıdan en önemli gıda ürünlerinden biridir ve gelecekteki gıda güvenliği için büyük öneme sahiptir. Yetiştirilen patates türlerinin haploid kromozom sayısı $n = 12$ 'dir, haploid genom büyüklüğü 840 Mb'dır ve %73 diploid ($2n = 2x = 24$), %4 triploid ($2n = 3x = 36$), %15 tetraploid ($2n = 4x = 48$) veya %2 pentaploid ($2n = 5x = 60$) türlerini içerir. Patates genetiği ve ıslahı çalışmaları, poliploid genomu ve yüksek heterozigotluk seviyeleri sebebiyle yavaş ilerlemektedir. Yüksek verimli sekanslama teknolojilerindeki son gelişmeler ve biyoinformatik araçların zenginliği, patates de dahil olmak üzere çeşitli türlerin genomik bilgilerinin analizine büyük ölçüde katkıda bulunmuştur. Ek olarak, gen ekspresyon veri tabanları gibi çeşitli halka açık biyolojik veri tabanları, aday genlerin içinde veya yakınında bulunan fonksiyonel markör geliştirmeyi mümkün kılmıştır. Temel enerji kaynağı tarım ürünleri arasında yer alan patates bitkisinin transkripsiyon faktörü kodlayan genlerin sistematik biçimde tanımlanıp karakterize edilmesi önemlidir. Bu tez kapsamında öncelikle patates türüne ait transkripsiyon faktörü kodlayan sekanslar tanımlanmış ve transkripsiyon faktörü aileleri bakımından karakterize edilmiştir. Bu genler üzerinden regülatör fonksiyona sahip olma potansiyeli yüksek DNA markörleri geliştirilmiştir. Bu yeni geliştirilen markör seti tezin temel çıktısı olmaktadır. Geliştirilen markör seti, tür içi ve türler arası polimorfizm düzeyi ve çeşitliliğini araştırmak amacıyla hem patates kromozomlarında hem de yakın akraba tür olan *Solanum lycopersicum* kromozomlarında haritalanmıştır. Polimorfik fonksiyonel markörlerin peptid sekansı düzeyinde protein işlevleri için homoloji temelli yapısal protein modelleri oluşturulmuştur. Bu veri seti de tezin önemli bir çıktısıdır.

Anahtar Kelimeler: Markör destekli seçim, *Solanum tuberosum* L., SSR markörü, Transkripsiyon faktörü

ABSTRACT

MS THESIS

IDENTIFICATION OF TRANSCRIPTION FACTOR GENES AND DEVELOPMENT OF REGULATORY FUNCTIONAL MARKERS IN POTATO (*Solanum Tuberosum* L.)

Şerife Nur TOSUN

THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE OF
NECMETTİN ERBAKAN UNIVERSITY
THE DEGREE OF MASTER OF SCIENCE IN BIOTECHNOLOGY

Advisor: Doç. Dr. Ayşe Özgür UNCU
2022, 101

Jury

Doç. Dr. Öğretim üyesi Ayşe Özgür UNCU
Doç. Dr. Musa SEYMEN
Dr. Öğr. Üyesi Fatih ERCİ

Potato (*Solanum tuberosum* L.) is an important member of the *Solanaceae* family cultivated for its starch-rich tubers. It is one of the world's most economically important food crops and is of great importance for future food security. The basic chromosome number of the cultivated potato species is $n = 12$, with a length of about 840 Mb, and is 73% diploid ($2n = 2x = 24$), 4% triploid ($2n = 3x = 36$), 15% tetraploid ($2n = 4x = 48$) or 2% pentaploid ($2n = 5x = 60$). Potato genetics and breeding studies are progressing slowly due to the polyploid genome and high levels of heterozygosity. Recent advances in high-throughput sequencing technologies and the wealth of bioinformatics tools have greatly contributed to the analysis of genomic information of various species, including potato. In addition, several publicly available biological databases, such as gene expression databases, have made it possible to develop functional markers located in or near candidate genes. It is important that genes encoding transcription factors in potato, which is among the main energy source agricultural products, are systematically identified and characterized. In this thesis work, first of all, potato transcription factor coding genic sequences were identified and characterized in terms of transcription factor families. DNA markers with high potential to have regulatory functions have been developed over these genes. This newly developed marker set is an important output of the thesis. The developed marker set was mapped both in potato chromosomes and in the chromosomes of the closely related species, *Solanum Lycopersium*, in order to investigate the level and diversity of polymorphism within and between species. Homology-based structural protein models have been constructed for protein functions at the peptide sequence level of polymorphic functional markers. This data set is also an important output of the thesis.

Keywords: Marker Assisted Selection, *Solanum tuberosum* L., SSR marker development, Transcription Factor

ÖNSÖZ

Lisans ve Yüksek lisans hayatımızda bize yardımcı olan, çalışmalarımızda bizi sonuna kadar destekleyen ve kıymetli tecrübelerini bizimle paylaşan değerli danışman hocam Dr. Öğr. Üyesi Ayşe Özgür UNCU' ya, maddi manevi desteklerini esirgemeyen çalışmalarımızda bize yol gösteren Doç. Dr. Ali Tevfik UNCU' ya teşekkür ederim.

Tüm bu süreçte bana eşlik eden, güzellikleri ve zorlukları birlikte paylaştığımız takım arkadaşlarım Fatıma ŞEN, Cansu YÜRÜKÇÜ ve Dilay MALYEMEZ'e, desteklerinden, yardımlarından dolayı teşekkür ederim.

Hayatımın her alanında, aldığım kararlarda ve çalışmalarımda maddi manevi yanımda olup beni yetiştiren, gerektiğinde fedakârlık gösteren aileme ve her açıdan beni destekleyen arkadaşlarıma da teşekkür ederim.

Şerife Nur TOSUN
KONYA-2022

İÇİNDEKİLER

| | |
|------------------------------------------------------------------------------------------------------------------------------------|-----------|
| ÖZET | iv |
| ABSTRACT..... | v |
| ÖNSÖZ | vi |
| İÇİNDEKİLER..... | vii |
| SİMGELER VE KISALTMALAR..... | ix |
| ŞEKİL DİZİNİ..... | xii |
| ÇİZELGE DİZİNİ..... | xiv |
| 1. GİRİŞ..... | 1 |
| 2. KAYNAK ARAŞTIRMASI..... | 4 |
| 2.1. Patates (<i>Solanum tuberosum</i> L.)..... | 4 |
| 2.1.1 Patatesin Besin Değeri ve Ekonomik Önemi..... | 4 |
| 2.1.2. Patatesin Kültüre Alınması | 5 |
| 2.1.3. Bitki Islahı..... | 6 |
| 2.1.4. Patates Islahında Moleküler Markörler..... | 8 |
| 2.2. Moleküler Markörler/DNA Markörleri | 9 |
| 2.2.1 PCR Temelli DNA Markörleri | 10 |
| 2.2.2. SSR/Mikrosatellit (Basit Dizi Tekrarı)..... | 10 |
| 2.3. Markör Geliştirme..... | 13 |
| 2.3.1. Moleküler Markör Gelişimi İçin Biyoinformatik | 14 |
| 2.3.2. SSR Markörü Geliştirme | 15 |
| 2.4. Ökaryot Organizmalarda Transkripsiyon | 18 |
| 2.4.1. Ökaryotlarda Korunmuş Başlıca TF Aileleri..... | 20 |
| 2.4.2. Bitki Spesifik TF Aileleri | 28 |
| 3. MATERYAL VE YÖNTEM..... | 36 |
| 3.1. Patates (<i>Solanum tuberosum</i>) Genomunda Transkripsiyon Faktörü Kodlayan Genlerin Tanımlanması ve Karakterizasyonu..... | 36 |
| 3.2. Markör Geliştirme..... | 37 |
| 3.2.1. Patates cDNA sekanslarında SSR Tanımlama..... | 38 |
| 3.2.2. Markör dizaynı..... | 38 |
| 3.2.3. Tür içi ve Türlerarası Genomik Markör Lokus Haritalaması | 39 |
| 3.3. Polimorfik Fonksiyonel Markörlerin Protein Anotasyonu ve 3D Modellemeleri..... | 39 |
| 3.4. DNA İzolasyonu | 40 |

| | |
|--------------------------------------------------------------------------------------|-------------------------------------|
| 3.5. PCR Amlifikasyonu | 40 |
| 3.6. Kapiler Elektforez ve Markör Doğrulama Analizi | 40 |
| 4. ARAŞTIRMA SONUÇLARI VE TARTIŞMA..... | 42 |
| 4.1. Transkripsiyon Faktörü Kodlayan Genlerin Tanımlanması ve Karakterizasyonu | 42 |
| 4.2. SSR Markörlerinin Tespiti | 45 |
| 4.3 Primer Dizaynı | 49 |
| 4.4 Markör Haritalama..... | 50 |
| 4.5. Markör Dizayn ve Haritalama Süreçlerinin Deneysel Doğrulaması | 69 |
| 4.6. Polimorfik Fonksiyonel Markörlerin Protein Anotasyonu ve 3D Modellemeleri..... | 70 |
| 5. SONUÇLAR VE ÖNERİLER | 89 |
| 5.1. Sonuçlar | 89 |
| 5.2. Öneriler | 91 |
| 6. KAYNAKLAR | 93 |
| ÖZGEÇMİŞ | Error! Bookmark not defined. |

SİMGELER VE KISALTMALAR

Simgeler

μ: mikro

°C: Santigrad derece

Kısaltmalar

ABA: Fitohormon Absisik Asit

ABFS: ABA Bağlanma Faktörleri

ABRE: ABA Yanıt Elementi

AFLP: Çoğaltılmış Parça Uzunluk Polimorfizmi

AGRIs: Arabidopsis Gen Düzenleme Sunucusu

AHA: Aktivatör Motifi

APAX2 : Ascorbate Peroksidasez

ATP: Adenozintrifosfat

BHLH: Basic helix-loop-helix

bZİP: Temel Bölge Lösün Zipper

CAPS: Kesilip Çoğaltılmış Parça Uzunluk Polimorfizmi

CPE: Çekirdek Promotör Eleman

cp-SSR: Kloroplast SSR

CREB: CRE Bağlayıcı Protein

DArT: Çeşitlilik Dizileri Teknolojisi

DATF: Arabidopsis Transkripsiyon Faktör Veritabanı

DBD: DNA Bağlanma Domaini

DDBJ: Japonya DNA Veritabanı

DNA: Deoksiribo Nükleik Asit

dNTP: Dinükleotidtrifosfat

DPE: Down Stream Destekleyici Eleman

DREB: Dehidrasyona Duyarlı Element Bağlayıcı Protein

EMBL: Avrupa Moleküler Biyoloji Laboratuvarı
EREB: Etilene Duyarlı Element Bağlayıcı Protein
ESTs: İfade Edilen Dizi Etiketleri
GMATA: Genom Geneli Mikrosatellit Arama Aracı
g-SSR: Genomik SSR
GTF: Genel Transkripsiyon Faktörleri
HSF: Sıcaklık Stresi Faktörü
HTH: Helix-Turn-Helix
InDel: İnsersiyon-Delesyon
ISSR: Basit Tekrarlı Diziler Arası Polimorfizm
JA: Jasmonik Asit
LRP: Lateral Kök Primodiası
MAPK: Mitojenle Aktive Olan Protein Kinazlar
MAS: Markör Destekli Seçim
Med: Aktive Edici Medyator
MISA: MicroSatellite
mRNA: Mesajcı RNA
mt-SSR: Mitekondriyal SSR
MYB: Miyoblastoz
MyoD: Miyoblast Belirleme Faktörü
NCBI: Ulusal Biyoteknoloji Bilgi Merkezi
NES: Nükleer Export Sinyali
NLS: Nükleer Lokalizasyon Sinyali
PIC: Polimorfik Bilgi İçeriği
PlnTFDB: Bitki Transkripsiyon Faktörü Veritabanı
QTL: Kantitatif Karakter Lokusu
RAPD: Rastgele Çoğaltılmış Polimorfik DNA
RARTF: RIKEN Arabidopsis Transcription Factor database
RFLP: Restriksiyon Parça Uzunluk Polimorfizmi

RNA: Ribonükleik Asit
RNAPII: RNA Polimeraz İki
ROS: Reaktif Oksijen Türleri
SA: Salisilik Asit
SNP: Tek Nükleotid Polimorfizmi
TAF: TBP ilişkili Faktör
TAIR: Arabidopsis Bilgi Kaynağı
TBP: TATA Bağlayıcı Protein
TF: Transkripsiyon Faktörü
Ts: Transkripsiyon Başlangıç Bölgesi
 μ l: Mikrolitre
 μ mol/L: Mikromol/litre

ŞEKİL DİZİNİ

| | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Şekil 2. 1. (CA) _n dinükleotid tekrarlarından oluşan SSR markörlerinin temsili görüntüsü | 11 |
| Şekil 2. 2. Sekanslama maliyetinin, teknolojiadaki gelişmelerle zaman içerisinde değişimi Ulusal İnsan Genom Araştırma Entitüsü | 12 |
| Şekil 2. 3. İn Silico yaklaşımlarla DNA markörlerinin geliştirilmesi | 14 |
| Şekil 2. 4. SSR markörlerinin geliştirilmesini gösteren bir şema..... | 16 |
| Şekil 2. 5. Pol II transkripsiyon ön başlatma kompleksinin şematik özet..... | 19 |
| Şekil 2. 6. Çift sarmallı DNA'ya bağlı GCN4 B-ZIP dimerinin X-ışını yapısı..... | 21 |
| Şekil 2. 7. bHLH transkripsiyon faktörlerinin yapısı | 23 |
| Şekil 2. 8. Bitkilerdeki MYB TF'lerin sınıflandırılması ve yapısının gösterimi. Bitkilerde tanımlanan bir ila dört MYB alan tekrarına sahip MYB TF'ler..... | 24 |
| Şekil 2. 9. MYB protein sınıflarının şematik gösterimi ve bitişik MYB tekrarlarının (R) sayısına bağlı olarak işlevleri..... | 25 |
| Şekil 2. 10. Arabidopsis HsfA2'nin fonksiyonel modülleri..... | 27 |
| Şekil 2. 11. Farklı ERF alt ailelerinin şematik gösterimi | 30 |
| Şekil 2. 12. WRKY ailesinin alan yapısı ve proteinlerinin temsili alan organizasyonları | 32 |
| Şekil 2. 13. WRKY transkripsiyon faktörlerinin düzenleyici mekanizması | 33 |
| Şekil 2. 14. NAC proteinlerinin yapısı | 34 |
| | |
| Şekil 3. 1. Uzun bir DNA dizisi için GMATA'daki SSR arama stratejisini gösteren şematik diyagram..... | 38 |
| | |
| Şekil 4. 1. Yüksek çözünürlüklü kapiler elektroforez ile görüntülenen bazı polimorfik SSR markörleri | 70 |
| Şekil 4. 2. Kromozom 1 üzerinde haritalanan >MK90 markörünün üç boyutlu protein modeli | 75 |
| Şekil 4. 3. Kromozom 2 üzerinde haritalanan >MK345 markörünün üç boyutlu protein modeli..... | 76 |
| Şekil 4. 4. Kromozom 3 üzerinde haritalanan >MK708 markörünün üç boyutlu protein modeli..... | 77 |
| Şekil 4. 5. Kromozom 4 üzerinde haritalanan >MK657 markörünün üç boyutlu protein modeli..... | 78 |
| Şekil 4. 6. Kromozom 5 üzerinde haritalanan >MK166 markörünün üç boyutlu protein modeli | 79 |
| Şekil 4. 7. Kromozom 6 üzerinde haritalanan >MK752 markörünün üç boyutlu protein modeli..... | 80 |
| Şekil 4. 8. Kromozom 7 üzerinde haritalanan >MK775 markörünün üç boyutlu protein modeli..... | 81 |
| Şekil 4. 9. Kromozom 8 üzerinde haritalanan >MK417 markörünün üç boyutlu protein modeli..... | 82 |
| Şekil 4. 10. Kromozom 9 üzerinde haritalanan >MK304 markörünün üç boyutlu protein modeli..... | 83 |
| Şekil 4. 11. Kromozom 10 üzerinde haritalanan >MK376 markörünün üç boyutlu protein modeli..... | 84 |

| | |
|-------------------------------------------------------------------------------------------------------|----|
| Şekil 4. 12. Kromozom 11 üzerinde haritalanan >MK810 markörünün üç boyutlu protein modeli..... | 85 |
| Şekil 4. 13. Kromozom 12 üzerinde haritalanan >MK810 markörünün üç boyutlu protein modeli..... | 86 |

ÇİZELGE DİZİNİ

| | |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Çizelge 4. 1. Blastx analizinde kullanılan sekans bilgileri. | 43 |
| Çizelge 4. 2. Patates cDNA sekansları ile Arabidopsis TF gen setleri arasında yapılan Blastx analizi sonucunda elde edilen e- Değeri ve Bit skor özet tablosu. | 43 |
| Çizelge 4. 3. Patates cDNA sekansları ile Pirinç TF gen setleri arasında yapılan Blastx analizi sonucunda elde edilen e- Değeri ve Bit skor özet tablosu. | 43 |
| Çizelge 4. 4. Blastx Analizi sonuçları | 44 |
| Çizelge 4. 5. SSR motiflerinin cDNA sekanslarında bulunma sayısı ve frekansı..... | 46 |
| Çizelge 4. 6. Seçilen <i>Solanum Tuberosum</i> cDNA sekanlarında bulunan SSR motifleri, bulunma sayıları ve frekansları verilmiştir. | 48 |
| Çizelge 4. 7. Markör dizaynı istatistikleri. | 49 |
| Çizelge 4. 8. Patates ve Domates kromozomlarında haritalanan markör sayıları | 50 |
| Çizelge 4. 9. Patates kromozomlarında yapılan E-map sonucu | 51 |
| Çizelge 4. 10. Dizayn edilen Primer dizileri. | 52 |
| Çizelge 4. 11. Markör lokuslar tarafından kodlanan peptit sekanslarının 3D modelleme (üç boyutlu modelleme) sonuçları. | 72 |
| Çizelge 4. 12. Seçilen markörlerin sekanslarında yapılan modifikasyon sonrası modelleme değerlerindeki değişimler..... | 87 |

1. GİRİŞ

Patates (*Solanum tuberosum* L.) Solanaceae familyasından yumruları yenen bir bitki türüdür. Yer altı depo organı olan yumrular eşeysiz üremesini sağlar. En önemli yetiştirilme sebebi nişastasıdır ve yüksek karbonhidrat içeriği nedeniyle enerji sağlar ayrıca tüketicilerin beslenmesine ve sağlığına katkıda bulunan polifenoller karotenoidler, E ve C vitaminleri gibi mineraller, lifler, proteinler ve antioksidan bileşikler içerir (Calliope ve ark.,2018). Patates (*Solanum tuberosum* L.), dünyanın ekonomik açıdan en önemli gıda ürünlerinden biridir. Küresel ısınmanın etkisini fazlaca gördüğümüz bu çağda ve gelecekte gıda güvenliği için büyük önem taşımaktadır. Patatesin geniş tarımsal ekosistem yelpazesi ve veriminin yüksek olması, özellikle üretim ve beslenme seçeneğinin az ya da hiç bulunmadığı bölgelerde tüketim için patatesi en iyi seçenek haline getirmektedir.

Solanum tuberosum L. türü içinde taksonomik gruplardan oluşan bir koleksiyon bulunmaktadır. Bu gruplardan birkaçı olan Andigena, Phureja ve Stenotomum, Güney Amerika'nın And bölgesinde kültüre alınmıştır ve çoğu, Kuzey Amerika ve Avrupa'daki tipik yaz aylarının uzun gün koşulları altında yumrular oluşturamaz. Buna karşılık, *Solanum tuberosum* L. dünya çapında yaygın olarak yetiştirilmektedir ve uzun gün koşullarında yumru kökler oluşturmak üzere ıslah edilmiştir. Neredeyse tüm patates çeşitleri, germplazm bakımı ve tarımsal üretim için vejetatif olarak çoğaltılan heterozigot ototetraploidlerdir ($2n = 4x = 48$) (Hirsch ve ark.).

Patates ıslahı çalışmaları, esas olarak poliploid genomu ve yüksek heterozigosite seviyeleri nedeniyle geride kalmıştır. Geleneksel markör ve genotipleme yaklaşımları, patates genetik araştırmalarının ilerlemesine destek olmuştur, ancak aynı zamanda, gen keşfi ve uygulamalı araştırma uygulamaları için bu çalışmalardan elde edilen sonuçların kullanılmasında sınırlamalara sahiptir. Patates genomunun sekanslanmasının, ardından markör ve genotipleme teknolojilerindeki gelişmeler, patates genetik çalışmalarının yürütülme işlemlerinde yöntem değişiklikleri sağlamıştır. Önceden bilgi sahibi olmadan, herhangi bir popülasyonda binlerce markör haritalama yeteneklerinin artmasıyla artık modern dizi bazlı markör ve genotipleme yöntemlerine patates de uygun hale gelmiştir. Bu gelişme, daha önce patatesten mümkün olmayan genetik çalışmaların kesinliğini ve çözünürlüğünü artırmıştır (Sharma ve ark., 2017).

Genetik çeşitliliğin değerlendirilmesi, bağlantı haritasının oluşturulması, genetik saflık testi, kantitatif karakter lokuslarını (QTL) haritalama ve markör destekli seçim (MAS) gibi çeşitli genetik ve ıslah uygulamalarında çalışmalarının etkinliğini artırmaya yönelik, ilgili genlere bağlı moleküler markörler geliştirilebilmesi son derece önemlidir.

Farklı iklim koşullarına dayanıklı patates çeşitlerinin ıslahını sağlamak yeni ve faydalı özellik alellerini belirlemek, bu önemli mahsulün mevcut genetik çeşitliliğini taramak için markör destekli ıslah (MAS) çalışmaları yürütülmektedir. Bir genomdaki markör alellerindeki çeşitlilik, kantitatif özelliklerin genotipik değişkenliğinin araştırılmasını sağlar.

Moleküler markörler bir popülasyonu oluşturan bireylerin, genetik materyalini oluşturan DNA'larının baz dizilimindeki farklılıkları değişik yollarla ortaya koyan ve nesiller boyu aktarılan nükleotid dizileridir. Hiperdeğişkenlik, multialelik özellik, kodominant kalıtım, tekrarlanabilirlik, organel genomları dahil geniş genom kapsama alanı ve yüksek verimli genotipleme gibi genetik özellikler sebebiyle bitki ıslah çalışmalarında en çok tercih edilen moleküler markörler SSR (Basit dizi tekrarı) markörleridir (Kalia ve ark.,2011).

Yüksek verimli sekanslama teknolojilerindeki son gelişmeler ile biyoinformatik araçların zenginliği, patates de dahil olmak üzere çeşitli türlerin genomik bilgilerinin analizine büyük ölçüde katkıda bulunmuştur. Ek olarak, gen ekspresyon veri tabanları gibi çeşitli halka açık veri tabanları, aday genlerin içinde veya yakınında bulunan fonksiyonel markör geliştirmeyi mümkün kılmıştır.

Gen ifadesinin düzenlenmesinde rol alan Transkripsiyon faktörleri (TF) sekansa özgü bir şekilde hedef genlerin promoterlerine doğrudan bağlanan proteinler olarak tanımlanır. Mikroarray analizi ve büyük ölçekli transkriptom analizi gibi moleküler tekniklerle, bitkilerde çok sayıda abiyotik strese yanıt veren genler tanımlanmıştır. Bu genler, önemli enzimlerin ve fonksiyonel proteinler üretilmesiyle hücrelerin stresten korunmasında rol oynamakla kalmaz, aynı zamanda stres yanıtında sinyal iletimini ve gen ifadesini düzenlemede de rol oynar (Guo ve ark., 2016).

Temel enerji kaynağı tarım ürünleri arasında yer alan patates bitkisinin transkripsiyon faktörü kodlayan genleri sistematik biçimde tanımlanıp karakterize edilmiş değildir. Bu genleri tanımlamak, transkripsiyon faktörü aileleri bakımından

karakterize etmek ve bu genler üzerinden regülatör fonksiyona sahip olma potansiyeli yüksek DNA markörleri geliřtirmek, patatesin polimorfizm seviyesi ve çeřitlilięi hakkında bilgiler saęlamak, patates yetiřtirme programlarını hızlandırmak için önemli ve güvenilir bir kaynak oluřturacaktır.

2. KAYNAK ARAŞTIRMASI

2.1. Patates (*Solanum tuberosum* L.)

Solanum tuberosum L., Şili'de ve yüksek And dağlarında yetişen yerli ekili popülasyonlar için geleneksel olarak kullanılan isimdir. *Solanum tuberosum* adı, Güney Amerika'nın dışındaki on altıncı yüzyılın sonundan bu yana yetiştirilen patates çeşitleri için de kullanılır. Modern çeşitler, farklı kültür grupları ve yabancı türler arasındaki kapsamlı ıslahın ürünleridir. Hepsi *Solanum* olarak sınıflandırılan yerli patatesleri ve yabancı patates türleri, patates ıslahı için yaygın olarak kullanılmaktadır. *Solanum* üyeleri Amerika'nın güneybatısından Güney Amerika'nın Güney Konisine kadar geniş bir şekilde dağılmıştır (Spooner ve ark., 2014).

Patates; biber, patlıcan, domates, petunya ve tütün gibi ekonomik açıdan önemli bitkileri içeren *Solanaceae* ailesine aittir. Patates yumruları, bu türün depolama organlarıdır ve ayrıca patatesler vejetatif olarak çoğalır (Monte ve ark., 2018).

Yaygın olarak yetiştirilen patates çeşitleri arasında, diploid ($2n = 2x = 24$), triploid ($2n = 3x = 36$), tetraploid ($2n = 4x = 48$) veya pentaploid ($2n = 5x = 60$) seviyelerinde ekili türler varken, temel kromozom sayısı 12 olan tetraploid ($2n = 4x = 48$) çeşidi bulunmaktadır. *Solanum tuberosum* 12 haploid kromozom sayısı ile yaklaşık 840Mb genom boyutuna sahiptir. Yaklaşık 8000 yıl önce ekiminin başladığı bilinen patates And bölgesinde kökeninden ve kültüre alınmasına ve dünyanın üçüncü en önemli gıda mahsulü haline gelene kadar izlediği yol, kitaplarda, makalelerde ve denemelerde kapsamlı bir şekilde belgelenmiştir (Ortiz ve Mares, 2017).

2.1.1 Patatesin Besin Değeri ve Ekonomik Önemi

Patates yumruları yaklaşık olarak %80 su ve %20 kuru maddeden oluşmakta olup, bu oran çeşide bağlı olarak değişebilir. 100 gramlık bir yumrudaki 20 gram kuru maddenin yaklaşık 18 gramı karbonhidrat ve 2 gram proteindir. Yumrulardaki birincil depolama proteinleri, çözünen protein içeriğinin %40'ını oluşturan patatinlerdir (Navarre ve ark., 2009).

Patates özellikle yüksek kalitede proteinler, karbonhidratlar, C vitamini, B6 vitamini, B3 vitamini ve potasyum, fosfor ve magnezyum gibi bazı minerallerin kaynağı

olarak kabul edilmektedir. Bu temel besinlerin ötesinde, patateslerin önemli miktarda antioksidan fitokimyasallar içerdiği bulunmuştur (Calliope ve ark.,2018).

Sebze ve meyveler arasında, patateslerdeki C vitamini konsantrasyonu orta ila düşük arasında değişmektedir. Orta derecede içeriğine rağmen, patatesler dünyanın birçok yerindeki insanların beslenmelerinde C vitaminine ihtiyacını karşılayan önemli bir gıdadır. Tek bir 148 g'lık porsiyon pişmiş patates, önerilen günlük C vitamini alımının yaklaşık %45'ini sağlayabilir (Love ve Pavek, 2008).

Patates, farklı diyet minerallerinin önemli bir kaynağıdır. Potasyumun günlük ihtiyacının %18'ini, demir, fosfor ve magnezyumun %6'sını, kalsiyum ve çinkonun %2'sini sağlayan besin olarak listelenmiştir (Navarre ve ark., 2009).

Patatesin besin kalitesi, büyüebildiği geniş tarımsal ekosistem yelpazesi ve birim arazi başına düşen yüksek kuru madde ve besin çıktıları, bu mahsulü insanlığın gıda ve beslenme güvenliği için en iyi seçenek haline getiriyor. Özellikle çok fazla üretkenlik ve beslenme seçeneğinin bulunmadığı bölgelerde, dünyanın hem gelişmekte olan hem de gelişmiş bölgelerindeki yayla tarımı için bu geçerlidir. Gelişmekte olan bölgelerde artan tüketim eğilimleri ve dünya çapında temel gıda olarak patatesin artan önemi nedeniyle, besleyici değeri hem gelişmiş hem de gelişmekte olan ülkelerde insan sağlığını iyileştirmenin bir aracı olarak önemli bir potansiyele sahiptir.

Patates dünyanın ekonomik açıdan en önemli gıda bitkilerinden biridir ve gelecekteki gıda güvenliği için büyük önem taşır. Önemli olmasına rağmen, patates genetiği ve ıslah çalışması, çoğunlukla poliploid genomu ve yüksek düzeyde heterozigotluk seviyesi nedeniyle geride kalmıştır (Sharma ve Bryan, 2017).

Dünya çapında en önemli tahıl dışı temel üründür ve geleneksel olarak pirinç, bakliyat ve tatlı patates gibi diğer kök ve yumrulu mahsullerin yetiştiği Asya ve gelişmekte olan ülkelerde giderek daha fazla yetiştirilmektedir. Çin 91.881 ton ile patates üretiminde dünyada ilk sırada her alırken Hindistan (50.190 ton), Ukrayna (20.269 ton), Rusya (20.074 ton) ve ABD (19.181 ton) onu takip eden ülkelerdir (FAOSTAT,2019).

2.1.2.Patatesin Kültüre Alınması

Solanum tuberosum adı, Güney Amerika'nın dışında, on altıncı yüzyılın sonundan bu yana dünyada yetiştirilen patates çeşitleri için de kullanılır. Modern çeşitler, farklı kültür grupları ve yabani türler arasındaki kapsamlı ıslahın ürünleridir. Yerel patatesleri

ve yabani patates türleri, patates ıslahı için yaygın olarak kullanılmaktadır (Spooner, 2014).

Yerel patates çeşitleri iki geniş coğrafi bölgede bulunur: batı Venezüella'nın güneyinden kuzey Arjantin'e kadar 3.000-4.000 m'de yüksek And Dağları'nda bulunan diploidler $2n = 24$, triploidler $2n = 36$ ve tetraploidler $2n = 48$ çeşitleri bulunan *Solanum tuberosum* Andigenum Group, "Andigenum" ve güney-orta Şili'de Chonos ve Guaitecas takımadaları ve yakındaki anakarada, deniz seviyesinden 500 m yükseklikte çoğunlukla tetraploidler içeren *Solanum tuberosum* Chilotanum Grubu, "Chilotanum" (Spooner 2012).

Solanum tuberosum, 8.000-10.000 yıl önce Güney Peru'nun And Dağları'na özgü yabani türlerden ($2n = 2x = 24$) kültüre alınmıştır ve Peru, Bolivya ve Ekvador 'nun dağlık bölgelerinde yaşayan eski toplumlar için gıda güvenliği ve kültürel mirasın bir kaynağı haline gelmiştir (Hardigan, 2017).

Hawkes (1990) tarafından yapılan kapsamlı taksonomi çalışması, 228 yabani ve 7 kültüre alınmış patates türü olmak üzere toplam 235 patates türü olduğunu bildirmiştir. O zamandan beri çok çeşitli türleri kapsayan çok sayıda örnekle gelişmiş moleküler tekniklerle yapılan taksonomik çalışmalar taksonomik sınıflandırmanın yeniden gözden geçirilmesi gerektiğini göstermiştir (Spooner 2014; Machida Hirano 2017).

Spooner ve ark. (2007, 2014) ve Ovchinnikova ve ark. (2011), saha çalışmaları ve entegre morfolojik ve moleküler çalışmalardan elde edilen verileri kullanarak, Hawkes (1990) tarafından tanımlanan toksonomik sınıflandırmayı, 107 yabani ve 4 kültüre alınmış yerel türe ait patates olarak türlerinin sayısını yarıdan fazla azaltarak sınıflandırmıştır.

Otopoliploidinin ortaya koyduğu birçok genetik karmaşıklık nedeniyle, patates genetik çalışmalarının çoğu, iki ebeveynli popülasyonlar kullanılarak diploid düzeyde gerçekleştirilmiştir. Tetraploidi ($2n = 4x = 48$) ve inbreeding depresyonu ile karakterize olan mevcut ıslah sistemi, üstün alel kombinasyonlarının homozigot halde sabitlenmesine izin vermez (D'hoop ve ark., 2014).

Gelişen moleküler biyoloji teknikleri, klasik ıslah yöntemlerinin yanı sıra moleküler ıslah tekniklerinin de gelişmesini sağlamıştır.

2.1.3. Bitki Islahı

Bitki ıslahı, en genel tanımıyla istenilen özelliklere sahip yeni bitki çeşitleri elde etmek amacıyla yapılan melezleme ve seleksiyon çalışmalarının tamamıdır. Bitki ıslahı insanlara ve hayvanlara yönelik besinlerin kalitesini arttırmayı ve çeşitli tarımsal uygulamalar için benzersiz üstün özellikte mahsul üretmeyi amaçlar. İnsan beslenmesi için kullanılan bitki türevi ürünlerin iyileştirilmesi, evcil hayvanların beslenmesi veya lif üretimi için bitki ıslahı binlerce yıldır yapılmaktadır. Geleneksel ıslahta bu, örneğin daha üretken veya patojene dirençli bitkiler elde etmek için esas olarak üstün bitkilerin diğer uyumlu bitkilerle çaprazlanmasıyla çok uzun bir süredir gerçekleştirilir (Hartung., ve Schiemann., 2014).

Islah programları genellikle tohum kalitesi ve miktarı, biyokütle üretimi ve haşere ve hastalık direnci gibi temel tarımsal özellikleri optimize etmeye çalışır. En sık ele alınan özellikler, biyotik ve abiyotik stres toleransı, tahıl verimi, tat veya spesifik biyolojik moleküllerin (proteinler, şekerler, lipidler, vitaminler, lifler) konsantrasyonları gibi son kullanım kalite özellikleri ve işleme kolaylığı ile ilgili özelliklerdir.

Şimdiye kadar, ıslah programları, çeşitleri seçmeye ve diğer çeşitlerle birlikte yetiştirildiğinde ortalama bir performans gösterme yeteneğine dayalı çeşitler yaratmaya çalışmıştır. İlişkili olarak gerçekleştirilmeye yönelik bu ortalama yeteneği tahmin etmek için, kullanılan yaklaşım, hatların hibritleşme yeteneğinin belirlenmesi için mısır (*Zea mays*) ile yapılan diallel testlerinden türetilmiştir. Özellikle bu, çok hatlı ve çok türlü karışımlar dahil olmak üzere çok genotipli karışımları geliştirmek için kullanılmıştır. Ancak, bu yaklaşımın belirli test materyalinin ötesinde ve belirli bir çalışma bağlamında sınırlı taşınabilirliği vardır. Bir dereceye kadar bu sınırlamalar, büyük numune boyutları ve geniş bir genotip yelpazesi kullanılarak dengelenebilir (Litrice., 2015).

Klasik ıslah yöntemlerine ek olarak daha kısa zamanda sonuç veren ve daha ekonomik olan moleküler ıslah yöntemleri geliştirilmiştir. Moleküler markörler bu yöntemlerin temelidir ve son 50 yılda bitki biyoteknolojisinin gelişmesine önemli katkılar sağlamıştır. Moleküler ıslah yöntemleri bitki ıslahında, genotipe dayanarak fenotipi tahmin etmek ve geliştirilmiş çeşitleri seçmek için kullanılabilir. Kalıtsal genetik faktörler ve ortaya çıkan fenotipler arasındaki bağlantıya dair anlayışımızı ilerletmek, azalan ekilebilir arazi ve iklim değişikliği karşısında küresel gıda kaynaklarını artırmak için gereken ölçekte genomik destekli ıslahın var olmasını sağlayacaktır. Bitki ıslahı programlarında üzerinde çalışılan karakterler, birçok gen tarafından ve çevre faktörlerinin etkisiyle kontrol edilen büyük kısmı kantitatif olan karakterlerdir. Gözlenen fenotip üzerine bu genlerin etkisi kalitatif karakterlere kıyasla küçüktür. Örneğin bitkilerde verim

ve tohum iriliği kantitatif karakterlerin tipik örneğidir. Klasik ıslah programlarında seleksiyon, gözlemlenebilen aday fenotiplerin seçilmesi şeklinde yapılmakta ancak gerçekte hangi genlerin seçildiği bilinmemektedir. Moleküler markörlerin geliştirilmesi bu sınırlandırmayı ortadan kaldıran önemli bir keşif olmuştur.

Patates ıslahında moleküler markörlerin kullanılması, genotiplerin seçimi için yeni fırsatlar sunmaktadır. Bugüne kadar, faydalı özelliklerle bağlantılı birçok markör bulunmuştur. Patates haritasında, çoğu haşere dirençli genler olan 25'ten fazla dominant gen, verim ve yumru kalitesi özelliklerini kontrol eden bazı kantitatif karakter lokusları (QTL) ile birlikte lokalize edilmiştir. Bu genlere bağlı moleküler markörler artık hızlı, pozitif yardımcı seçim için mevcuttur. Ayrıca, yararlı genleri transfer etmek için, vahşi ve ekili genotipler arasında spesifik olmayan çaprazlamalar sıklıkla gerçekleştirilir ve bu durumda türe özgü moleküler markörlerin kullanımı, vahşi genomik içeriğin birkaç geri çapraz nesilde azaltılmasına yani negatif destekli seçime izin verir (Barone., 2004).

Bitki genom analizi için geliştirilen çok sayıda moleküler markörün geliştirilmesinden bu yana, fenotip yerine genotipi seçme olasılığı yakından incelenmiş ve moleküler markör destekli seçim (MAS) kavramı ortaya çıkmıştır. Restriksiyon parça uzunluk polimorfizmi (RFLP), rastgele çoğaltılmış polimorfik DNA (RAPD), çoğaltılmış parça uzunluk polimorfizmi (AFLP) ve basit dizi tekrarı (SSR), MAS için kullanılan markörler arasındadır. Markör lokusların alelleri, evrim sürecinde hayatta kalma başarısını olumsuz etkilemeden biriken nokta mutasyonu veya kromozom yeniden düzenlemeleridir.

MAS için en uygun markörlerin seçimi farklılık gösterebilir ve bunların tespiti gereken iş gücüne, tek veya çoklu lokusların ortaya çıkma olasılığına, kodominant veya dominant doğaya ve maliyetlere bağlıdır. MAS, farklı genotiplerin çaprazlanmasından sonra meydana gelen doğal DNA varyasyonlarının analizinde yer aldığı için, biyoteknolojinin bitki ıslahında en verimli uygulamalarından biridir (Barone., 2004).

2.1.4. Patates Islahında Moleküler Markörler

Patates ıslahında moleküler markörlerin kullanımı, çeşit tanımlaması (Gebhardt ve ark. 1989a), filogenetik çalışmalar (Kardolns ve ark., 1998), genomlar arası rekombinasyon analizi (Williams ve ark. 1993), özellikleri kontrol eden genler tanımlama (Gebhardt 1994)) gibi birçok amaç için rapor edilmiştir.

Moleküler markörler ile farklı türler arasında faydalı genlerin transferini amaçlayan ıslah programlarında ilk adım, transferi yapılacak genlere bağlı bir veya daha fazla markörün tanımlanması ve moleküler harita üzerinde lokalizasyonun belirlenmesidir.

Patates genom haritası, farklı moleküler markörlere sahip zengin haritalardan biridir ve bu nedenle, MAS için kapsamlı fırsatlar sunar. Başlangıçta, farklı genetik arka planlar üzerinde RFLP markör lokusları takip edilerek eşzamanlı olarak oluşturulmuş iki patates haritası oluşturulmuştur (Bonierbale ve ark., 1988; Gebhardt ve ark., 1989b).

Bu haritalar daha sonra birbirleriyle karşılaştırılmış ve ayrıca domates RFLP haritası ile hizalanmıştır (Gebhardt ve ark., 1991; Tanksley ve ark., 1992).

Yeni moleküler markörlerin geliştirilmesiyle patates haritası oldukça zenginleştirilmiş ve patates genomunun yaklaşık %90'ını kapsayan 350'den fazla marköre sahip, faydalı özelliklerin ifadesini kontrol eden genleri lokalize etmek için değerli bir araç haline getirilmiştir (Gebhardt ve ark., 2001).

Son yıllarda bitki genetik koleksiyonlarındaki genetik değişkenliği ölçmek için çeşitli moleküler teknikler geliştirilmiştir. Moleküler markörler, çeşitliliği belirlemek ve biyolojik çeşitliliği tahmin etmek için birçok fayda sağlayabilir. Bu teknikler, araştırmacıların, taksonomik düzeyde girişleri belirlemelerine, türler içindeki ve türler arasındaki nispi çeşitliliği değerlendirmelerine ve ıslah amacıyla çeşitli girişleri belirlemelerine olanak tanır.

2.2. Moleküler Markörler/DNA Markörleri

Genetik markör, belirli bir geni veya özelliği kontrol eden, bilinen bir kromozom konumuna sahip bir gen veya DNA dizisidir. Genetik markör kavramı yeni bir kavram değildir, Gregor Mendel, 19. yüzyılda yaptığı deneyde fenotip tabanlı genetik markörleri kullanmıştır. Daha sonra, Drosophila için fenotip tabanlı genetik markörler genetik bağlantı teorisinin kurulmasına yol açmıştır. Fenotip temelli genetik markörlerin eksiklikleri, moleküler markör olarak bilinen daha genel ve kullanışlı doğrudan DNA temelli markörlerin geliştirilmesini sağlamıştır. Genetik markörler genel olarak iki kategoriye ayrılır: klasik markörler ve DNA / moleküler markörler. Morfolojik, sitolojik ve biyokimyasal markörler klasik markör tipleridir (Agarwal ve ark., 2008; Nadeem ve ark., 2018).

Moleküler bir markör, genom düzeyindeki farklılıkları temsil eden belirli bir DNA segmenti olarak tanımlanır. Moleküler markörler nükleotit sekanslarıdır ve farklı bireylerin nükleotit sekansları arasında bulunan polimorfizm yoluyla araştırılabilir. İnsersiyon, delesyon, nokta mutasyonları duplikasyon ve translokasyon bu polimorfizmlerin temelidir, ancak mutasyonlar genlerin aktivitesini etkilemeyebilir (Nadeem ve ark., 2018).

Moleküler markörler, bir özelliğin fenotipik ekspresyonu ile ilişkili olabilir veya olmayabilir, hücrenin büyümesi, farklılaşması, gelişimi veya savunma durumundan bağımsız olarak tüm dokularda kararlı ve saptanabilir olduklarından, geleneksel fenotip bazlı alternatiflere göre çok sayıda avantaj sunarlar. Temel moleküler markör teknikleri hibridizasyon temelli markörler ve PCR temelli markörler olarak iki kategoriye ayrılabilir.

2.2.1 PCR Temelli DNA Markörleri

PCR tekniği, herhangi bir canlı organizmanın az miktarda DNA 'sını amplifiye edebilen bir teknik olarak 1983 yılında Cary Mullis tarafından geliştirilmiştir (Mullis ve ark., 1986). Genel olarak PCR, DNA'nın in vitro koşullarda enzimatik olarak sentezlenmesi işlemidir. PCR'da kullanılan temel bileşenler; çoğaltılması istenen kalıp DNA veya RNA, taq DNA polimeraz enzimi, primerler, dinükleotidtrifosfat (dNTP), tampon sıvı, tamponlar ve Mg iyonlarıdır. PCR tekniğinde üç ana adım vardır: DNA ipliklerinin birbirinden ayrılarak açılması (Denatürasyon), primerin bağlanması (Annealing) ve uzama (Extension).

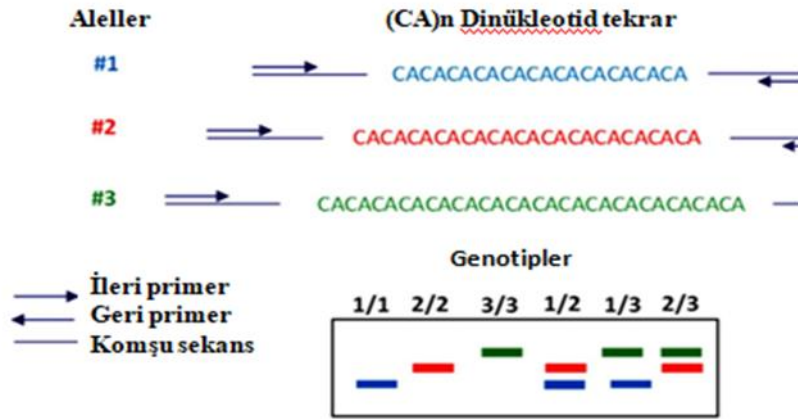
DNA dizilişi her genotipte farklı olduğu için, aynı primerler kullanılsa bile her genotipte farklı DNA ürünleri elde edilir ve bu farklı ürünler genetik markör olarak kullanılabilir. Günümüzde PCR'a dayalı olarak geliştirilen PCR-RFLP, kesilip çoğaltılmış parça uzunluk polimorfizmi (CAPS), RAPD, SSR, AFLP, basit tekrarlı diziler arası polimorfizm ISSR gibi belli başlı markör teknikleri kullanılmaktadır.

2.2.2. SSR/Mikrosatellit (Basit Dizi Tekrarı)

Mevcut moleküler markörler arasında, 1-6 nükleotit uzunluğunda DNA motifinin ardışık tekrarları olan SSR'lar veya basit dizi tekrarları hiperdeğişkenlik, multialelik doğa, kodominant kalıtım, tekrarlanabilirlik, organel genomları dahil geniş

genom kapsama alanı, ve yüksek verimli genotipleme gibi genetik özellikler sebebiyle bitki genetiğinde ve ıslahında önem kazanmıştır (Kalia ve ark., 2011; Zhao ve ark., 2014).

(CA) n ve (GA) n gibi dinükleotid tekrarları en bol tekrarlardır. İnsanlarda (CA) n tekrarı her 30 kb'de bir meydana gelir. PCR primer çiftleri, SSR'ların yan bölgelerinin bağlanır ve PCR amplifikasyonundan sonra agaroz veya poliakrilamid jellerde gözlenir. Şekil 2.1' de SSR'ların temsili bir görüntüsü verilmiştir.



Şekil 2. 1. (CA)_n dinükleotid tekrarlarından oluşan SSR markörlerinin temsili görüntüsü (Alzohairy., 2015).

SSR markörleri tarafından ortaya konan yüksek dereceli alelik varyasyon, mayoz sırasında replikasyon kayması veya eşit olmayan krossingoverın neden olduğu bir lokusta tekrar motif sayısındaki değişiklikten kaynaklanır. Bitki genleri içindeki SSR motiflerinin fonksiyonlarının anlaşılması kısıtlı olmasına rağmen, SSR'lar bitki genom analizinde yaygın olarak kullanılmaktadır. Kodlama bölgelerindeki SSR'lerin varlığı, aminoasit sekanslarında tekrarlayan modellerin ortaya çıkmasına neden olur ve bu nedenle gen ifadenmesinin veya moleküler fonksiyonların düzenlenmesini hakkında bilgi verir (Senan, ve ark., 2014).

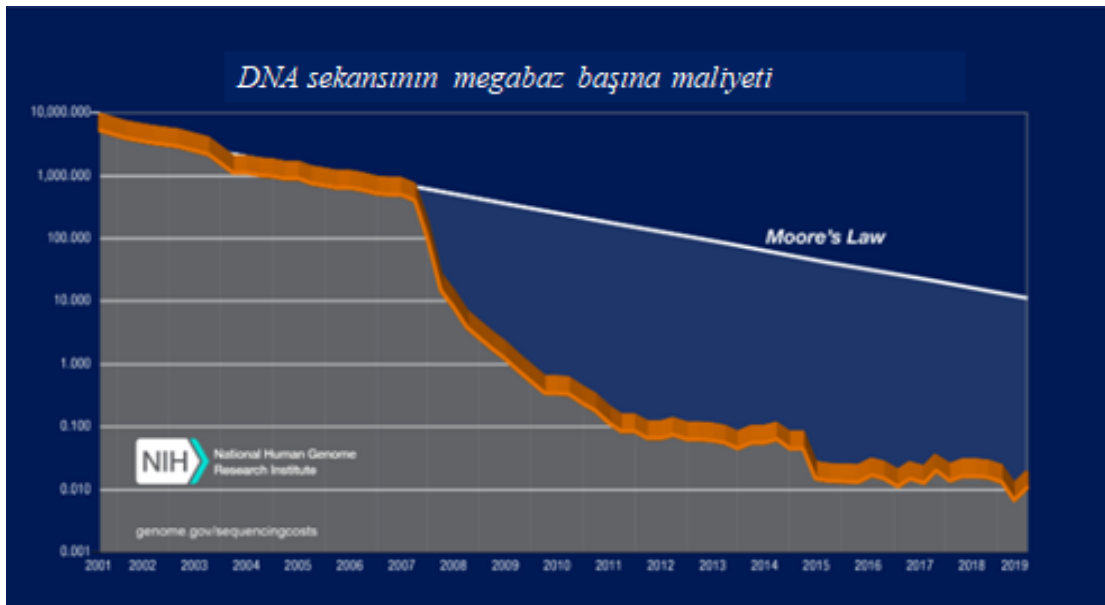
SSR'lar doğrudan genomik DNA kütüphanelerinden veya spesifik SSR'lar için zenginleştirilmiş kütüphanelerden geliştirilebilir. Alternatif olarak, SSR'lar, GenBank ve EMBL gibi halka açık veritabanları araştırılarak veya türler arası aktarılabirlik yoluyla da bulunabilir.

Günümüzde, EST veritabanları önemli bir aday gen kaynağıdır, çünkü bunlar doğrudan ilgilenilen bir özellik ile ilişkili markörler geliştirilebilir ve yakın akraba türlerine aktarılabir. Çok sayıda SSR tabanlı teknik geliştirilmiştir (Kalia ve ark., 2011).

Aynı zamanda son yıllarda yapılan bazı çalışmalar Transkripsiyon faktörü (TF) genlerine dayanan kodominant SSR markörlerinin, QTL haritalaması ve markör destekli seleksiyon için yararlı araçlar olarak hizmet edebileceğini göstermiştir.

1980'lerde moleküler markör teknolojisinin gelişmesiyle, bitki ıslahının kaderi değişmiştir. Farklı tipte moleküler markörler geliştirilmiş ve sekanslama teknolojilerindeki ilerlemeler ürün gelişimini hızlandırmıştır. Moleküler bitki ıslahı, genetik, genomik seleksiyon ve genom düzenlemesinde kaydedilen ilerleme, moleküler markörlerin daha kapsamlı bir şekilde anlaşılmasına katkıda bulunmuş, bitkiler için çeşitlilik ve büyük ölçüde tamamlanmış yetiştirme stratejileri hakkında daha derin bilgiler sağlamıştır.

Sanger sekanslama teknolojisinden sonra, birinci nesil sekanslamanın düşük iş hacmi ve yüksek maliyeti, metodolojide temel bir değişime yol açarak 2. nesil ve 3. nesil olarak bilinen yeni nesil sekanslama teknolojilerinin gelişmesini sağlamıştır. Yeni nesil sekanslama teknolojileri ile birim zamanda elde edilen sekanslama verileri artarken sekanslama maliyeti de azalmıştır. Şekil 2.2. 'de zaman içerisinde değişen sekanslama maliyeti gösterilmiştir. Yeni nesil sekanslama teknikleri ile, aynı anda milyonlarca farklı dizinin elde edilmesi mümkün olmuştur. Bugün yüksek miktarda sekans verisinin erişilebilir olması sınırsız markör geliştirme kapasitesini kazandırmıştır. Sekanslama yöntemlerinin gelişmesi ile biyoinformatik analizlerde önem kazanmıştır.



Şekil 2. 2. Sekanslama maliyetinin, teknolojiye gelişmelerle zaman içerisinde değişimi Ulusal İnsan Genom Araştırma Entitüsü (Wetterstrand, 2020).

2001' den 2007'ye kadar olan veriler kimyasallar ve kapiler tabanlı araçların kullanıldığı Sanger sekanslama yani 1. Nesil sekanslama platformlarından elde edilen verilerdir. 2008'den başlayıp günümüze kadar olan veriler ise 'ikinci nesil' (veya 'yeni nesil') sekanslama platformlarını kullanarak DNA dizisi oluşturma maliyetlerini temsil eder.

2.3. Markör Geliştirme

DNA polimorfizminin tespiti ve kullanımı için moleküler markörlerin geliştirilmesi ve kullanılması, moleküler genetik alanındaki en önemli gelişmelerden biridir. Çeşitli tipte moleküler markörlerin varlığı ve bunların ilkeleri, metodolojileri ve uygulamalarındaki farklılıklar, bu tür yöntemlerden bir veya daha fazlasını seçerken dikkatli bir değerlendirme gerektirir. Genetik çeşitliliğin değerlendirilmesi, bağlantı haritasının oluşturulması, genetik saflık testi, kantitatif karakter lokuslarını (QTL) haritalama ve markör destekli seçim (MAS) gibi çeşitli genetik ve ıslah uygulamalarında çalışmalarının etkinliğini artırmaya yönelik, ilgili genlere bağlı moleküler markörler geliştirilebilmesi son derece önemlidir.

Son otuz yılda, Restriksiyon Parça Uzunluk Polimorfizmi (RFLP), Rastgele Amplifiye Edilmiş Polimorfik DNA (RAPD), Basit Dizi Tekrarı (SSR), Çeşitlilik Dizileri Teknolojisi (DArT), Çoğaltılmış Parça Uzunluğu Polimorfizmi (AFLP), gibi bir dizi DNA markör sistemi moleküler bitki ıslahı için markör geliştirmede kullanılmıştır. Bu yöntemler etkilidir, ancak yoğun emek gerektirir ve zaman alıcıdır.

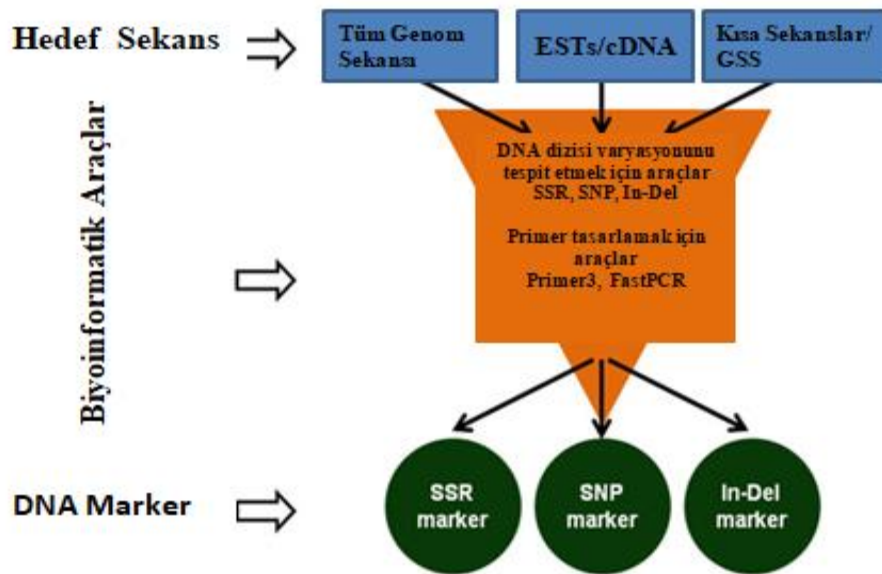
Yeni nesil sekanslama teknolojilerinin ortaya çıkışı ile pirinç, sorgum, mısır, kavak, üzüm, papaya, patates, medicago, hint fasulyesi ve soya fasulyesi gibi önemli ürün türlerinin bazı tam genomlarının sekanslanması ile sonuçlanmıştır. Buna ek olarak, birçok transkripsiyon projesi, neredeyse ekonomik açıdan önemli tüm türler için ifade edilen dizi etiketleri (ESTs) veya cDNA dizi koleksiyonlarının artmasını sağlamıştır. Bu diziler esas olarak, Ulusal Biyoteknoloji Bilgi Merkezi (NCBI) 'daki GenBank, Avrupa Moleküler Biyoloji Laboratuvarı (EMBL) nükleotit dizi veritabanı ve Japonya DNA Veritabanı Bankası (DDBJ) gibi genel veritabanlarında depolanmaktadır. Farklı ürün türlerinin sekans verilerinin fazla miktarda depolanması nedeniyle, birçok ürüne özgü veritabanları kurulmuştur. NCBI veritabanı, tüm nükleer genomun yanı sıra bitkilerinin organel genomlarına, nükleotit dizilerine, genom araştırma sekansları (GSS) ve EST dizilerinden

oluşmaktadır. DNA sekanslarının varlığı, hedef sekansın tanımlanması için önemli bir kaynak oluşturur.

2.3.1. Moleküler Markör Gelişimi İçin Biyoinformatik

Biyoinformatik, in silico analiz yoluyla işe yarar sekans verilerini elde etmeye yardımcı olur, böylece moleküler araştırmacılar, özellikle hızlı ve ucuz bir şekilde moleküler markörlerin geliştirilmesinde ve ürün geliştirmek için bu verileri kullanabilir.

Biyoinformatik araçlar, Web tabanlı araçlar, verihatları, Perl komut dosyaları, Java komut dosyaları, vb. şeklindedir. DNA dizisi polimorfizmleri bu biyoinformatik araçlar tarafından tanımlandıktan sonra, bu tür sekans varyasyonları GeneTool, Primer3, FastPCR, gibi çeşitli primer dizayn yazılımları aracılığıyla uygun primerler dizayn edilerek, PCR tabanlı markörlere dönüştürülebilmektedir Şekil 2.3'de İn Silico yaklaşımlarıyla DNA markörlerinin geliştirilmesi genel bir şema ile gösterilmiştir. Genom araştıma sekansları (GSS), ifade edilen dizi etiketleri (ESTs), tam uzunlukta cDNA'lar, tüm nükleer ve organel genom dizileri, genom ek açıklamalarından kodlama sekansları, Basit Dizi Tekrarları (SSR), İnsersiyon-Delesyonlar (InDel) ve Tek Nükleotid Polimorfizmleri (SNP'ler), DNA dizisi polimorfizmlerinin belirlenmesi için hedef olarak hizmet vermektedir (Rajendrakumar., 2015).



Şekil 2. 3. İn Silico yaklaşımlarla DNA markörlerinin geliştirilmesi (Rajendrakumar, 2015).

Hedef Sekans: Genom araştırma sekansları (GSS) ve ifade edilen sekans etiketleri (ESTs), tam uzunlukta cDNA'lar ve herhangi bir PCR amplicon sekansları gibi tüm nükleer /organel genom sekansları veya kısa nükleotit dizileri olabilir.

Biyoinformatik Araçlar: DNA markörü geliřimi için yararlı biyoinformatik araçlar, SSR, SNP, İnDel, vb. gibi DNA dizisi varyasyonlarının tespiti için gerekli araçları ve GeneTool, Primer3, FastPCR GMATA gibi primerlerin dizaynı için gerekli araçları içerir. DNA varyasyonunu hedefleyen primerlerin dizaynı, PCR tabanlı markörlere dönüřtürülmesi için gereklidir.

2.3.2. SSR Markörü Geliřtirme

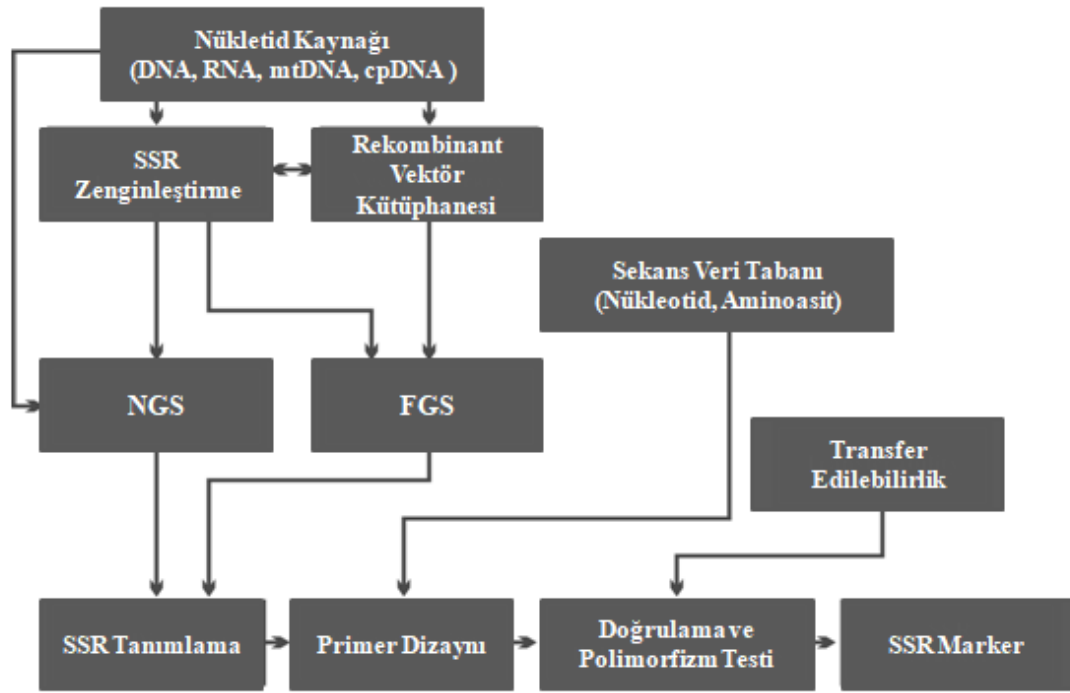
Basit dizi tekrarları hiperpolimorfik yapısı nedeniyle, PCR tabanlı moleküler markörlerin geliřimi için uygun hedefler olarak kabul edilirler. SSR markörleri, yüksek polimorfizm oranına sahip olması lokus spesifik ve kodominant doğası, genomda bol bulunması ve tekrarlanabilirlik düzeyleri nedeniyle moleküler ıřlahçıların tercihi haline gelmiřtir. SSR'lar, kodlama bölgesinde bulunan ve transkripsiyonlu RNA dizilerinden tanımlanan kaynaklarına, yani genomik SSR'lar (g-SSR) ve ifade edilen dizi etiketleri SSR'lara (EST-SSR) göre çeřitlendirilir.

Yeni nesil sekanslama teknolojilerinden önce, SSR'ların geliřim sürecini, DNA parçalanması için restriksiyon enzimleri kullanarak rekombinant DNA molekülleri oluřtururken, hedeflenen SSR motifleri için genomik kütüphaneler oluřturma zorunluluęu nedeniyle yoğun emek isteyen, ekonomik açıdan maliyetli ve zaman alıcı süreçlerdi. Güçlü, hızlı, uygun maliyetli ve güvenilir bir teknoloji olan yüksek verimli Yeni Nesil Sekanslama teknolojileri, çok miktarda sekans verisi oluřturarak moleküler markörlerin keřfi ve geliřtirilmesi alanını deęiřirmiřtir.

SSR markörlerinin geliřtirilmesi ařaęıdaki temel ařamalara ayrılabilir

- SSR markörlerin meydana geldięi nükleotid dizileri hakkında ön bilgi elde edilmesi;
- SSR markörünün çevreleyen bölgeleri tamamlayıcı oligonükleotidlerin (veya primerlerin) tasarımı;
- PCR ve reaksiyon ürününün elektroforezi ile primerlerin doęrulanması;
- Bireyler arasında polimorfizmlerin tespiti (Vieira ve ark, 2016).

řekil 2.4' de SSR markörlerinin geliřtirilmesini gösteren bir řema verilmiřtir.



Şekil 2. 4. SSR markörlerinin geliştirilmesini gösteren bir şema (Vieira ve ark., 2016).

Oluşum ve geliştirme kaynağına göre SSR markörleri:

- a. Genomik veya nükleer SSR (gSSR'ler)- nükleer genomdan izole edilen SSR (genomik DNA kütüphanesi olan veya olmayan bir organizmanın genomik DNA'sı).
- b. EST veya genik SSR (EST-SSR'ler)- Kamuya açık veritabanlarında depolanan EST dizilerinden yararlanılarak geliştirilen SSR'lar.
- c. Organel SSR-kloroplast SSR'lar (cpSSR'lar) ve mitokondriyal SSR'lar (mtSSR'lar) - bir organizmanın kloroplastından veya mitokondriyal genomundan geliştirilen SSR'lar (Taheri ve ark., 2018).

SSR tanımlanması için MISA (MicroSatellite) ve SSR Locator gibi araçlar geliştirilmiştir ancak, bu araçlar büyük genomları verimli bir şekilde işleyememektedir ve üretilen istatistikler yeterince güvenli değildir. Son zamanlarda, yeni bir yazılım olan genom genelinde mikrosatellit arama aracı (GMATA), büyük genomlarda SSR analizi ve primer tasarımı için yeni stratejiler uygular, bu da GMATA'nın mevcut araçlardan daha hızlı hesaplama yapmasını ve daha doğru sonuçlar vermesini sağlar. Ayrıca, bu yazılım yüksek kaliteli istatistiksel grafikler verir. Özellikle, GMATA bir genom tarayıcısında diğer genom özellikleri ile birlikte SSR lokus ve SSR markör bilgileri görüntülemeyi sağlayan sonuçlar üreten ilk araçtır (X. Wang ve Wang, 2016).

Son yıllarda yapılan bazı çalışmalarda Transkripsiyon faktörü (TF) genlerine dayanan kodominant SSR markörlerinin, QTL haritalaması ve markör destekli seleksiyon için yararlı araçlar olarak hizmet edebileceği ve monokot, dikot türlerinin ıslahında özel bir öneme sahip olacağı düşünülmüştür. Bitkilerin stres tepkileri dahil olmak üzere farklı hücrel süreçleri kontrol eden ana düzenleyiciler olarak işlev gören TF'lerdeki fonksiyonel alanların rolü hakkında önemli miktarda kanıt mevcuttur. Dizileme teknolojilerindeki gelişmeler artık, TF gen tabanlı moleküler markörler gibi aday gen tabanlı genomik kaynakları geliştirmek için transkriptom dizilerini kolayca ve maliyet etkin bir şekilde oluşturmamız için yeterli bilgi sağlıyor.

Sharma ve ark. (2020) yaptıkları çalışmada buğdayda 370 cDNA dizisinden toplam 846 SSR tanımlamış ve daha sonra buğdayda yüksek polimorfizm ve çapraz türler arası aktarılabirlik gösteren toplam 35 yeni bZIP türevi SSR tanımlamışlardır, tanımlanan SSR'lar 92 Hint buğday çeşidinin ve ilgili türlerin çeşitlilik ve genetik yapı analizi için kullanılmıştır.

Zhang ve ark. (2015) yaptığı çalışmada, *M. truncatula* genomunda 75 bZIP transkripsiyon faktörü kodlayan gen belirlenmiştir. Bu bZIP protein dizilerinin filogenetik analizi, bunların 12 gruba sınıflandırılmasına izin vermiştir. Ek olarak, 53 MtbZIP genine özgü 104 SSR tespit edilmiş ve 97 primer çifti geliştirilmiştir. 20 yonca çeşidinde rastgele seçilmiş 30 primer çiftinin doğrulanması, bunların 17'sinin SSR lokusu için doğru olduğunu ve yonca çeşitleri arasında oldukça aktarılabir olduğunu göstermiştir.

Biswas ve ark. (2018) yaptığı çalışmada, Zambak (*Lilium sp.*), genomunda transkripsiyon faktörü (TF) ile ilişkili SSR markörleri (TFSSR'ler) geliştirmek için *Lilium sp.* transkriptomunda büyük ölçekli veri analizi yapılmış ve sekans verilerinden çıkarılan 216.768 unigen arasında 6966 unigende SSR bulundurduğu tespit edilmiştir. 71 SSR, TF genleriyle ilişkilendirilmiş ve bunlar, primerleri tasarlamak ve markör olarak potansiyellerini doğrulamak için kullanılmıştır. Bu 71 SSR, bHLH, MYB, C2H2, ERF, C3H, NAC, bZIP ve benzeri dahil 31 transkripsiyon faktör ailesi ile analiz edilmiştir.

Parmar ve ark. (2020) yaptığı çalışmada TF'lerini kodladığı varsayılan 2776 transkriptde 3687 SSR lokusu tanımlanmıştır. ORF'de (% 44.37), UTR'lerde (5'UTR:% 32.57; 3'UTR:% 23.0) yüksek tekrar bolluğu, yeni SSR lokuslarının anlamlı fonksiyonel ve düzenleyici etkisini göstermiştir. 589 polimorfik yeni deneysel olarak doğrulanmış SSR markörleri (ORF: 297; 5'UTR: 211; 3'UTR: 81), lokus başına 2 ila 17 alel sayısı ile 2864 alel varyantı tespit edilmiştir.

Kujur ve ark. (2013) yaptığı çalışmada Nohutun 707 TF'sinden 1108 transkripsiyon faktör geninden türetilmiş SSR (TFGMS) ve 161 transkripsiyon faktörü fonksiyonel alanla ilişkili SSR (TFFDMS) markörü geliştirmişler ve markörler tarafından tespit edilen amplifikasyon verimliliği (% 96,5) ve yüksek intra-spesifik polimorfik potansiyel (% 34) olarak hesaplanmıştır. Hem fiziksel hem de fonksiyonel transkript haritalarının oluşturulması ve popülasyon yapısının anlaşılması dahil olmak üzere verimli büyük ölçekli genotipleme uygulamalarındaki önemli faydalarını göstermiştir.

2.4. Ökaryot Organizmalarda Transkripsiyon

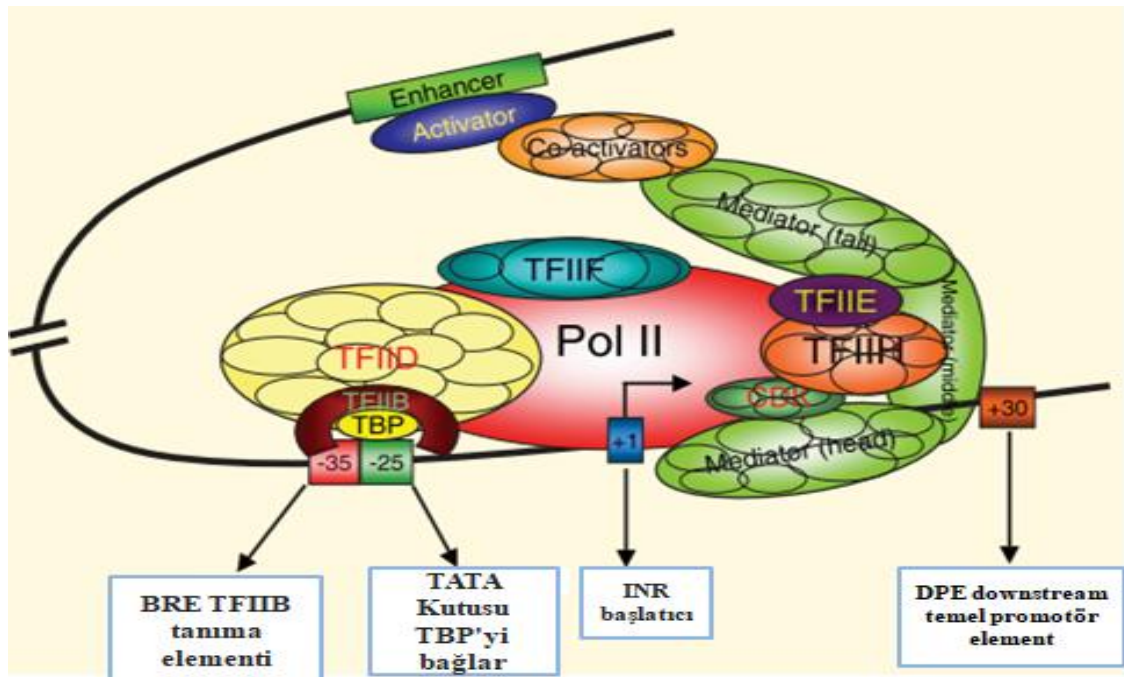
Transkripsiyon, bir DNA şablonundan RNA sentezi olarak tanımlanır. Kalıp DNA sekansına bağlanıp başlama, uzama ve sonlanma aşamalarından oluşur. Bakterilerde, tüm genlerin transkripsiyonu tek bir RNA polimeraz tarafından katalize edilirken, ökaryotik hücre çekirdeğinde, üç farklı RNA polimeraz Pol I, Pol II ve Pol III farklı RNA sınıflarının sentezini gerçekleştirir. Pol I ve Pol III, ribozomal RNA'ları, tRNA'larını ve küçük nükleer RNA'ları kodlayan sınırlı sayıda geni kopyalar. Buna karşılık Pol II, protein kodlayan binlerce genin transkripsiyonundan sorumludur ve bunu, 3 milyon daltondan (MDa) fazla birleşik moleküler kütleyle sahip yaklaşık 60 polipeptit içeren protein makinesi kullanarak yapar (Krishnamurthy ve Hampsey, 2009).

Nükleer olarak kodlanmış genler, core promotor elemanlarında (CPE'ler) transkripsiyon başlangıç kompleksinin başarılı oluşumunda rol oynayan birçok protein ile karmaşık etkileşimler gerektiren RNA polimeraz II (RNAPII) tarafından kopyalanır. Hedef genlerin promoter bölgelerindeki cis-acting elementlere spesifik olarak bağlanarak gen ekspresyonunu düzenleyen proteinlere, transkripsiyon faktörleri (TF) adı verilir. CPE'lere bağlanan ve bazal transkripsiyonda yer alan TF'ler genel transkripsiyon faktörleri (GTF) olarak adlandırılır. GTF'ler, kromatografik elüsyon profilleri ve keşif sırasına göre TflIA, TflIB, TflID, TflIE, TflIf ve TflIH olarak adlandırılmıştır (Şekil 2.5.) (Hong., 2016).

Ökaryotik transkripsiyon, bir aktivatörün enhancer DNA elemanlarına sekansa özgü bağlanması ile başlar, daha sonra genel transkripsiyon faktörlerinin (GTF'ler) ve RNA polimeraz II'nin (RNAPII) hedef gen promoterlerine sıralı olarak görevlendirilmesi ile devam eder (Shandilya ve Roberts, 2012).

Transkripsiyon süreci dinamik bir kromatin yapısı gerektirdiğinden, histon değiştirici enzimler, adenozin trifosfat (ATP) bağımlı kromatin yeniden modelleyicileri

ve histon şaperonları gibi çeşitli hücrel faktörler yer alır. Ek olarak, aktive edici veya inhibe edici olarak hareket eden DNA'ya bağlı TF'ler, hedef aktive edici medyatör (Med) kompleksi ve promotör inhibe edici gibi kofaktörlerle etkileşime girer. Transkripsiyon aktivatörlerinin down streamden uzak düzenleyici elemanlara sekansa özgü bağlanmasıyla başlar daha sonra GTF'lerin ve RNAPII'nin hedef gen promotörüne sıralı olarak görevlendirilmesini teşvik eder. Transkripsiyon başlangıç bölgesine (Ts) en yakın bölgelerde bulunan ve transkripsiyonun doğru şekilde başlatılmasını yönlendiren spesifik DNA elemanlarına CPE'ler denir. CPE'ler, hedef gen başlangıç bölgesinin up veya down stream bölgesine yerleştirilir, bazı durumlarda CPE'ler kodlama bölgesinde de bulunur. CPE'ler TATA içeren ve TATA'sız promotörler olarak sınıflandırılır (Krishnamurthy ve Hampsey, 2009).



Şekil 2. 5. Pol II transkripsiyon ön başlatma kompleksinin şematik özet

(Krishnamurthy ve Hampsey, 2009)

Çekirdek promotör elemanları, -35, -25, +1 (transkripsiyonun başlangıç bölgesi) ve +30 pozisyonlarında gösterilmektedir. Ön başlatma kompleksi, TFIID'nin TATA bağlayıcı protein (TBP) alt biriminin TATA kutusuna bağlanmasıyla başlar, ardından genel transkripsiyon faktörleri, TFIIB, Pol II / TFIIF, TFIIE ve TFIIH'nin aşamalı bağlanması ile gerçekleştirilir.

Transkripsiyon başlangıç bölgesinden 25 baz çifti ortalanmış olan TATA kutusu, tanımlanmış ilk RNA Pol II promoter elementiydi. Daha sonra, +1 de transkripsiyonun başlangıç bölgesini kapsayan başlatıcı (Inr) ve +30'da ortalanmış down stream destekleyici eleman (DPE) dahil olmak üzere diğer yapısal ve düzenleyici olarak farklı promoter elementleri tanımlandı (Krishnamurthy ve Hampsey, 2009).

TATA dizisi, birkaç TBP ile ilişkili faktör (TAF) ile birlikte TFIID kompleksini oluşturan TATA bağlayıcı protein (TBP) ile bağlanır. Daha sonra TFIIB, DNA-TBP kompleksi ile birleşerek, TATA'nın hem up stream alanında hem de down stream alanından DNA'yı bağlar. Daha sonra, TFIIF, RNAPII'nin başlama öncesi kompleksinin oluşumunu daha da stabilize eden TFIIE ve TFIIH bağlanmasını izleyen çekirdek promotere RNAPII bağlanmasını kolaylaştırır.

2.4.1. Ökaryotlarda Korunmuş Başlıca TF Aileleri

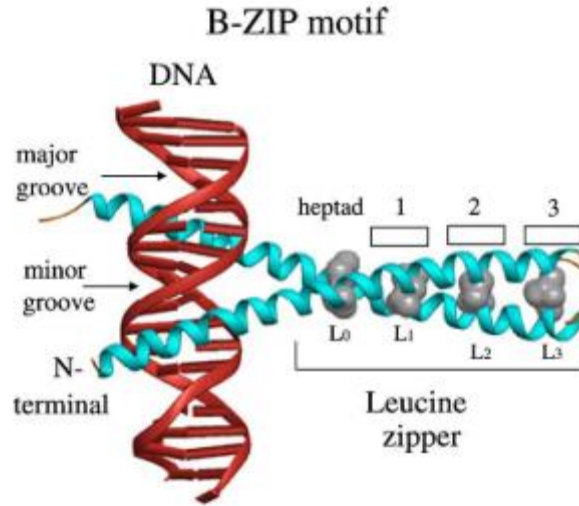
Gen ekspresyonu, hücrelerdeki en karmaşık aktivitelerden biridir çünkü sinyal iletim yollarının bir araya gelmesi, proteinlerin hücreler arasında hareketini, kromozom yapısındaki değişiklikleri, RNA sentezini ve RNA işlemeyi içerir. Bitki büyümesini ve gelişimini moleküler düzeyde anlamak için, transkripsiyon mekanizmaları hakkında ayrıntılı bilgi gereklidir. Bunu başarmak için kapsamlı bir dizi teknik ve yaklaşım bir araya getirilmeli ve tamamlayıcı bir şekilde kullanılmalıdır (Schwechheimer ve ark., 1998).

TF'ler genel olarak hedef genlerinin ekspresyonunu (ifadelenmesini) düzenlemek için sekansa özgü bir şekilde hedef genlerin promoterlerine doğrudan bağlanan proteinler olarak tanımlanır. TF'ler promoterlerde spesifik DNA sekanslarını tanıyan DNA'ya bağlanma domainleri (DBD) içerir. Bununla birlikte, transkripsiyonel kompleksler oluşturmak için DNA bağlayıcı TF'ler ile doğrudan etkileşime giren bir DBD'siz düzenleyici proteinler vardır ve bunlar da aynı zamanda TF'ler olarak kategorize edilir (Hong, 2016).

2.4.1.1. bZIP TF Ailesi

Temel bölge / lösün zipper (bZIP) TF'ler, DNA'yı bağlayan temel bir bölgeye ve bir lösün zipper dimerizasyon motifine sahiptir (Şekil 2.6). bZIP domainlerine sahip proteinler, bu güne kadar analiz edilen tüm ökaryotlarda mevcuttur. Jun / Fos veya CRE

bağlayıcı protein (CREB) gibi bazıları, hayvanlarda kapsamlı bir şekilde çalışılmıştır ve TF-DNA etkileşimlerini, üçlü kompleks oluşumunu ve TF post-translasyonel modifikasyonları anlamak için modeldir (Jakoby ve ark., 2002).



Şekil 2. 6. Çift sarmallı DNA'ya bağlı GCN4 B-ZIP dimerinin X-ışını yapısı.

B-ZIP α -helixleri mavi renktedir ve lösinler gri ile gösterilen "d" pozisyonundadır. Proteinin N-terminali, temel bölge ve lösin zipperi etiketlenmiştir. Lösin zipperinin ilk üçü numaralandırılmıştır. (Vinson, ve ark., 2006).

Arabidopsis'te, proteinleri bZIP imza motifiyle kodlayan 75 varsayılan gen (AtbZIP1 – AtbZIP75) mevcuttur. AtbZIP proteinleri, temel bölgelerinin sekans benzerliklerine, korunmuş motiflere ve boyutlara göre 10 grupta (A – I ve s) kümelenir. Grup-A bZIP proteinleri arasında ABA yanıt elementi (ABRE), bağlanma faktörleri (ABfs) ve ABA'ya duyarlı element bağlama proteinleri (AREB'ler) bulunur (Hong, 2016).

Arabidopsis, maya, solucan ve insandan yaklaşık dört kat fazla bZIP genine sahiptir (Riechmann ve ark., 2000).

Bu Arabidopsis thaliana bZIP (AtbZIP) faktörlerinden birkaçının genetik ve moleküler çalışmaları, bunların patojen savunması, ışık ve stres sinyali, tohum olgunlaşması ve çiçek gelişimi gibi çeşitli biyolojik süreçleri düzenlediğini göstermiştir (Jakoby ve ark., 2002).

2008'de yapılan bir çalışmada, 4 bZIP geninin tüm bitkilerin en son ortak atasının genomu tarafından kodlandığını ortaya çıkarmıştır (Guedes Corrêa ve ark., 2008).

2.4.1.2. bHLH TF Ailesi

1989'da Murre, McCaw ve Baltimore, önemli sayıda aynı amino asidi paylaşan hayvanlardan on DNA bağlayıcı proteinde bulunan bir bölge tanımlamıştır (Murre ve ark., 1989).

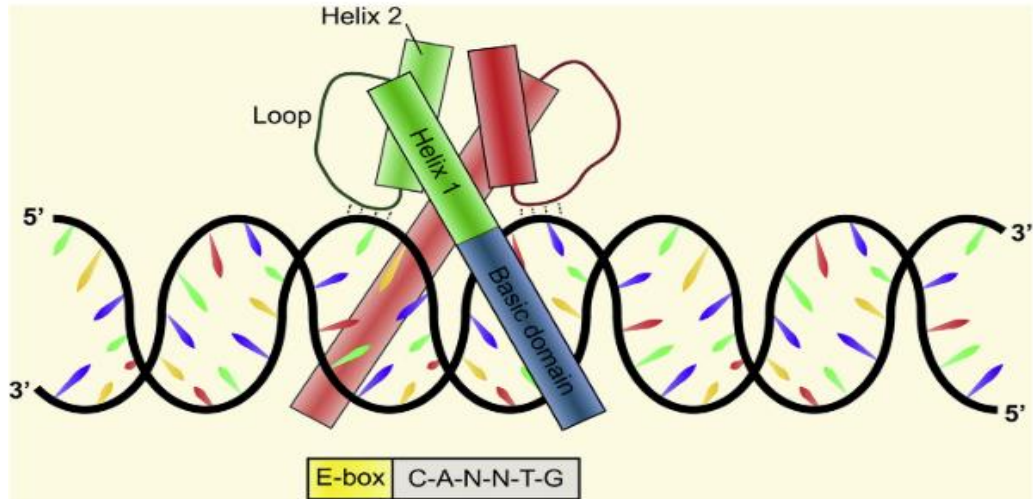
Bu bölge, basic helix-loop-helix (bHLH) alanı olarak bilinir. Bu alanı içeren proteinler, hücre proliferasyonunu ve hücreyel farklılaşma yollarını düzenlemede geniş işlevlere sahiptir ve insan c-Myc, miyoblast belirleme faktörü (MyoD) ve ayrıca nöral gelişimdeki ilk adımları tanımlayan *Drosophila melanogaster*'dan Achaete ve Scute'u içerir (Heim ve ark., 2003).

Basic helix-loop-helix (bHLH) TF ökaryotlarda yaygın olarak bulunur ve en büyük TF ailelerinden birini oluşturur (Riechmann ve ark., 2000).

bHLH transkripsiyon faktörleri, dimerizasyona aracılık eden iki alfa sarmalından ve DNA'daki E-box dizilerini bağlayan temel bir alandan oluşur (Şekil 2.7). bHLH TF'ler, yüksek oranda korunmuş bir bHLH alanının varlığı nedeniyle bu şekilde adlandırılır ve 50-60 amino asitlik bHLH alanı, bir temel bölge ve bir HLH bölgesi içerir.

15 amino asitlik temel bölge, bHLH alanının N-terminalinde bulunurken, 40-50 amino asitlik bHLH bölgesi, bHLH alanının iki alfa heliks ve değişken uzunluklara sahip daha az korunmuş C terminalindedir (Hong, 2016; Y. Wang ve Liu, 2020). Bazı bHLH proteinleri, en yaygın form olan G-box (5' CACGTG-3') ile E-box (5'-CANNTG-3') olarak adlandırılan bir konsensüs cis elemanı içeren dizilere spesifik olarak tanıyıp bağlanabilir (Hong, 2016).

Aynı bHLH TF'den veya farklı bHLH TF'lerden gelen iki sarmal, bir homodimer veya bir heterodimer oluşturmak için etkileşime girebilir ve bu daha sonra, hedef genlerinin ekspresyonunu düzenlemek için promoterdeki farklı bölgeleri bağlar. Diğer bHLH TF dizileri değişkendir ve daha az korunur (Y. Wang ve Liu, 2020).



Şekil 2. 7. bHLH transkripsiyon faktörlerinin yapısı
(Dennis ve ark., 2019).

bHLH transkripsiyon faktörleri, ifade kalıplarına göre sınıf I veya sınıf II olarak geniş bir şekilde sınıflandırılabilir. Sınıf I bHLH proteinleri her yerde ifade edilir ve Tcf2a (E12, E47), Tcf4 ve Tcf12 tarafından kodlanır. Sınıf II bHLH faktörlerinin dokuya özgü ifade profilleri vardır ve miyogenez dahil çeşitli gelişim süreçlerinde yer alırlar (Dennis ve ark., 2019).

2.4.1.3 MYB TF Ailesi

MYB (miyeloblastoz) transkripsiyon faktör ailesi tüm ökaryotlarda mevcuttur. "Onkogen" v-MYB, kuş myeloblastoz virüsünde tanımlanan ilk MYB genidir (Ambawat ve ark., 2013; Cao ve ark., 2020; Li ve ark., 2015). Daha sonra birçok omurgalıda v-MYB ile ilgili üç gen, yani c-MYB, A-MYB ve B-MYB tanımlanmış ve hücre proliferasyonu, farklılaşması ve apoptozunun düzenlenmesinde rol oynamıştır (Ambawat ve ark., 2013; Katiyar ve ark., 2012).

Hayvanlarda, bitkilerde, mantarlarda ve balçık küflerinde birçok MYB geni tanımlanmıştır. Bitkilerdeki MYB transkripsiyon faktörlerinin yapıları ve fonksiyonları, hayvanlar ve mayalardakine kıyasla daha fazla korunduğu tespit edilmiştir (Li ve ark., 2015).

Bitkilerde transkripsiyon faktörünü kodlayan ilk gen, mısır (*Zea mays*) çekirdeklerinin aleuronundaki antosiyaninlerin sentezi için gerekli bir MYB alan proteinini kodlayan COLORED1 (C1) lokusudur (Paz-Ares ve ark., 1987).

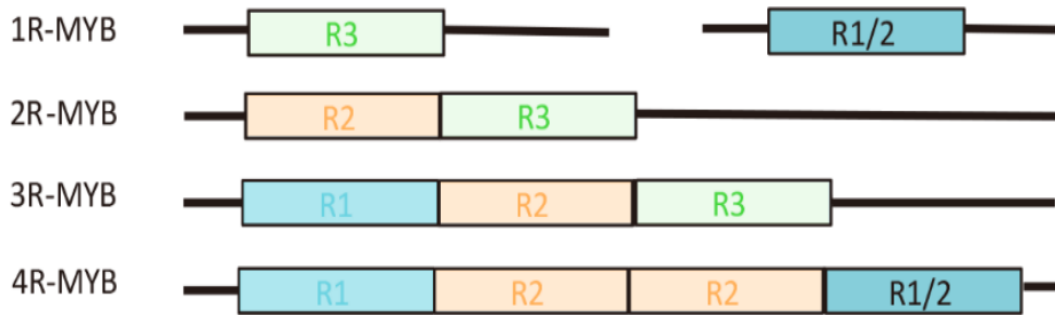
MYB ailesi, Arabidopsis'teki toplam TF'lerin yaklaşık % 9'unu temsil eden bitkilerdeki en büyük TF ailelerinden biridir (Riechmann ve ark., 2000).

Genel olarak, MYB alanı genellikle bitkilerde 1-4 tekrardan oluşur (Cao ve ark., 2020; Hong, 2016; Li ve ark., 2015).

Her bir tekrar, üç alfa helix kodlar ve yaklaşık 50-53 amino asit içerir. Bu üç helix arasında, ikinci ve üçüncü helixler bir helix- turn- helix (HTH) yapı oluşturur (Katiyar ve ark., 2012).

MYB ailesi, tekrarların sayısına ve konumuna göre dört alt aileye ayrılır; 4R-MYB, R1R2R3-MYB, R2R3-MYB ve 1R-MYB (Şekil 2.8), (Cao ve ark., 2020; Li ve ark., 2015).

MYB proteini ilk olarak c-Myb'nin üç tekrarı, R1, R2 ve R3 olarak adlandırılmıştır ve diğer MYB proteinlerinden gelen tekrarlar, c-Myb'nin R1, R2 veya R3'e benzerliklerine göre adlandırılır (Hong, 2016; Katiyar ve ark., 2012).

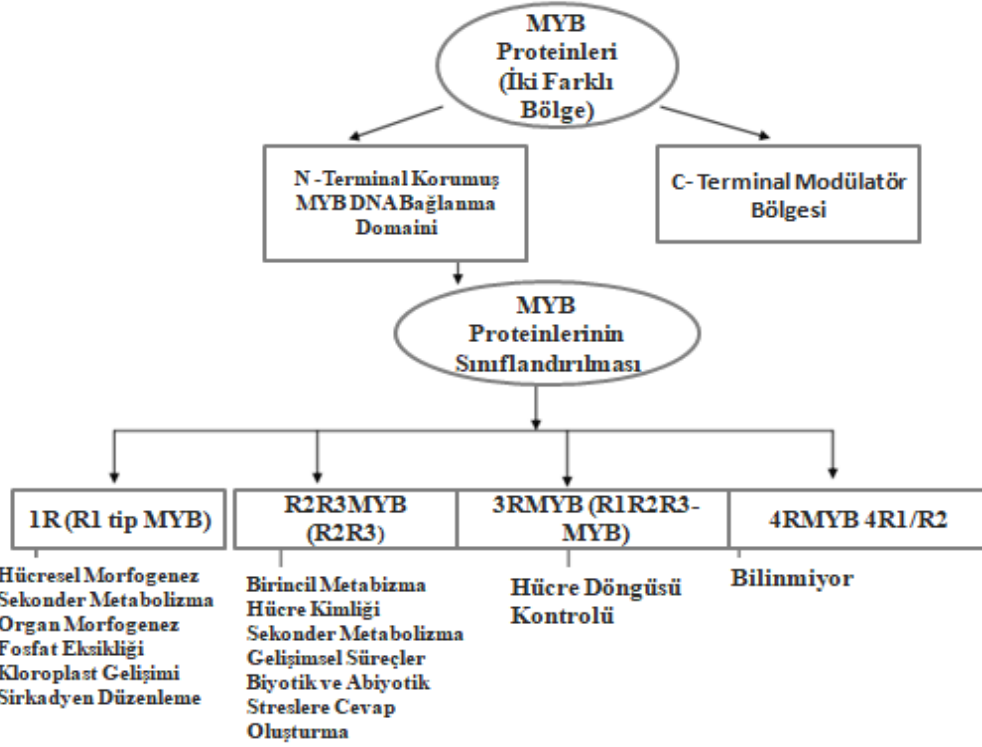


Şekil 2. 8. Bitkilerdeki MYB TF'lerin sınıflandırılması ve yapısının gösterimi. Bitkilerde tanımlanan bir ila dört MYB alanı tekrarına sahip MYB TF'ler (Cao ve ark., 2020).

Bitki MYB proteinleri üç ana grup olarak sınıflandırılmıştır, iki bitişik tekrar ile R2R3-MYB; Üç bitişik tekrar ile R1R2R3-MYB; ve toplu olarak MYB ile ilişkili proteinler olarak adlandırılan ve genellikle tek bir MYB tekrarı içeren heterojen bir gruptur. Bitkilerde MYB ailesi, özellikle geniş R2R3-MYB ailesi aracılığıyla seçici olarak genişlemiştir. Arabidopsis genomunda 138'i R2R3-MYB, 5'i R1R2R3-MYB, 52'si MYB ile ilgili ve 2'si atipik MYB genleri olarak sınıflandırılmıştır. R2R3-MYB proteinleri bitkilere özgüdür ve ayrıca dikotların ve monokotların genomlarında 100'den fazla R2R3-MYB üyesiyle bitkilerde en bol bulunan türdür (Hong, 2016).

Dört sınıfın tümü bitkilerde bulunur ve MYB proteinlerinin en yüksek çeşitliliğine sahip taksonu temsil eder. En küçük sınıf, üyeleri dört R1 / R2 benzeri tekrar içeren 4R-

MYB grubudur. İkinci sınıf, R1R2 benzeri tekrarlar içerir. Tek bir 4R-MYB proteini, birkaç bitki genomunda kodlanmıştır. İkinci sınıf, tipik olarak daha yüksek bitki genomlarında beş gen tarafından kodlanan R1R2R3 tipi MYB (3R-MYB) proteinlerini içerir. Yüksek oranda korunmuş MYB alanının aksine, R2R3-MYB proteinlerinin diğer bölgeleri oldukça değişkendir. Geçtiğimiz on yılda, R2R3-MYB genleri kapsamlı bir şekilde incelenmiş ve MYB ailesinin üyelerinin fenilpropanoid metabolizma, biyotik ve abiyotik stres, hücre şekli farklılaşması, bitki savunması sırasında hormon tepkileri gibi çeşitli biyolojik işlevlerde rol oynadıkları bulunmuştur Şekil 2.9.'da MYB proteinlerinin işlevlerini gösteren bir şema verilmiştir (Ambawat ve ark., 2013).



Şekil 2. 9. MYB protein sınıflarının şematik gösterimi ve bitişik MYB tekrarlarının (R) sayısına bağlı olarak işlevleri (Ambawat ve ark., 2013).

2.4.1.4. HSF TF Ailesi

Isı stresi yanıtının (HSR) temel düzenleyicileri, ısı şoku transkripsiyon faktörleri (HSF'ler) olarak adlandırılan transkripsiyon faktörleridir. HSF'ler, ökaryotlar arasında evrimsel olarak korunur ve HSR'nin ana düzenleyicileridir. Hücreler ısı stresi tespit

ettiğinde, HSF'ler hızla aktive olur ve ısı şoku proteinlerini (HSP'ler) kodlayan birçok genin ifadesini artırır (Guo ve ark., 2016; Ohama ve ark., 2016; Ohama ve ark., 2017).

Isı stresi, tüm organizmaların karşılaştığı evrensel streslerden biridir. Isı stresi, protein denatürasyonuna, reaktif oksijen türlerinin oluşumuna ve membran destabilizasyonuna neden olarak hücrel bileşenlere zarar verir. Bu nedenle, organizmaların, ısı stresine maruz kaldıktan sonra hücrel homeostazı sürdürmek için koruma mekanizmalarını hızla aktive etmesi önemlidir (Ohama ve ark., 2016).

Bitki HSF'leri, çeşitli abiyotik streslere yanıt veren genlerin ifadenmesine aracılık eden bir sinyal transdüksiyon zincirinin terminal bileşenleridir. Birçok çalışma, sıcaklık stresi (HS) dahil olmak üzere çeşitli abiyotik streslerde HSF'lerin merkezi rollerini bildirmiştir ancak, stres yanıtlarındaki çoğu HSF analizi, abiyotik stres faktörlerinin bir kombinasyonunu değil, bireysel stresleri incelemektedir. Doğal koşullarda bitkiler rutin olarak kuraklık, ısı ve tuzluluk streslerinin kombinasyonu gibi farklı abiyotik streslerin bir kombinasyonuna maruz kalırlar (Guo ve ark., 2016).

HSF'ler, bir DBD ve bir oligomerizasyon alanı (OD) ile modüler bir yapıya sahiptir. Ek olarak, alt aileye bağlı olarak, bir nükleer lokalizasyon sinyali (NLS), bir nükleer export sinyali (NES) ve bir aktivatör motifi (AHA motifi) içerirler (Şekil 2.10). OD veya HR-A / B bölgesi, değişken uzunluktaki (15-80 amino asit) esnek bir bağlayıcıyla N-terminal DBD'ye bağlanır. HR-A / B bölgesindeki hidrofobik amino asitlerin yedili bir modeli, lösün zipper tipi protein etkileşimlerine izin veren sarmal bir yapının oluşumuna yol açar (Hong, 2016).

DNA'ya bağlanma domaini (DBD), tüm HSF'nin N-terminaline yakın konumdadır. Sıcaklık stres element (HSE)'ler, ökaryotik HS indüklenebilir genlerin TATA kutusunun yukarısındaki palindromik bağlanma motiflerinin (5'-AGAAnnTTCT-3') tekrarlayan modellerinden oluşur. DNA sarmalının zıt bölgelerinde bulunan ana olukta bulunan G ve C nükleotidleri HSE işlevi için gereklidir. Genellikle ikiden fazla HSE motifi gereklidir ve ek olarak, HSE ince yapısının ayrıntıları ve promotör veya kromatin bağlamı, HSF oligomerlerinin verimli bağlanması için çok önemlidir (Guo ve ark., 2016; Scharf ve ark., 2012).

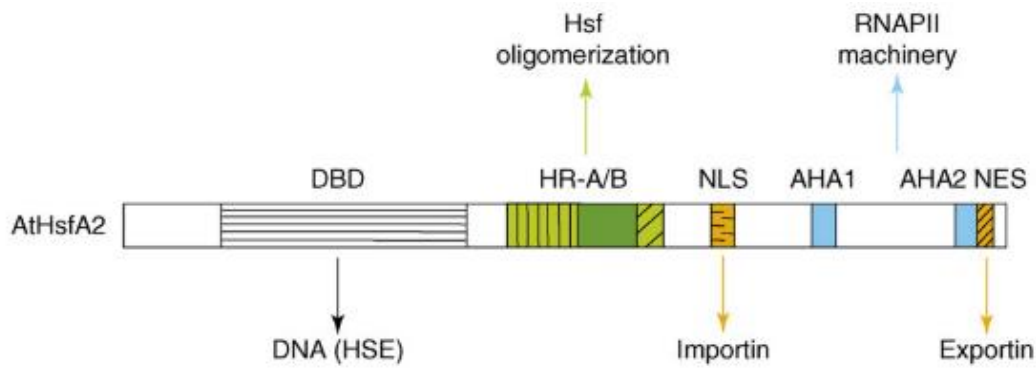
Bitki HSF'lerinin C-terminal aktivasyon bölgeleri, birçok durumda aktivatör işlevi için çok önemli olan kısa peptid motifleri (AHA motifleri) ile karakterize edilir (Guo ve ark., 2016).

Hidrofobik amino asit kalıntılarının (HR-A / B bölgesi) çift taraflı bir heptad modeline sahip oligomerizasyon alanı (OD), DBD'ye esnek bir bağlayıcıyla bağlanır ve,

lösün zipper tipi protein etkileşim alanlarının karakteristiğine sahip kıvrımlı bir alan oluşumuna yol açar (Scharf ve ark., 2012).

DBD ve HR A / B bölgeleri arasındaki esnek bağlayıcı bölgenin uzunluğuna ve HR-A / B bölgelerine eklenen amino asit kalıntılarının sayısına bağlı olarak, bitki HSF'leri HSFA, B ve C olmak üzere üç sınıfa ayrılır. Aromatik, büyük hidrofobik ve asidik amino asit kalıntılarında oluşan AHA motifleri, HSFA'ya özgü motiflerdir, ancak HSFB veya C sınıflarında bulunmazlar (Guo ve ark., 2016).

Bitkiler, diğer ökaryotlardan çok daha fazla HSF kodlayan genlere sahiptir (Arabidopsis'te 21, omurgalılarda 4 ve Drosophila'da 1). Arabidopsis'teki 21 HSF nin 15'i Sınıf A'ya, 5'i Sınıf B'ye ve 1'i Sınıf C'ye aittir (Guo ve ark., 2016; Hong, 2016).



Şekil 2. 10. Arabidopsis HsfA2'nin fonksiyonel modülleri (Döring, 2007).

Diğer birçok gen düzenleyici proteine benzer şekilde, HSF'ler modüler bir yapıya sahiptir N-terminal DNA'ya bağlanma domaini (DBD), bir helix-turn-helix motifi ile karakterize edilir ve HSF'ye yanıt veren genlerin hızlandırıcı bölgesindeki ısı gerilimi elemanlarına (HSE) bağlanır. HSF'nin homo- veya hetero oligomerizasyonu için hidrofobik amino asit kalıntılarının (HR-A / B bölgesi) bir heptad tekrar modeline sahip bitişik alan gereklidir. Nükleer lokalizasyon (NLS; importin ile etkileşim) için bir bazik amino asit kümesi gereklidir ve C-terminalindeki lösün bakımından zengin bir bölge nükleer ihraç sinyali (NES; exportin ile etkileşim) olarak işlev görür. Asidik bir çevreye (AHA motifleri) gömülü aromatik ve büyük hidrofobik amino asit bakımından zenginleştirilmiş kısa peptit motifleri, RNA polimeraz II (RNAPII) mekanizmasının bileşenlerini toplayarak A sınıfı HSF'nin transkripsiyonel aktivitesi için gereklidir.

HSFA2, stresli bitkilerde en yüksek düzeyde indüklenen HSF'dir ve ısı toleransında bir rol oynar ve APX2 (Ascorbate Peroxidase2) gibi genel stresle ilişkili

şaperon kodlayıcı olmayan genlerin ekspresyonunda daha geniş bir rol oynar. B Sınıfı HSF'ler, HS ile indüklenebilir genlerin transkripsiyonunun aktif baskılayıcıları olarak işlev görür. Sınıf B HSF'lerin (IKlfGVWl) baskılayıcı alanında bulunan lfgV-tetrapeptid motifinin, diğer bitki TFs'lerinde baskılayıcı işlevlere sahip olduğu bilinmektedir (Hong, 2016).

2.4.2. Bitki Spesifik TF Aileleri

Bitkiler, çeşitli biyotik ve abiyotik streslere yanıt vermek ve değişen bir ortamda optimum büyümeyi sürdürmek için ayrıntılı mekanizmalar geliştirmişlerdir. Mesajcı RNA (mRNA) transkripsiyonu, hücrel aktivitenin değişen bir ortama göre düzenlenmesini sağlayan önemli kontrol noktası olmuştur.

Abiyotik stres koşulları altında üretilen fitohormon absisik asit (ABA), yaprak stomalarının kapanmasına neden olur ve stresle ilişkili birçok genin aktivasyonunu tetikleyerek, abiyotik stres faktörlerine verilen yanıtlarda anahtar rol oynar. Mikroarray analizi ve büyük ölçekli transkriptom analizi gibi moleküler tekniklerle, bitkilerde çok sayıda abiyotik strese yanıt veren genler tanımlanmıştır. Bu genler, önemli enzimlerin ve fonksiyonel proteinler üretilmesiyle hücrelerin stresten korunmasıyla birlikte aynı zamanda stres yanıtında sinyal iletimini ve gen ifadesini düzenlemede de rol oynar (Guo ve ark, 2016).

TF'lerin farklı ailelere sınıflandırılması genellikle karakteristik DBD'lerine dayanır. Arabidopsis genom dizisinin tamamlanmasından sonra, başlangıçta yaklaşık 30 farklı TF ailesine ait yaklaşık 1500 TF tanımlanmıştır (Riechmann ve ark., 2000).

Diğer bitki türlerinde sınıflandırmanın temeli olarak Arabidopsis TF veri tabanları kullanılmaktadır. Arabidopsis TF aileleri için RARTF, AGRIs, DATF, PlnTFDB olmak üzere dört temsili veri tabanı vardır (Mitsuda ve Takagi, 2009).

Bugüne kadar karakterize edilen bitkiye özgü TF ailesi üyelerinin çoğu, bitkiye özgü organların geliştirilmesi ve uyum sağlamak için yanıt sistemi ile ilgili gen düzenlemelerinde yer almaktadır.

Bitkiler, farklı biyotik ve abiyotik stres türleri ile ve ayrıca yaralanma ile başa çıkmak için birçok farklı strateji geliştirmişlerdir. Bu stratejiler muhtemelen hareketsiz yaşam tarzları nedeniyle geliştirilmiştir. Özellikle kök, sürgün ve yaprakları çiçekler ve meyveleri birbirine bağladığından, besinlerin ve suyun taşınması için kullanılan damar dokusunu tutan gövdede meydana gelirse, yaralar bitkinin hayatta kalmasını tehdit eder.

Ek olarak, yaralar patojenler için olası giriş bölgelerini temsil eder. Hücresel düzeyde bitkiler, savunma sistemlerinin aktivasyonu ile yaralanmalara yanıt verirler. Fizyolojik tepkiler, hücre duvarının onarımı ve güçlendirilmesini ve yara sinyal yollarının aktivasyonunu içerir. Bu hem lokal hem de sistemik savunma ile ilgili proteinlerin indüksiyonu ve etilen ve jasmonik asit (JA) gibi yarayla ilgili hormonların aktivasyonu yoluyla gerçekleşir. Buna paralel olarak, bitki yara iyileşme mekanizmalarını başlatır ve bu da genellikle hücre bölünmesinin yerel olarak yeniden etkinleştirilmesini gerektirir. Bitkiler tarafından yaralanmanın ardından enfeksiyonu önlemek için kullanılan sinyalleme basamakları iyi belgelenmiş olsa da, bu yolların hücre onarım mekanizmalarıyla nasıl iletişim kurduğu açık değildir (Heyman ve ark., 2018; Mizoi ve ark., 2012).

2.4.2.1. AP2/ERF TF Ailesi

AP2 / ERF ailesi, yaklaşık 60-70 amino asitten oluşan ve DNA bağlanmasında rol oynayan AP2 / ERF tipi DBD'ler içeren bitkiye özgü büyük bir TF grubudur (Hong, 2016; Mizoi ve ark., 2012; Yamasaki, ve ark., 2013).

Farklı ERF alt ailesi üyeleri, patojen atakları, abiyotik stres sinyalleri (soğuk gibi), jasmonik asit (JA), reaktif oksijen türleri (ROS), absisik asit (ABA), etilen, salisilik asit (SA) ve doku hasarı gibi kaynaklı çeşitli uyarılara yanıt verir Up streamde bir ROS aktivasyonu, su basması ve kuraklığa direnç, vasküler kalınlaşmanın aktivasyonu, sürgün dallanması, lateral kök primordiası başlangıcı, yaralanmayı takiben doku rejenerasyonu, spontan kallus oluşumu ve kök kök hücre proliferasyonu yer alır (Heyman ve ark., 2018).

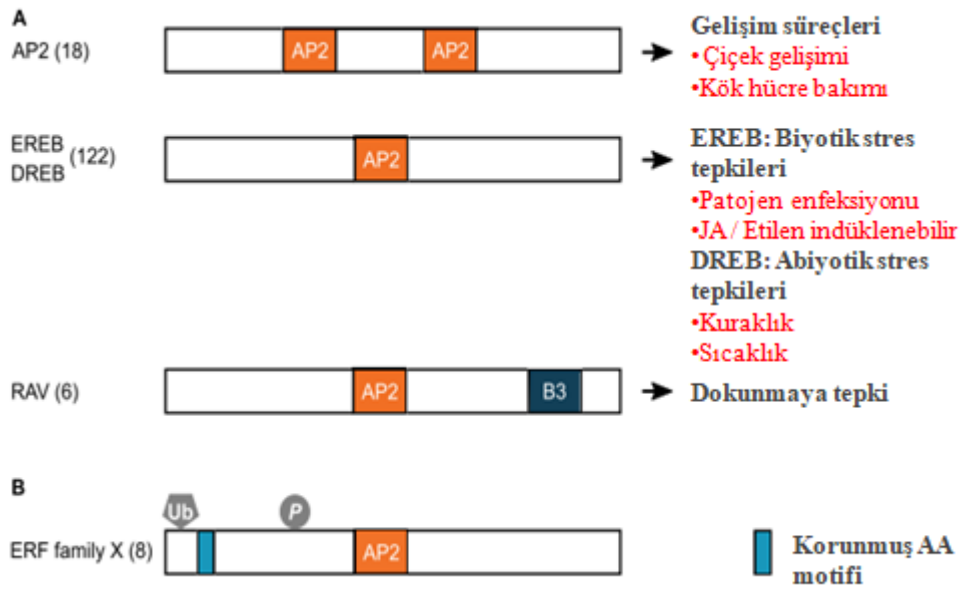
AP2 alanı ilk olarak çiçek gelişiminde rol oynayan Arabidopsis homeotik gen APETALA 2'de bulunmuştur ve benzer bir alan tütünde (*Nicotiana tabacum*) etilene duyarlı element bağlayıcı proteinlerde (EREBP'ler) bulunmuştur. AP2 / ERF ailesi proteinleri, bitkiye özel transkripsiyon faktörleridir ve bulunduğu en düşük bitki, yeşil alg *Chlamydomonas reinhardtii*'dir (Mizoi ve ark., 2012).

AP2 alanı olarak adlandırılan en az bir DNA bağlanma alanı içeren AP2 / ERF proteinleri, ERF, AP2 ve RAV ve dehidrasyona duyarlı element bağlayıcı protein (DREB) alt aileleri aileleri olmak üzere 4 ayrı aileye bölünmüştür (Şekil 2.11) (Licausi ve ark., 2013; Mizoi ve ark., 2012).

Tek bir AP2 alanına sahip ve genomik sekansı az miktarda intron içeren çoğu protein, ERF ailesine atanır. AP2 ailesi, iki AP2 alanının art arda tekrarlanması ve ERF

proteinlerinin AP2 alanına göre daha yüksek benzerlik gösteren tek bir AP2 alanına sahip az sayıda proteinin art arda tekrarlanmasıyla karakterize edilen üyelerden oluşur. Üçüncü bir protein sınıfı, bir B3 DNA'ya bağlanma domaini ile ilişkili bir ERF alanına sahiptir ve RAV ailesini oluştururlar(Licausi ve ark., 2013).

ERF ailesi alt üyeleri, etilene yanıt veren gen transkripsiyonunda yer alan bir DNA dizisi olan GCC kutusunun (5'AGCCGCC3') tanınmasıyla biyotik streslere verilen yanıtlarda yer alır. Cbf / DREB alt ailesi üyeleri, bir çekirdek motif 5'A / GCCGAC3' ile dehidrasyona duyarlı bir öğeyi tanıyarak abiyotik streslerde önemli roller oynarlar. DREB1 / Cbf alt grubunun üyeleri (DREB1A / Cbf3, DREB1B / Cbf1 ve DREB1C / Cbf2) soğukla indüklenebilir ve soğuk stres tepkilerinin başlıca düzenleyicileridir, DREB2 alt grubu (DREB2A ve DREB2B) ise dehidrasyon ve ısı stresi yanıtlarında önemli roller oynamaktadır (Hong, 2016).



Şekil 2. 11. Farklı ERF alt ailelerinin şemantik gösterimi (Heyman ve ark., 2018).

2.4.2.2. WRKY TF Ailesi

WRKY TF ailesi, üzerinde en iyi çalışılmış bitkiye özgü TF ailelerinden biridir ve Arabidopsis'te 74 üyeden oluşur. WRKY protein ailesi adını, N terminalinde korunmuş bir amino asit sekans motifi WRKYGQK ve C terminalinde yeni bir zing- finger benzeri motif içeren yüksek oranda korunmuş 60 amino asit uzunluğundaki WRKY alanından

almıştır. Bu iki motif, W-kutusu (TTGACT / C) olarak adlandırılan konsensüs cis-acting ögesine bağlanmak için hayati öneme sahiptir (Hong, 2016; Yamasaki ve ark., 2013).

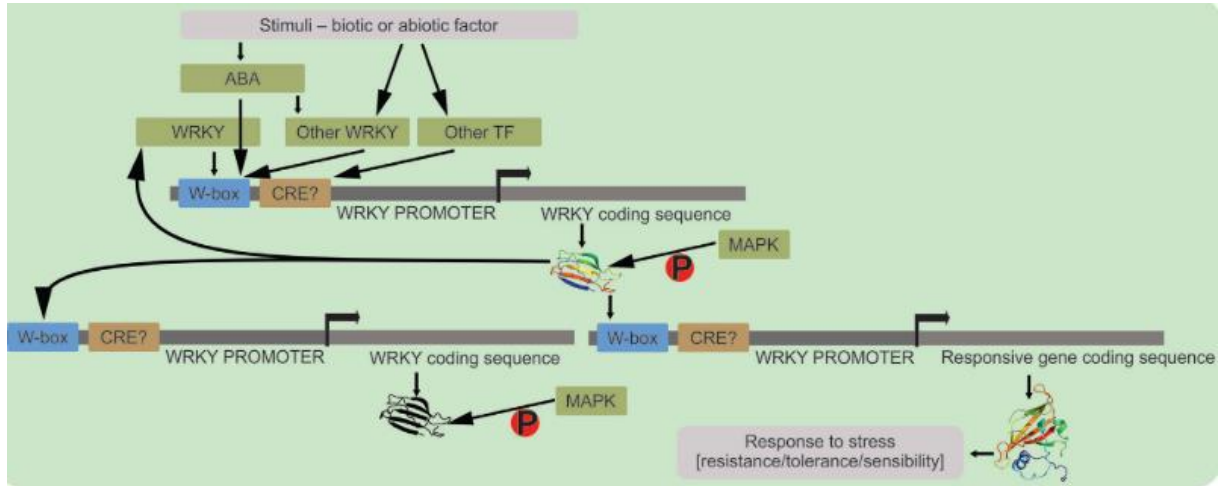
Hem WRKY alanlarının sayısına hem de zing finger motifinin özelliklerine bağlı olarak, WRKY proteinleri üç alt ailede kategorize edilmiştir: birinci Grup I iki WRKY alanına sahiptir, grup II aynı Cys2 – His2 zing finger motifine sahip bir WRKY alanına sahiptir ve Grup III farklı bir Cys2 – His2 zing finger motifi içeren bir WRKY alanına sahiptir. Grup II WRKY'ler ayrıca WRKY alanı dışındaki ek korunmuş motiflere dayalı olarak a – e'ye ayrılır (Chen ve ark., 2017; Hong, 2016).

Bitki WRKY alt grupları, DNA tanıma sekanslarında farklılık gösterir ve 5'-TTGAC (C / T) -3' un up stream G bazı, sırasıyla Iib ve IId gruplarına ait olan AtWRKY6 ve AtWRKY11 için en çok tercih edilir, ancak sırasıyla grup I, III ve IIc'ye ait olan AtWRKY26, AtWRKY38 ve AtWRKY43 için tercih edilmez. Protistler, balçık küfleri ve yeşil alglerde tanımlanan WRKY proteinleri grup I'e aittir ve iki WRKY alanına sahiptir. İki WRKY alanının, eski bir WRKY proteinindeki tek bir WRKY alanından çoğaltma yoluyla oluşturulduğunu düşünülmektedir. Spesifik DNA bağlanmasından sadece C-terminal WRKY alanının sorumlu olduğu düşünülmesine rağmen, DNA'ya doğrudan bağlanan proteinlerin N-terminal WRKY alanında da korunduğu bulunmuştur (Yamasaki ve ark., 2013).

WRKY proteinlerinin boyut olarak büyük ölçüde değişebilmesine rağmen, bunların tümü, iki kısımdan, DNA bağlayıcı heptapeptit WRKYGQK ve zing finger bağlama motifinden oluşan korunmuş bir WRKY alanını barındırır. İki motif birlikte yaklaşık 60 amino asit uzunluğunda N-terminalini ve WRKY alanlarındaki C-terminalini kapsar. Bazı WRKY genleri, bazılarının ZF-SBP, CBS, kinaz, PAH, ULP_PROTEASE, TIR, NAC, LRR, ATP_GRASP, B3 gibi yeni alanları kodlayan birleşik genler olduğu üçlü veya dördü WRKY alanlarını kodlar (Şekil 2.12.) (Chen ve ark., 2017; Mohanta, Park ve ark., 2016).

WRKY TF'lerin, tohum çimlenmesi, kök gelişimi, bitki büyümesi, tohum gelişimi ve yaşlanma, dahil olmak üzere hücrel ve fizyolojik süreçlerde önemli roller oynadığı bildirilmiştir. Ayrıca, böcek otçullar, bakteriyel patojenler, mantarlar ve virüslerin neden olduğu biyotik strese çeşitli tepkiler verirler. İndol-3-asetik asit, jasmonik asit, salisilik asit, absisik asit ve gibberellik asit gibi farklı sinyal moleküllerine de yanıt verirler. Bu nedenle, bitkilerdeki WRKY TF'lerin temel biyolojisi ve genomiklerini anlamak çok önemlidir (Mohanta ve ark., 2016).

yanıtını modüle ederek, uyarıcı algısı ile moleküler hücre yanıtları arasında bir bağlantı kurar. Aktive edildiklerinde, WRKY TF'ler, W-box dizisini bağlayarak farklı bir WRKY'yi düzenleyebilir veya çevresel uyaranlara karşı tolerans, direnç veya duyarlılık kazandıran başka bir yanıt veren geni düzenleyebilir (Şekil 2.13) (Finatto ve ark., 2018).



Şekil 2. 13. WRKY transkripsiyon faktörlerinin düzenleyici mekanizması (Finatto ve ark., 2018).

2.4.2.3. NAC TF Ailesi

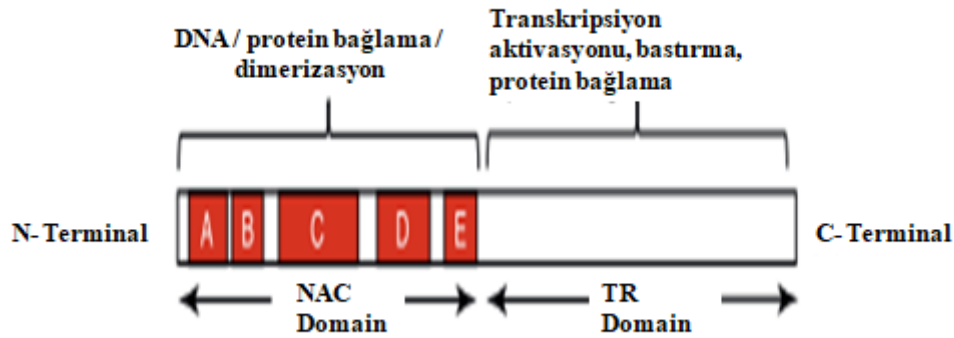
Bir NAC proteinini kodlayan cDNA, ilk olarak Arabidopsis'te dehidrasyona duyarlı (RD26) geni olarak tanımlanmıştır (Nakashima ve ark., 2012).

NAC genlerinin bitki alemindeki dikkate değer çeşitliliği ve amplifikasyonu, bunların sayısız işlevini yansıtır. Bunlar, bitki gelişimi, yaşlanma, besin dağılımı, hücre duvarı biyosentezi ve abiyotik ve biyotik stres tepkilerini içerir (Nakashima ve ark., 2012; Puranik ve ark., 2012; Welner ve ark., 2016). Özellikle stres tepkileri ve duvar biyosentezindeki merkezi roller, gıda, yem ve biyoyakıt üretimi için mahsul optimizasyonu içerisindeki uygulamalar için NAC proteini için büyük bir potansiyel göstermektedir (Welner ve ark., 2016).

NAM / ATAF1 / CUC2 (NAC) proteinleri, önemli model organizmalar Arabidopsis thaliana da dahil olmak üzere birçok bitkide 100'den fazla temsilci ile geniş ve yaygın bir transkripsiyon faktörü (TF) ailesi oluşturur (Nakashima ve ark., 2012; Welner ve ark., 2016).

NAC ailesi proteinleri, N-terminal bölgesinde bulunan ve beş alt alana (A'dan E'ye) bölünmüştür (Ooka ve ark., 2003). NAC proteinleri modüler bir organizasyona

sahiptir. Çoğunlukla, NAC alanı olarak adlandırılan bir N-terminali -150-amino asit korunmuş DNA-bağlanma alanı (DBD), transkripsiyonel düzenleyici aktiviteye sahip daha uzun bir C-terminal alanına (CTD) bağlıdır (Şekil 2.14) (Ooka ve ark., 2003; Puranik ve ark., 2012; Welner ve ark., 2016). Bununla birlikte, yapıdaki birkaç varyasyon da tespit edilmiştir. Bu atipik NAC proteinleri, yalnızca NAC alanını, art arda tekrarlanan iki NAC alanını kodlayabilir (Puranik ve ark., 2012).



Şekil 2. 14. NAC proteinlerinin yapısı
(Puranik ve ark., 2012).

NAC alanı, *Petunia NAM* ve *Arabidopsis ATAF1 / 2* ve *CUC2* proteinlerinden alınan konsensüs sekanslarına göre tanımlandı (alan, genlerin ilk harflerinden adlandırıldı) *Arabidopsis CUC2* dahil birçok NAC proteini, bitki gelişiminde önemli işlevlere sahiptir. Bazı NAC genleri, yaralanma ve bakteriyel enfeksiyon sırasında up stream regüle edilirken, diğerleri viral dirence aracılık eder (Hong, 2016; Nakashima ve ark., 2012).

Birkaç tam bitki genomik sekansının yardımıyla yapılan kapsamlı araştırmalar, *Arabidopsis*'te 117, pirinçte 151, üzümde 79, kavakta 163 ve soya fasulyesi ve tütünde 152 NAC geni tanımlamıştır. Genom çapında transkriptom analizi ile sağlanan biyoinformatik destek,% 20-25 NAC genlerinin en az bir veya diğer strese işlev gördüğünü öngörmüştür (Puranik ve ark, 2012).

2.4.2.4. TCP TF Ailesi

Bitkiye özgü TCP transkripsiyon faktörü ailesi genleri, dallanma, yaprak gelişimi, hormon yolları, tohum çimlenmesi ve sirkadiyen saat gibi birçok bitki büyümesi ve gelişim sürecine dahil olur (Aggarwal ve ark., 2010; Danisman, 2016; Hong, 2016).

Hücresel düzeyde, çoğu TCP geni, ya siklin transkripsiyonunu doğrudan kontrol ederek ya da hücrelerin farklılaşmasını indükleyerek bitkilerin organlarındaki hücre çoğalmasını modüle eder (Aggarwal ve ark., 2010).

TCP proteinleri başlangıçta tanımlanmış ve karakterize edilen ilk üç aile üyesinden sonra adlandırılmıştır: Mısırdan (*Zea mays*) Teosinte Branched 1 (TB1), aslanagzından (*Antirrhinum majus*) Cycloidea (CYC), ve pirinçten (*Oryza sativa*) Proliferating Hücre Faktörleri 1 ve 2 (PCF1 ve PCF2) (Hong, 2016; Liu ve ark., 2019).

TCP proteinleri, TCP alanlarının sekans homolojisine göre Sınıf I ve Sınıf II olmak üzere iki sınıfa ayrılır. Bu sınıflar, ilgili NLS'lerin bileşimi, bHLH alanındaki ikinci helix uzunluğu ve bHLH alanı dışında bilinmeyen işlevselliğe sahip arginin açısından zengin bir alanın varlığı bakımından farklılık gösterir (Danisman, 2016).

Sınıf I, TCP alanında korunmuş bir dört amino asit delesyonuna sahiptir ve bazı Sınıf II genleri, bir glutamik asit-sistein-glutamik asit ve / veya TCP alanı dışında arginin açısından zengin bir R alanı gibi ek motifler sergiler (Liu ve ark., 2019).

Bu sözde R alanı, sınıf I TCP'lerde bulunmaz ve protein-protein etkileşimlerine aracılık eden hidrofilik bir a-helix veya coiled-coil yapı oluşturduğu tahmin edilmiştir (Danisman, 2016).

Sınıf II'nin üyeleri, TCP alanlarındaki dizi farklılıklarına göre CIN ve CYC sınıflarına ayrılabilir. CIN alt sınıfı *Antirrhinum CINCINNATA* (CIN) ile temsil edilir ve CYC alt sınıfı CYC ve TB1 ile temsil edilir. Sınıf II TCP proteinlerinin, diğerlerinin yanı sıra gelişim örneğinin, yaprak farklılaşması, aksiller meristem gelişimi, sürgün dallanması hormon sinyalleme ve savunma dahil olmak üzere birçok bitki işlemini düzenlediği bilinmektedir (Liu ve ark., 2019).

Çeşitli deneysel yaklaşımlarda, sınıf I ve sınıf II TCP proteinlerinin hedef gen promoterlerindeki GC açısından zengin dizileri tanıdığı gösterilmiştir. *Arabidopsis thaliana* genomu, sekans benzerliklerine göre Sınıf I ve Sınıf II TCP'lere bölünen 24 TCP TF'yi kodlar. Tüm TCPTF'ler, nükleer hedefleme, DNA bağlama ve aracılık eden protein-protein etkileşimlerinden sorumlu 59 amino asit uzunluğunda, kanonik olmayan bir bHLH alanı olan TCP alanını paylaşır (Hong, 2016).

TCP genlerinin çeşitli biyolojik süreçlerdeki kritik rolleri göz önüne alındığında, çeşitli bitkilerde TCP homolog genleri analiz edilmiş ve bu gen ailesinin analizleri tür genomları arasında gerçekleştirilmiştir. Örneğin *Physcomitrella patens*'de 6 TCP geni, mısırdan 29 buğdayda (*Triticum aestivum*) 66 *A. thaliana*'da 24, *Populus euphratica*'da 33 ve *Gossypium raimondii*'de 74 TCP geni tanımlanmıştır.

3. MATERYAL VE YÖNTEM

3.1. Patates (*Solanum tuberosum*) Genomunda Transkripsiyon Faktörü Kodlayan Genlerin Tanımlanması ve Karakterizasyonu

Patates (*Solanum tuberosum*) (Ensembl Plant Erişim: PGSC0003DMG400029744) cDNA sekans koleksiyonu transkripsiyon faktörü kodlayan genlerin tanımlanmasında kullanılmak üzere Ensembl Plant Genom Veritabanı'ndan (<https://plants.ensembl.org/index.html>) FASTA formatında indirilmiştir. Gen tanımlama analizleri, homoloji temelli metodoloji ile gerçekleştirilmiştir.

Plant Transcription Factor Database (PlantTFDB)'den (<http://plantfdb.gao-lab.org/>) alınan karakterize edilmiş Arabidopsis (*Arabidopsis thaliana*) ve pirinç (*Oryza sativa*) transkripsiyon faktörü gen setlerinin homologları patates cDNA sekanslarında konumlandırılmıştır. Bu amaçla erişilebilir bir analiz sunucusu olan Galaxy'de (<https://usegalaxy.eu/>) bulunan NCBI BLAST+ blastx aracı kullanılmıştır.

Galaxy, erişilebilir, tekrarlanabilir, açık kaynaklı, web tabanlı bir platformdur. Genomik araştırmalar için geliştirilmiş birçok aracı içinde bulunduran ve kalıcı olarak veri depolama imkânı sunan bu platform yıllar içinde gen ekspresyonu, genom montajı, proteomik, epigenomik, transkriptomik gibi birçok disiplin için kullanılmaya devam etmiştir (Afgan ve ark., 2018).

BLAST (Temel Yerel Hizalama Arama Aracı) bir web arayüzü aracılığıyla ya da bağımsız olarak kullanılabilen bir, dizi benzerliği arama programıdır (Altschul ve ark., 1990). Blast programları belirli bir sorgu dizisinin homologlarını belirlemek amacıyla dizi benzerliği araması için uyarlanmıştır. Nükleotit veya protein sorgularının tüm kombinasyonlarını nükleotit veya protein veritabanlarıyla karşılaştırmak için çeşitli Blast türleri vardır. Blastx analizi BLAST analizinin önemli algoritmalarından biridir, bir nükleotid sorgu dizisi, altı okuma çerçevesinin tamamına çevrilerek, protein dizisi veritabanıyla karşılaştırılır (McGinnis ve Madden 2004; Pirooznia ve ark., 2008).

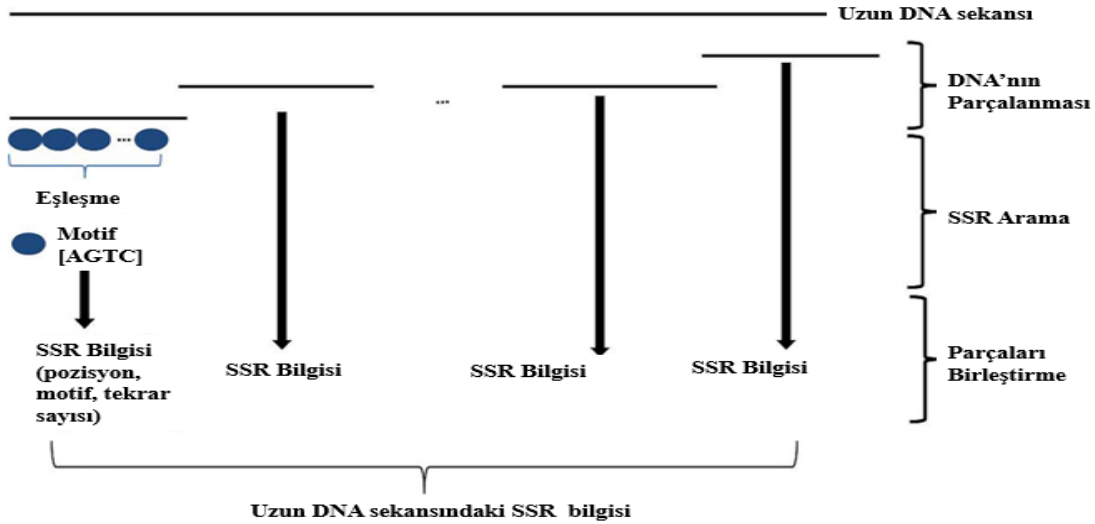
56.210 sekans içeren *Solanum tuberosum* cDNA koleksiyonu, transkripsiyon faktörü kodlayan genlerin tanımlanması için Blastx analizinde nükleotid sorgu sekansı olarak kullanılmıştır. Blastx analizi “e değeri” eşik değeri “1e-10”, maximum hits değeri

“5“olarak yapılmıştır. Hem Arabidopsis hem de pirinç transkripsiyon faktörü gen setlerinin patates homologları belirlenmiştir. Blastx analizi sonucunda cDNA sekans erişim numaraları, cDNA' ya denk düşen TF erişim numarası, özdeş eşleme yüzdesi, mismatch, gaps, e-değeri gibi verileri içeren dosyalar çıktı olarak alınmıştır. Galaxy 'ye yüklenen *S. tuberosum* cDNA sekans koleksiyonundan, tahmin edilen cDNA sekansları Phyton programa dili kullanılarak filtrelenip seçilmiştir.

3.2. Markör Geliştirme

Transkripsiyon faktörü kodladığı tespit edilen *Solanum tuberosum* cDNA sekanslarında, SSR'ların tanımlanması amacıyla genom genelinde SSR arayan, yan dizilerden primerlerin tasarlanmasını sağlayan, e-map işlemlerini gerçekleştiren GMATA (Genome-wide Microsatellite Analyzing Tool Package) (X.Wang ve Wang, 2016) yazılımı kullanılmıştır. GMATA yazılımı ile birlikte Perl, R ve Java yazılımlarının da arka planda çalışması gerekli programlar olduklarından kurulumu yapılmıştır.

GMATA, tüm genom düzeyinde SSR tanımlayabilen, bulduğu tüm lokuslara primer dizayn eden, elektronik harita oluşturan, polimorfizm taraması ve markör aktarılabiliğinin tespitini yapabilen bir yazılımdır. DNA ön işleme, SSR arama, SSR görüntüleme, SSR istatistikleri, markör dizaynı ve markör e-haritalanmasını içeren altı modülden oluşur. GMATA, DNA dizisindeki SSR'leri verimli bir şekilde tespit etmek ve SSR arama hızlarını artırmak için uzun bir DNA dizisini (varsayılan > 2 Mb'dir) uygun bir uzunluğa bölme stratejisi uygular. Her parçanın sonundaki kısa örtüşen bölgeler (varsayılan 20 bp), dizinin sonunda doğru SSR'ların tespit edilmesini sağlar (X.Wang ve Wang, 2016). Şekil 3.1 'de GMATA'nın SSR arama stratejisi verilmiştir.



Şekil 3. 1. Uzun bir DNA dizisi için GMATA'daki SSR arama stratejisini gösteren şematik diyagram (Wang ve Wang, 2016).

3.2.1. Patates cDNA sekanslarında SSR Tanımlama

Solanum tuberosum cDNA sekanslarında SSR tanımlamak için GMATA'nın "SSR Arama" modülü kullanılmıştır. FASTA formatında indirilen cDNA sekans dosyaları GMATA'ya aktarılmış ve motifler tanımlanmıştır. Uygulanan SSR arama parametreleri; en az 5 tekrarlı, en kısa 2 nükleotid ve en uzun 6 nükleotid olacak şekilde belirlenmiştir.

SSR tanımlama aşamasının sonucunda cDNA dizileri üzerindeki SSR başlangıç ve bitiş konumlarını, SSR motiflerini ve tekrar birimlerinin sayısını, SSR lokus bilgilerini içeren (.ssr), giriş dizilerinin toplam sayısı ve uzunluğunu içeren (.sat1) uzantılı dosyalar çıktı olarak alınmıştır.

3.2.2. Markör dizaynı

Tanımlanan SSR lokuslarını kapsayan sekans parçalarının, primer dizaynında kullanılması için GMATA'nın "Markör Dizaynı" modülü kullanılmıştır. Seçilmiş cDNA sekansları içeren FASTA formatındaki dosya ve "SSR Arama" modülü tarafından çıktı olarak alınan .ssr uzantılı SSR lokus dosyaları, sırasıyla "sekans dosyası" ve "SSR lokus dosyası" olarak "Markör Dizaynı" modülüne aktarılmıştır. Markör dizaynı için minimum ampikon boyutu 100 baz çifti, maksimum ampikon boyutu 250 baz çifti olarak, optimum tavlama (annealing) sıcaklığı 60 °C, yan dizi uzunluğu 400 baz çifti olarak tanımlanmıştır.

Markör dizaynı sonucunda her iki tarafta 400 baz komşu sekansı içinde ortalananmış tekrar lokuslarını içeren (.seq) uzantılı çıktı dosyası, ileri ve geri primer dizilerini, tavlama sıcaklıklarını, sekanslar üzerindeki primer konumlarını ve beklenen ampikon boyutlarını içeren (.mk) ve (.sts) uzantılı çıktı dosyaları alınmıştır.

3.2.3. Tür içi ve Türlerarası Genomik Markör Lokus Haritalaması

GMATA'nın "e-Map" modülü çalıştırılarak e-PCR işlemi gerçekleştirilmiştir. E-PCR algoritması ampikonlar oluşturmak ve onları patates cDNA ve genom sekansları üzerindeki markör konumlarına atamak amacıyla kullanılmıştır.

E-PCR için, seçilmiş patates cDNA sekansları içeren FASTA formatındaki dosya ve "Markör Dizayn" modülü sonucu elde edilen .sts uzantılı çıktı dosyası sırasıyla "sekans dosyası" ve "markör dosyası" olarak "e- mapping" modülüne aktarılmıştır.

Maksimum "mismatch" ve maksimum "indel" parametreleri '0' olarak ayarlanmıştır. (.emap), (.amp), (.frg) dosyaları çıktı olarak alınmıştır. cDNA üzerinde hesaplanan ampikon boyutları ve hedef konumları ile markörlerin ayrıntılı amplifikasyon modellerini içeren (.emap) çıktı dosyası alınmıştır.

Markörlerin polimorfizm potansiyeli, GMATA programının "e-Map" modülüne, hem patates (*Solanum tuberosum*) hem de yakın akraba tür domates (*Solanum lycopersicum*) kromozom sekansları yüklenerek belirlenmiştir. E-Map modülü ile *Solanum tuberosum* ve *Solanum lycopersicum* kromozomları üzerinde primer sekanslarının amplifikasyon profilleri elde edilmiştir.

3.3. Polimorfik Fonksiyonel Markörlerin Protein Anotasyonu ve 3D Modellemeleri

Solanum tuberosum kromozomlarında haritalanan ve polimorfizm potansiyelini türler arası 'mapping' analizi ile gösteren markörlerin peptid sekansı düzeyinde anotasyonları ve 3D (üç boyutlu) modelleme analizleri Swiss Prot server'ı kullanılarak gerçekleştirilmiştir, proteinin işlevi, post-translasyonel modifikasyonlar, domainler (kalsiyum bağlama bölgeleri, ATP bağlama bölgeleri, zing finger, homebox, SH2 ve SH3 alanları vb.) ve 3D yapıda polimorfizme dayalı olası değişimler belirlenmiştir. 3D yapı

oluřturma analizleri, polimorfik bulunan markör lokuslarının alel alternatiflerinin tamamı için gerekleřtirilmiřtir.

3.4. DNA İzolasyonu

DNA izolasyonunda bitki materyali olarak yerel eřide ait yaprak doku örneęi kullanılmıřtır. İzolasyonlar CTAB (cetyltrimethylammoniumbromide) protokolüne göre gerekleřtirilmiřtir. Bu amala 100 mg yaprak doku örneęi 800 µl CTAB ekstraksiyon tamponu [100mmol/LTris-HCl (pH 8.0), 20mmol/L EDTA (ethylenediaminetetraaceticacid, pH 8.0), 1.4mol/L NaCl, 2% (w/v) CTAB,1% PVP (polyvinylpyrrolidone)]. İinde paralanarak, 65°C sıcaklıkta 1 saat süre ile inkübe edilmiřtir. Lizis iřleminin ardından örneklere 600 µl kloroform:izoamil alkol (24:1) eklenmiř, örnekler 10 dakika süre ile 14,000 rpm hızda santrifüjlenmiř, üst faz temiz bir tüpe aktarılıp, izopropanol ile izole edilen DNA'nın öktürülmesi iřlemi gerekleřtirilmiřtir. İzopropanolde öktürülen DNA, pelet haline getirilmek üzere 10 dakika süre ile 14,000 rpm hızda santrifüjleme iřlemi yapılmıř, DNA peletleri %70'lik etanolde yıkanmıř, izole edilen DNA, TE tamponunda elüe edilerek, -20°C'de saklanmıřtır.

3.5. PCR Amlifikasyonu

PCR amplifikasyonları 1X Q5 reaksiyon tamponu, 0.2 mmol/L dNTP, 0.5 U Q5High-Fidelity DNA Polimeraz, 0.50 µmol/L ileri ve geri primer ve 5 ng kalıp DNA içerecek řekilde 25 µL hacimde hazırlanmıřtır. Tüm markörler için standart döngü kořulları řu řekilde uygulanmıřtır: Bařlangı denatürasyonu 98°C/30 saniye, 35 döngü 98°C/10 saniye, 60°C/20 saniye, 72°C 30 saniye, son uzama için 72°C 10 dakika. PCR ürünleri Qiaxcel Advanced Kapiler elektroforez sistemi (Qiagen, Hilden, Almanya) kullanılarak görüntülenmiřtir.

3.6. Kapiler Elektroforez ve Markör Doğrulama Analizi

Markör fragmanları, PCR amplifikasyonunun ardından bir Qiaxcel Advanced Kapiler elektroforez sistemi'nde yürütülerek ayrılmıřtır. Qiaxcel DNA High Resolution Kit (Qiagen) kapiler kartuřu kapiler elektroforez analizleri için kullanılmıřtır. Fragman

boyutlandırma ve boyut standartı hizalama için sırasıyla QX DNA Size Markör 25–500 bp v2.0 (Qiagen) ve QX Alignment Markör 15 bp-600 bp (Qiagen) kullanılmıştır. OM800 yüksek çözünürlüklü yöntemi, kapiler elektroforez analizlerinde 10 s'lik bir numune enjeksiyon süresi ile kullanılmıştır.

4. ARAŞTIRMA SONUÇLARI VE TARTIŞMA

4.1. Transkripsiyon Faktörü Kodlayan Genlerin Tanımlanması ve Karakterizasyonu

Solanum tuberosum transkriptomunu kullanarak transkripsiyon faktörü kodlayan sekansları belirlemek amacıyla, Arabidopsis (*Arabidopsis thaliana*) ve pirinç (*Oryza sativa*) transkripsiyon faktörü gen setlerinin homologları patates cDNA sekanslarında konumlandırılmıştır. Pirincin monokot türler için Arabidopsisin dikot türler için model organizma olması ve çok fazla çalışılmalarından dolayı transkripsiyon faktörleri için en güvenilir verileri elde edebileceğimiz türlerdir. cDNA sekansları genomik kütüphanelerden daha küçük olduğu için çalışmalar daha hızlı ilerler. Aynı zamanda genom üzerinde tahmin edilen gen bölgesini içeren sekanslar üzerinde markör geliştirmede, gen olarak tahmin ettiğimiz lokusun transkripsiyonel olarak aktif olduğunu garanti edemeyiz, yalnızca ilgili gen konumunun “aday gen” olduğunu söyleyebiliriz. cDNA koleksiyonu ise transkriptom olduğu için transkript edildiği kesindir.

Bu doğrultuda Blastx analizine “nükleotid query” sekans olarak 56.210 sekans içeren *Solanum tuberosum* cDNA sekansları kullanılmıştır. Sırasıyla 2296 sekans içeren Arabidopsis ve 2408 sekans içeren pirinç peptid sekansları ile Blastx analizi yapılmıştır. Blastx analizinde kullanılan sekans bilgileri Çizelge 4.1’ de verilmiştir. Blastx analizi sonucunda elde edilen çıktı hizalamanın eşleşme yüzdesi, mismatch -gap değerleri, e-value ve bit skor değerleri elde edilmiştir bu verilerle ilgili özet bilgi Çizelge 4.2. ve Çizelge 4.3.’de verilmiştir. *Solanum tuberosum* cDNA koleksiyonu ve Arabidopsis transkripsiyon faktörü sekansları ile yapılan Blastx analizi ile ilgili veriye <https://doi.org/10.6084/m9.figshare.19607067> linkinden, *Solanum tuberosum* cDNA koleksiyonu ve *Oryza sativa* Japonica transkripsiyon faktörü peptid sekansları arasında yapılan Blastx analizi ile ilgili veriye <https://doi.org/10.6084/m9.figshare.19615719> linkinden erişilebilir.

Çizelge 4. 1. Blastx analizinde kullanılan sekans bilgileri.

| Tür Adı | Blastx Analizi | Sekans Sayısı |
|--------------------------------------|--------------------|---------------|
| Patates (<i>S. tuberosum</i>) | Nükleotid Query | 56.210 |
| Arabidopsis (<i>A. thaliana</i>) | Protein veritabanı | 2296 |
| Pirinç (<i>O. sativa Japonica</i>) | Protein veritabanı | 2408 |

Çizelge 4. 2. Patates cDNA sekansları ile Arabidopsis TF gen setleri arasında yapılan Blastx analizi sonucunda elde edilen e- Değeri ve Bit skor özet tablosu.

| Patates cDNA ID | Arabidopsis TF ID | e-Değeri | Bit Skor |
|----------------------|-------------------|-----------|----------|
| PGSC0003DMT400006735 | AT1G52150.1 | 0 | 1438 |
| PGSC0003DMT400036765 | AT5G62000.4 | 0 | 944 |
| PGSC0003DMT400035592 | AT5G20730.1 | 0 | 602 |
| PGSC0003DMT400056951 | AT5G56270.1 | 1,07E-180 | 536 |
| PGSC0003DMT400060864 | AT5G06839.2 | 1,43E-148 | 424 |
| PGSC0003DMT400062861 | AT3G48160.1 | 3,25E-107 | 321 |
| PGSC0003DMT400079507 | AT4G39070.1 | 2,19E-69 | 210 |
| PGSC0003DMT400076759 | AT2G41240.2 | 1,76E-32 | 113 |
| PGSC0003DMT400011891 | AT4G12020.1 | 2,64E-13 | 70,9 |
| PGSC0003DMT400047302 | AT1G64100.2 | 1,00E-10 | 61,2 |

Çizelge 4. 3. Patates cDNA sekansları ile Pirinç TF gen setleri arasında yapılan Blastx analizi sonucunda elde edilen e- Değeri ve Bit skor özet tablosu.

| Patates cDNA ID | Pirinç TF ID | e-Değeri | Bit Skor |
|----------------------|------------------|-----------|----------|
| PGSC0003DMT400030829 | LOC_Os03g43930.1 | 0 | 1362 |
| PGSC0003DMT400068039 | LOC_Os10g33960.1 | 0 | 914 |
| PGSC0003DMT400062050 | LOC_Os02g03690.1 | 0 | 653 |
| PGSC0003DMT400022996 | LOC_Os01g62460.3 | 4,33E-180 | 536 |
| PGSC0003DMT400020839 | LOC_Os05g48870.1 | 7,92E-134 | 419 |
| PGSC0003DMT400063315 | LOC_Os08g41950.1 | 1,23E-72 | 226 |
| PGSC0003DMT400017414 | LOC_Os01g10504.1 | 1,82E-41 | 140 |
| PGSC0003DMT400084206 | LOC_Os12g13170.1 | 2,22E-26 | 108 |
| PGSC0003DMT400092871 | LOC_Os05g23780.1 | 4,22E-14 | 63,5 |
| PGSC0003DMT400018557 | LOC_Os02g02424.1 | 9,91E-11 | 60,5 |

e-Değeri belli bir sorgu sekansının, belli bir veri tabanında ürettiği hizalamaların rastgele olma olasılığını gösterir. Sorgu sekansı veri tabanında aratıldığında e-değerinin “1” olması neredeyse tamamen rastgele olma olasılığının yüksek olduğunu gösterir. E-Değerinin “0” olduğu hizalamalar en iyi değerlerdir ve hizalamanın rastgele olma olasılığının çok düşük olduğunu gösterir. Aynı şekilde bit skor yükseldikçe de hizalamanın rastgele olma olasılığı düşer. Yapmış olduğumuz her iki Blastx hizalmasında elde ettiğimiz e-değerler “0” veya 0’a oldukça yakın değerlerdir bu da Blastx analiz

sonuçlarının çok yüksek güvenilirlikte olduğunu ve gerçek homologileri bildirdiğini göstermiştir.

Blastx analizi sonucunda; Arabidopsis transkripsiyon faktörü peptit sekansları, patates cDNA sekanslarında konumlandırılmış 22.274 pozitif blast hits (eşleşme) sayısı elde edilmiştir. Eşleşmeler arasında 17.185 yinelenen 5089 benzersiz cDNA tahmin edilmiştir. Arabidopsis gen seti için patates cDNA koleksiyonunda 20.290 yinelenen transkripsiyon faktörü eşleşmesi, 1984 benzersiz transkripsiyon faktörü tahmin edilmiştir.

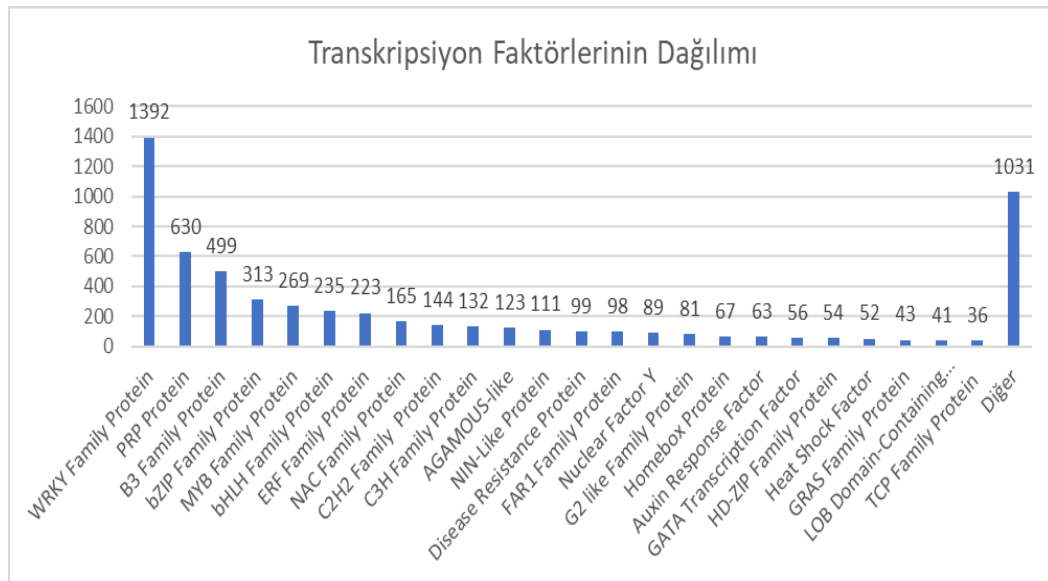
Pirinç transkripsiyon faktörü peptit sekanslarının, patates cDNA sekanslarında konumlandırılması sonucunda 17.893 pozitif blast hits sayısı elde edilmiştir. 13.227 yinelenen, 4379 benzersiz cDNA tahmin edilmiştir. cDNA kodlayan pirinç gen seti içinde 15.815 yinelenen transkripsiyon faktörü 1791 benzersiz transkripsiyon faktörü tahmin edilmiştir. Blastx analizi sonuçları ile ilgili veriler Çizelge 4.4’de verilmiştir.

Arabidopsis ve pirinçte tahmin edilen toplam 9468 benzersiz cDNA sekansları Excel ile filtrelenerek eşleşen ve eşleşmeyen sekanslar belirlenmiş ve her cDNA sekansından 1 tane olmak üzere markör geliştirilecek 6052 cDNA sekansı tahmin edilmiştir. Tahmin edilen sekanslar, Ensembl plant veri tabanından Galaxy platformuna aktarılan 56.210 cDNA sekans içeren *Solanum tuberosum* cDNA sekans koleksiyonundan, Python programa dili kullanılarak filtrelenmiştir.

Çizelge 4. 4. Blastx Analizi sonuçları

| | Blastx (<i>S. tuberosum</i> ve <i>A. thaliana</i>) | Blastx (<i>S. tuberosum</i> ve <i>O. sativa</i>) |
|----------------------------------------------------------------------|---------------------------------------------------------|-------------------------------------------------------|
| Pozitif Blast Hit Sayısı | 22.274 | 17.606 |
| TF kodlayan benzersiz cDNA sekansı | 5089 | 4379 |
| cDNA sekansına denk düşen TF kodlayan benzersiz peptit sekansı | 1984 | 1791 |

Blastx analizi ile elde edilen hem arabidopsis hem de pirinç transkripsiyon faktörü gen setlerinin, patates homologları tespit edilmiştir. Her cDNA’ya denk düşen TF erişim numaralarının hangi transkripsiyon faktörü ailesinin üyesi olduğu plantTFDB’den alınan peptit dosyaları ile belirlenmiştir. Transkripsiyon faktörlerinin gen ailelerine göre dağılımı Grafik 4.1’de verilmiştir.



Grafik 4.1. Patates cDNA sekanslarında tespit edilen TF kodlayan gen ailelerinin dağılımı.

Transkripsiyon faktörlerinin gen ailelerine göre sınıflandırma dağılımına baktığımızda en fazla bulunan TF'nin 1392 ile WRKY TF ailesi olduğu tespit edilmiştir.

4.2. SSR Markörlerinin Tespiti

Transkripsiyon faktörü kodladığını tespit ettiğimiz sekanslarda SSR lokuslarını belirlemek amacıyla GMATA yazılımının "SSR Arama" modülü kullanılmıştır. Seçilen 6052 cDNA sekansında 2-6 nükleotid içeren en az 5 tekrarlı SSR motifleri tanımlanmıştır. Uygulanan parametreler doğrultusunda toplam 1435 SSR lokusu tespit edilmiştir. En yüksek sıklıkta bulunan SSR tipi 773 ile trinükleotid tekrar motifi olduğu, toplam tanımlanan SSR'ların %51'ini oluşturduğu belirlenmiştir. Trinükleotid SSR'ların en çok bulunan motif olması, mononükleotid tekrarlarından sonra *Solanaceae* genomlarında en bol bulunan ikinci SSR tipini temsil ettiğinden ve kodlayan sekanslarda mononükleotid tekrarları genellikle tolere edilmediğinden (Chen ve ark., 2016), literatür ile uyumlu bir sonuçtur. Genomlardaki SSR dağılımının rastgele olmadığı ve trinükleotid SSR'ların kodlayan sekanslarda genellikle sık görüldüğü bilinmektedir (Uncu, 2018; Metzgar ve ark., 2002).

Trinükleotidler hariç tüm SSR tipileri, kodlamayan bölgelerle karşılaştırıldığında öngörülen protein kodlama dizilerinde önemli ölçüde daha az sıklıkta bulunmaktadır. Bu durum, kodlama bölgesinde çerçeve kayması mutasyonuna karşı negatif seçime atfedilebilir. Mutasyon baskısının ve spesifik tek amino asit uzantıları için muhtemelen

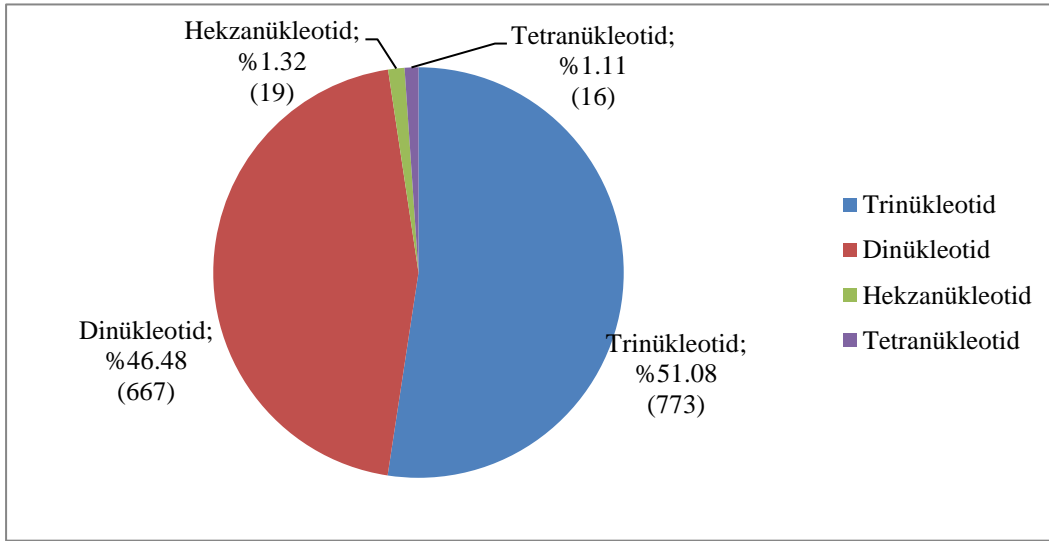
pozitif seçimin bir sonucu olarak trinükleotid tekrar sayısının kodlama bölgesindeki frekansı ikiye katlanır (Morgante ve ark., 2002).

Trinükleotid olmayan SSR türlerini etkileyen uzunluk mutasyonları, DNA'yı kodlamada meydana geldiklerinde çerçeve kayması mutasyonlarına yol açacaktır ve bu mutasyonlara karşı seçim, bunların sabitleme şanslarını azaltacaktır. Sabitleme hızına bağlı olan zenginleştirme, kodlama bölgelerinde de aynı şekilde azalacaktır. Buna karşılık, üçe bölünebilen tekrar uzunluğuna sahip SSR'lar (trinükleotid ve hekzanükleotid tekrarları) çerçeve kayması mutasyonları oluşturmayacağı düşünülmektedir. Kodlayan bölgelerde trinükleotid olmayan SSR'ların tek motifleri içeren uzunluk mutasyonlarına maruz kaldıklarında negatif olarak seçilmiş çerçeve kaymaları oluşturması ve bu tür mutasyonların, SSR uzunluğundaki değişikliklere katkıda bulunan en yaygın çeşitlilik olduğudur. Buna karşılık, trinükleotid SSR'lar, tek motifli uzunluk mutasyonları yoluyla çerçeve kaymaları üretmez. Trinükleotid SSR'ların üretim hızının kodlama durumundan önemli ölçüde etkilenmez. Bu durum hem kodlayan hem de kodlamayan trinükleotid SSR'ların benzer tekrar genişleme oranlarına tabi olduğunu göstermektedir. Daha spesifik olarak, Trinükleotid SSR'ların genişlemesi kodlama yapan ve kodlamayan bölgelerde diferansiyel seçici basınçlara tabi değildir, oysa trinükleotid olmayan SSR'lar kodlama bölgelerinde daha fazla saflaştırma seçimine tabidir (Metzgar ve ark., 2000)

Çalışma kapsamında elde edilen sonuçlara göre patates transkripsiyon faktörü kodlayan sekanslarında en sık görülen ikinci SSR motifi 667 ile dinükleotid tekrarları olmuş ve toplam SSR'ların %46,4 'ünü oluşturduğu belirlenmiştir. Veri kümesindeki en az bulunan SSR motifinin tetranükleotid olduğu belirlenmiştir. SSR motiflerinin tekrar eden nükleotid sayısına göre dağılımı ve bulunma frekansı Çizelge 4.5'te ve Grafik 4.2'de verilmiştir.

Çizelge 4. 5. SSR motiflerinin cDNA sekanslarında bulunma sayısı ve frekansı.

| Motif (-mer) | Toplam | Frekans |
|----------------|--------|---------|
| Trinükleotid | 773 | 51.0 |
| Dinükleotid | 667 | 46.4 |
| Hekzanükleotid | 19 | 1.3 |
| Tetranükleotid | 16 | 1.1 |
| Toplam | 1435 | 100 |



Grafik 4.2 Patates cDNA sekanslarında tanımlanan SSR motiflerinin grafik üzerinde dağılımı.

Solanum tuberosum cDNA sekans koleksiyonunda tanımlanan tekrar lokusları incelendiğinde “TA” nükleotid tekrarları 145, “AT” nükleotit tekrarları 131 kere olmak üzere en çok tekrar eden motifler olarak bulunmuştur ve tanımlanan tüm SSR'lerin %19.22 'ini temsil etmektedir. En çok tekrar eden trinükleotid tekrar “CAA”, tetranükleotid tekrar “CCCA”, hekzanükleotid tekrar “GCAGGT” olarak tespit edilmiştir. Genel olarak bakıldığında da AT/TA motifleri bakımından zengin olduğu gözlenmiştir. Bitki genomların AT tekrar motifleri bakımından zengin iken insanlarda AC veya TG tekrarı en yaygın olanıdır. Bunlar, tüm bitki ve hayvan genomlarını ayıran genel özelliklerdendir. Ayrıca bitkilerde en bol bulunan SSR tipi olan AT dinükleotidlerini palindromik oldukları için kütüphanelerden izole etmek zordur (Powell., 1996).

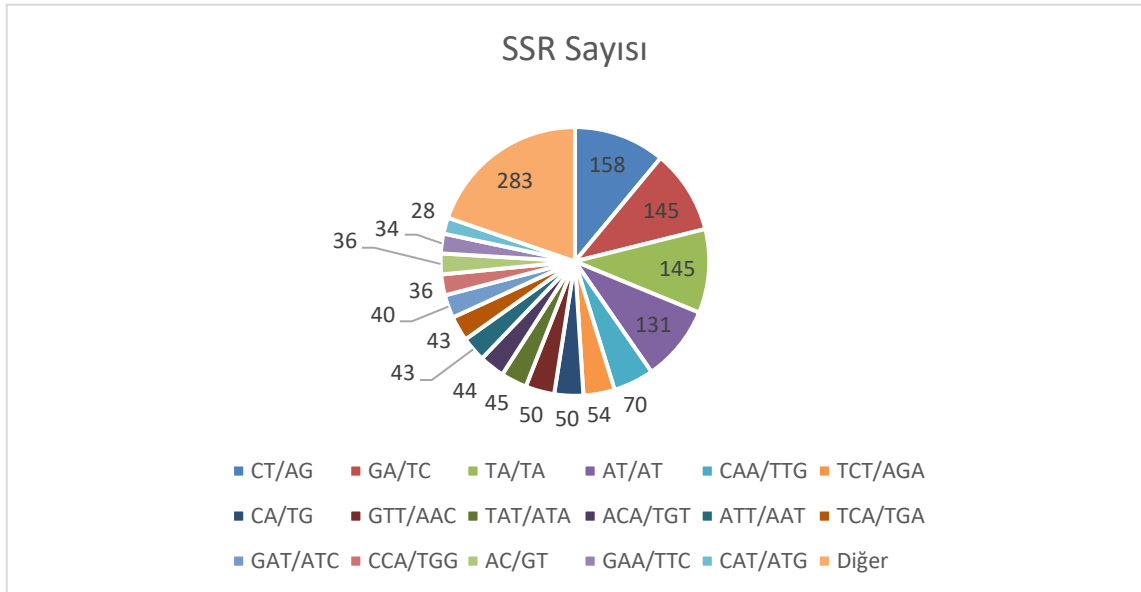
Biber genomunda yapılan bir çalışmada farklı motiflerin dağılımı açısından “AT” motifi biberdeki toplam motiflerin yaklaşık %95'ini oluşturmuş aynı zamanda incelenen tüm genomlar arasında en sık görülen motif olduğu bulunmuştur. Ek olarak, AT/AT motifi, en baskın dinükleotid olarak biber, domates ve patatesteki toplam motifinin %70'inden fazlasını oluşturduğu tespit edilmiştir (Cheng., 2016).

1435 lokus için toplam 53 farklı motif tespit edilmiştir. Dinükleotid tekrarlar tüm SSR motiflerinin yaklaşık %40'ını oluşturmuştur. SSR motifleri içinde toplam 158 olmak

üzere en fazla bulunan dinükleotid tekrar “CT/AG” olarak tespit edilmiştir. En fazla bulunan trinükleotid tekrar sayısı ise 70 ile “CAA/TTG” olarak tespit edilmiştir. SSR motifleri ile bilgiler Çizelge 4.6’te ve Grafik 4.3’te verilmiştir.

Çizelge 4. 6. Seçilen *Solanum Tuberosum* cDNA sekanlarında bulunan SSR motifleri, bulunma sayıları ve frekansları verilmiştir.

| Motif | Toplam | Frekansı (%) | Motif | Toplam | Frekansı (%) |
|---------|--------|--------------|---------------|--------|--------------|
| CT/AG | 158 | 11,0 | TGC/GCA | 16 | 1,1 |
| GA/TC | 145 | 10,1 | GTG/CAC | 10 | 0,7 |
| TA/TA | 145 | 10,1 | CCG/CGG | 6 | 0,4 |
| AT/AT | 131 | 9,1 | CCCA/TGGG | 6 | 0,4 |
| CAA/TTG | 70 | 4,9 | AGT/ACT | 5 | 0,3 |
| TCT/AGA | 54 | 3,8 | TGAG/CTCA | 3 | 0,2 |
| CA/TG | 50 | 3,5 | GCAGGT/ACCTGC | 3 | 0,2 |
| GTT/AAC | 50 | 3,5 | AAAT/ATTT | 2 | 0,1 |
| TAT/ATA | 45 | 3,1 | TCAC/GTGA | 2 | 0,1 |
| ACA/TGT | 44 | 3,1 | GCC/GGC | 2 | 0,1 |
| ATT/AAT | 43 | 3,0 | GATGAG/CTCATC | 2 | 0,1 |
| TCA/TGA | 43 | 3,0 | CAAAAT/ATTTTG | 2 | 0,1 |
| GAT/ATC | 40 | 2,8 | TCG/CGA | 2 | 0,1 |
| CCA/TGG | 36 | 2,5 | GC/GC | 2 | 0,1 |
| AC/GT | 36 | 2,5 | ACTTGA/TCAAGT | 2 | 0,1 |
| GAA/TTC | 34 | 2,4 | ACCAAA/TTTGGT | 2 | 0,1 |
| CAT/ATG | 28 | 2,0 | TGTA/TACA | 2 | 0,1 |
| TAA/TTA | 25 | 1,7 | TCTGGT/ACCAGA | 1 | 0,1 |
| CTT/AAG | 24 | 1,7 | TAGT/ACTA | 1 | 0,1 |
| GGA/TCC | 23 | 1,6 | CTCCA/TGGGAG | 1 | 0,1 |
| CTG/CAG | 22 | 1,5 | CTGATG/CATCAG | 1 | 0,1 |
| AGG/CCT | 19 | 1,3 | TAGTAA/TTACTA | 1 | 0,1 |
| GAG/CTC | 19 | 1,3 | GAAGTT/AACTTC | 1 | 0,1 |
| CTA/TAG | 19 | 1,3 | TGATTT/AAATCA | 1 | 0,1 |
| GCT/AGC | 18 | 1,3 | CCAACA/TGTTGG | 1 | 0,1 |
| ACC/GGT | 18 | 1,3 | CCCTCG/CGAGGG | 1 | 0,1 |
| GTA/TAC | 18 | 1,3 | | | |



Grafik 4.3. SSR motiflerinin dağılımı.

4.3 Primer Dizaynı

Toplam tespit edilen 2-6 nükleotid tekrarlı 1435 SSR lokusunun amplifikasyonu için PCR primerleri dizayn edilmiştir. Bu amaç doğrultusunda GMATA'nın "Markör Dizaynı" modülü kullanılmıştır. Markör dizaynı modülü sonucunda (.seq) uzantılı çıktı dosyası, her iki tarafta 400 baz komşu sekansı içinde ortalanmış tekrar lokuslarını içermektedir. İlgili veriye <https://doi.org/10.6084/m9.figshare.19607178> linkinden erişilebilir.

İleri ve geri primer dizilerini, tavlama sıcaklıklarını, sekanslar üzerindeki primer konumlarını ve beklenen ampikon boyutlarını içeren (ssr.mk) ve (.sts) uzantılı çıktı dosyaları alınmıştır. İlgili veriye <https://doi.org/10.6084/m9.figshare.19607085> linkinden erişilebilir. Çizelge 4.7'de markör dizaynı istatistikleri verilmiştir.

Çizelge 4. 7. Markör dizaynı istatistikleri.

| Markör dizaynı istatistikleri | |
|------------------------------------------|------|
| Toplam SSR lokusu sayısı | 1435 |
| Dizayn edilen toplam primer sayısı | 1210 |
| Primerlerin çoğalttığı SSR lokusu sayısı | 845 |

4.4 Markör Haritalama

Amplikonlar oluşturmak ve onları patates cDNA sekansları üzerindeki markör konumlarına atamak için e-PCR algoritması kullanılmıştır. Bunun için GMATA'nın "e-Mapping" modülü kullanılmıştır. (.emap) (.amp), ve (.frg) dosyaları çıktı olarak alınmıştır. Markörlerin tür içi ve türler arası polimorfizm potansiyeli, GMATA programı "Mapping" modülüne hem Patates (*Solanum tuberosum*) hem de yakın akraba tür domates (*Solanum lycopersicum*) kromozom sekansları yüklenerek belirlenmiştir. Mapping modülü *S. lycopersicum* kromozomları üzerinde primer sekanslarının amplifikasyon profillerinin elde edilmesinde kullanılmıştır. *Solanum tuberosum* cDNA sekanslarında geliştirilen 845 SSR markörünün *Solanum tuberosum*'un 12 kromozomu üzerinde haritalanması sonucunda 535, *Solanum lycopersicum*'nin 12 kromozomu üzerinde haritalanması sonucunda 99 SSR markörü listelenmiştir. Bu sonuçlar geliştirilen yeni markörlerin polimorfik olduğunu ve türler arası aktarılabilir olduğunu göstermiştir. *Solanum tuberosum* için ilgili veriye <https://doi.org/10.6084/m9.figshare.19640712> linkinden, *Solanum lycopersicum* için ilgili veriye <https://doi.org/10.6084/m9.figshare.19640751> linkinden erişebilir. Çizelge 4.8'de patates ve domates kromozomlarında haritalanan markör sayıları verilmiştir.

Çizelge 4. 8. Patates ve Domates kromozomlarında haritalanan markör sayıları

| Kromozom | <i>Solanum tuberosum</i> | <i>Solanum lycopersicum</i> |
|-----------------|---------------------------------|------------------------------------|
| 1 | 54 | 17 |
| 2 | 65 | 12 |
| 3 | 56 | 6 |
| 4 | 60 | 14 |
| 5 | 64 | 10 |
| 6 | 47 | 7 |
| 7 | 24 | 3 |
| 8 | 32 | 5 |
| 9 | 37 | 8 |
| 10 | 16 | 6 |
| 11 | 40 | 2 |
| 12 | 40 | 9 |
| Toplam | 535 | 99 |

Bununla birlikte amplikonların başlangıç ve bitiş konumları ve beklenen amplikon boyları elde edilmiştir. Amplikon boylarına baktığımız zaman exon-intron bağlantı noktalarına düşen ve düşmeyen markörler tespit edilebilmektedir. Örneğin, >MK587 ve >MK431 markörlerinin cDNA ve genomdan elde edilen amplikon büyüklükleri

karşılaştırılarak intron bölgelerine denk düştükleri öngörülmektedir. Çizelge 4.9’ da E-map sonuçlarının bir alt kümesini gösteren özet bir tablo verilmiştir.

Çizelge 4. 9. Patates kromozomlarında yapılan E-map sonucu

| Kromozom | Markör ID | Kromozom ID | Amplikon Boyu |
|----------|-----------|-------------|---------------|
| 1 | >MK123 | CM034976.1 | 886/240 |
| | >MK409 | CM034976.1 | 194/208 |
| 2 | >MK587 | CM034977.1 | 1188/215 |
| | >MK805 | CM034977.1 | 234/234 |
| 3 | >MK266 | CM034978.1 | 140/155 |
| | >MK296 | CM034978.1 | 186/186 |
| 4 | >MK139 | CM034979.1 | 120/134 |
| | >MK187 | CM034979.1 | 147/147 |
| 5 | >MK431 | CM034980.1 | 1329/200 |
| | >MK441 | CM034980.1 | 103/103 |
| 6 | >MK3 | CM035017.1 | 242/245 |
| | >MK30 | CM035017.1 | 249/247 |
| 7 | >MK252 | CM035018.1 | 462/196 |
| | >MK272 | CM035018.1 | 228/228 |
| 8 | >MK258 | CM035019.1 | 153/156 |
| | >MK276 | CM035019.1 | 178/182 |
| 9 | >MK579 | CM035020.1 | 1649/210 |
| | >MK585 | CM035020.1 | 124/124 |
| 10 | >MK246 | CM035021.1 | 1165/242 |
| | >MK259 | CM035021.1 | 152/152 |
| 11 | >MK827 | CM035022.1 | 131/131 |
| | >MK829 | CM035022.1 | 214/217 |
| 12 | >MK199 | CM035023.1 | 3044/239 |
| | >MK504 | CM035023.1 | 153/159 |

Markör haritalama sonucunda alınan çıktı dosyası (.emap), cDNA üzerinde hesaplanan amplikon boyutları ve hedef konumları ile markörlerin ayrıntılı amplifikasyon modellerini içermektedir. E-mapping sonucunda 1210 adet primer çiftinin toplam 845 SSR lokusunu çoğalttığı bazı SSR lokusları için birden fazla alternatif primer dizayn edildiği tespit edilmiştir. Çizelge 4.10’da dizayn edilen primerlerin listesi ve sekans bilgileri verilmektedir.

Çizelge 4. 10. Dizayn edilen Primer dizileri.

| MK ID | İleri Primer (5'>3') | Geri Primer (5'>3') | MK ID | İleri Primer (5'>3') | Geri Primer (5'>3') |
|-------|------------------------------|-----------------------|--------|-----------------------------|------------------------|
| >MK1 | GTCACGAAGAAAGCGGTAG | TGATCCCATGAATGCAAGAA | >MK423 | TGGGTGCTTTCCAAAACCTA | TTGAGTAACCCCATTTGATGC |
| >MK2 | CACCCACAACCAAGAAAAC | GCCTTCCCAAATTCCTCATA | >MK424 | GGACAAGGATTCAGAAGCACA | GGAAAATTGCCCTGATGATG |
| >MK3 | GCTTAGAGGCAGCAAAGCAA | CTTTTCCATCCCAAATCGTC | >MK425 | TGAGAACCCGACTTTCCAAC | GGTCAACATACGTGGCATCA |
| >MK4 | GCCTAATAAGCGCCATGCTA | CTCCATTCCCATGCATTTCT | >MK426 | CTCTGTTCCACAGCACCTCA | GGTGGTGAGCACTAGGAGGA |
| >MK5 | CGAGGTAGTGCTGGAAAAGG | CGCACTGCATGTTTCTGAGT | >MK427 | ACGTATATAAGGCGCTAAGGAG T | CAACGAGCTAGGAATTGAGTGA |
| >MK6 | TCACCCTTAATTCTAGGAGGAC A | ACGAGTTGCTTCCAGATGCT | >MK428 | GGCAAGTTCCAGACTTCGAG | GTGGCATGGACTAAGGGGTA |
| >MK7 | ATTTGGTGGTCCGACTTTTG | CGAGGTGACACTGGTCTACG | >MK429 | GCACGACAAATCATCCTGAA | TCTGGTTTCCCTCAGTCCTC |
| >MK8 | GGTTCAAAGGCTCAGCTCAA | CATATTGCACGGCAGGATAA | >MK430 | AAGCTTACACGGTGGACTCG | AGAGCATCAGCCAAACCAAG |
| >MK9 | GGATGAAGAAGAGGGATTAGT AGA | TTCATCTCTTGACAACCCTGA | >MK431 | CCACAGATCTCATACTCAAAGC A | GCAGCAGTCTTAGGGACACC |
| >MK10 | TCCCTTGTTGAAGAAAATCC | GCATCTTCATGAGGTCTCCAA | >MK432 | GGCATCATGAGCAAGGAAAT | ATGTTGCTGCTTCTGGAGGT |
| >MK11 | ATTTGCCTCTCCTCCTCCT | TTGTGCCTCATGCAACATTT | >MK433 | GTGGGAGGTGGATGTAGGAG | GGAGGAGGAAAATGGCTAGG |
| >MK12 | ATTTCCGAAGGGCAAAACT | AACTTGAACGACCGCATCTT | >MK434 | CCAGTGCAGGCTACTGATGA | CACCCCAATGTTGGAACCTCT |
| >MK13 | CTGGATTCACATGGCTGTTT | GCATCGTTCCCATCCATAAC | >MK435 | CCCAAAACACTTTCGGTTTC | GTTTCCCCATTTCCGTTACA |
| >MK14 | ACACATTGACGCGCAAATTA | GCGGGTGTCCATGTATTCTT | >MK436 | TGCTCAGCTCCGTCTAGTCA | CATGACGCATCAAGAATAGTCC |
| >MK15 | CCAGAAGTAAACCGGTCGAA | ATCCTTAATCCGGCGAAGAT | >MK437 | GAGCTGGAGAAAATGCATGA | AGGAAAATCTGAACCTGACC |
| >MK16 | ACTGCCAGCCGATTAAATGT | GTGCTGCTTTGAAACTTGGA | >MK438 | TCACAAGATGCCTCCTCCAG | TTGAGCAGGTTGATGTGGAC |
| >MK17 | TCGATCTTACTGCACCGAAA | GTGCTGCTTTGAAACTTGGA | >MK439 | GGGACTCTCCATCATCAAA | TCCAGAGGATGAATCGTTGTT |
| >MK18 | ACTTGGGAACCAACCATGTC | GCTTGAGGAAGAGGGAAGGT | >MK440 | AACGACAACGACGATGACAA | GACTCCGATCCTGAGTGCTG |
| >MK19 | ACCTCCCTCTCCTCAAGC | TTCATGGCTCCTAGCTCCTC | >MK441 | TGCCAGTCCGATACTCTTCC | GCTTCGTCAGTGCTGTGGTA |
| >MK20 | AGTGGCAAGGTCTGCGTACT | ACCCCTCTTTCCAGGTTTTG | >MK442 | AACACAATTGCAGCCAATCA | AGTCCTCAGGGATTCTTGGT |
| >MK21 | AGTGGCAAGGTCTGCGTACT | TTGTAAGCCTGCAGTTGACC | >MK443 | CCCTTGACACAGGCTTCTTC | TTTGGAAAGTACGAGCAACA |
| >MK22 | CGCGAGTCACACATCATTTT | TCATTTCCAACCAAACCACA | >MK444 | CAATGGGATTGCAACAACAA | GGCTAATGCCAATGAAGGAG |
| >MK23 | CATTGGAACCTCAGCCAGAT | CTTGGCTGCAACAACCTACCA | >MK445 | TCCCCATTCAACCCAAGATA | CAACAAACCAAACACAGATCG |
| >MK24 | CCACGAGTCCAACCTTTGTT | CAGACCTGTCCGAGATAGCC | >MK446 | TTATGGCGATGCTTCATTCA | AGCTCTTCCAATGCCAGAAA |

| | | | | | |
|-------|-------------------------------|-------------------------------|--------|-----------------------------|------------------------|
| >MK25 | TCTTCCACGAGTCCAACCTT | CAGCGCAGGTACAACAGAAA | >MK447 | GGCGATTTAGAACGTTTGGA | TCCTTCTCAACGCAAGAACA |
| >MK26 | TGGCCTAAATGGTTCCTCAA | TCACCCTTCCTGCAATTACC | >MK448 | GGCACCCTATGTGCTCCTT | ACATTGGCAACGACAGCTC |
| >MK27 | GGTAATTGCAGGAAGGGTGA | TGACAACAAAGCAAGGCAAT | >MK449 | CTGATTTGCACCAGTCTCA | CAGCAGGTTGTGACATGGAC |
| >MK28 | CCCCTCTCTCTCCTCTC | GGCATCTGGTCAAGGTCAAT | >MK450 | GCAAGGTACCCGCAAATGTA | TCAACAATTCCATCTCCATCC |
| >MK29 | TGGCTCCTAAGGCAGAAAAG | AGGATGCACTTGTTTCAGCA | >MK451 | GCATTTAGCCACAAGTGCAA | TAACCAATGTGCTTGGCATC |
| >MK30 | TCTCAATAGCCCTCTTATGTAG ATG | GGCCAATACCTTTACTGCTAGTC | >MK452 | CGTTCAGCTCACTCAGTTTCC | GGGGTTGAGATTTGGGTTTT |
| >MK31 | ATCGTCATTCATTGGGAAGC | TTTGGACTTGCTTGTGCTGA | >MK453 | GATGCAGCTGATGTCCAAGA | TCACTGCTGCTAATGCTGCT |
| >MK32 | CCTTCAAGAATCAACGACGA | GGATCCATCCGAAACATGA | >MK454 | CCCAAGCCCAAATACAAGAA | GGAGAAGCATGAGCACCATT |
| >MK33 | CAGATCATGTTTCGGATGGA | TGTAGGAACGCGCATGTAAG | >MK455 | CACACTGAGGCTTTGAGCAG | ACTGACCTGACGAGCCCTTA |
| >MK34 | CCAACATGCCTAATGCTGAA | GCTTAGTGCCTGAAGCTCGT | >MK456 | CATTTGCAGCGTTATGGAGA | TCCTTCCTTCAATTCATACCC |
| >MK35 | ATGGTGGGGAAGCTGAAGAT | CAAAGACTTGTCGCTGATGC | >MK457 | GAATCTAGCCCCACGAACAA | TCCTACATTTGCTGCTGTCG |
| >MK36 | ATGGTGGGGAAGCTGAAGAT | CCACAAAGACTTGTGCGTGA | >MK458 | CAAAATCGTTGGGTGGTTTT | CCGGAGTCGGTGTGTTATT |
| >MK37 | AATCCGTGCTTCGAGTCTGT | GCAGTCAGTAAGGCTTTCCA | >MK459 | CTCACAAGCCCAATTTACACC | GCTCCACCATTATTCTCTTTGC |
| >MK38 | CATCAACACGGCCATTACAG | ACGCTTTGGCTAATGGATTG | >MK460 | TAGGAGGAGGAGGTGGTGGT | CCGCGGATTATATGTTGCTG |
| >MK39 | ATTTGAGGGCTCGATGAAGA | CGTAGGACGATGCAATGAGA | >MK461 | TCGACAGCAACTCTGTTCAA | CCGCGTAAAATATGATGAGGA |
| >MK40 | GGGTTTCTGTCATCTCTCTCT C | TCCGTCGTCATCCATCAGTA | >MK462 | TCAACAACAACCGCCTGATA | GCCTCTTCTGCTGCTACCTC |
| >MK41 | TCATGACAACAACCTCATAGGG | GATGATCACCCAAAATTCCA | >MK463 | CATAGGACTGGCATGGGAAT | GTTGATCCCCCTCTCTCGATG |
| >MK42 | GGAGGTGGATCACATGAAGG | TGGATTTGGTTTCTCTCTGGA | >MK464 | TATTGGACCCATTGCCCTTT | CCCATGTTGGGGCTTACATA |
| >MK43 | GGAGTTCTCTCGCCACAATC | AGGGCTGGCAGTTGTTTAGA | >MK465 | TCTTCACCTGTTGCAGCACT | TTACGTGCATCACCATCACC |
| >MK44 | CAAATGATGATGCCACAAGC | ACGATGACGATGTTGAAACG | >MK466 | AGCGACACGGATAGTGAGGA | CTCCACCAGAAACATCAGC |
| >MK45 | GCAAAACCCAAGAAACCAAA | TGGTGCAACAACATCAGTAACA | >MK467 | TGCACTACTTTCACTATGTGGT A | TGCAAAATTATACCACCAGCAG |
| >MK46 | GGAAGAATACGCCAAAGTCG | TCCAACAACACCCCAAGAAT | >MK468 | CCCAAAACACACTGCACTCAC | TTTGCGGAATAGACTGAGACA |
| >MK47 | CAGCTTGGGCTAGTTTCAGG | CAACGTTTGATCAGATGGGTAA | >MK469 | CTCCGAAGAGGATTCTGCTG | GCTTTTATTACGCCGGTTA |
| >MK48 | CCATTTCCATGGCTTCAACT | GGAAGAGGAACGGAAGAAGG | >MK470 | CCCTTTGAACGGTGCTATGT | AGGAGAAGGAAATGGGTTCCG |
| >MK49 | ACGACGAGGATGAAGTCTC | TCCATTTCTTACCATTCC | >MK471 | CGGATGCAGTGAGCAGTTTA | GATTATCTGGTGTGATGGGAAA |
| >MK50 | TGCCTTCTCAATGTGTGGA | AAAATAGTCCGATATGTAGCTAG GG | >MK472 | CCATTCGAAACACTCAAACC | TCAGATGCATTCTGATCTCC |

| | | | | | |
|-------|-----------------------------|------------------------|--------|-------------------------|-------------------------|
| >MK51 | CTGTTGGAGGTGGATGTAGGA | GTTGATGATGGTGGCTAGGG | >MK473 | GAAAGTGTTAGTGGGTGTCAGAT | CAAAGAGCTTAATGGCAGGA |
| >MK52 | GAGGAATTGCAAATGTCAAGG | TTTGAAGGATCAAGCCAAGC | >MK474 | GGAAGGTAGTGGCAGTGGAA | TTCACACCACACTAGTCATTTGC |
| >MK53 | TTTGAAGGATCCGAACAACA | CGATCCAACCTCCCATAA | >MK475 | TCTGCATAGGAGCATTACAAA | CGCGGATATTTTACCAAGT |
| >MK54 | AACACGCACACACAGAGAGC | TGCGTTGGAGAAAATGAACA | >MK476 | TCCAGCTCTTCCTTCGAAAA | TCAGATCTTCGGGATCCTTG |
| >MK55 | TAGGAGAAGAACCCGTTCCA | GCGGCAAACATAATTCATCA | >MK477 | TAGTGGGATTTTCGCTGGGTA | CGGGATTTGAAGGCATCTTA |
| >MK56 | ATCCCCAAAACCCAATTCT | AGTCATCTCCGGCTCCCTAC | >MK478 | CTCCCTTGCAAATTCCTCCT | TTGGAATTGTAGCCCAAAGC |
| >MK57 | CCCTTTGCAAGTTCTTCTTCA | AGCAGCAACAGCAACCCTAA | >MK479 | AACTGTATGTGCCTCGGTGA | TGATCCCAAACAATCCAATC |
| >MK58 | CTCTTCAAACCGGGTCATGT | TGTAGCCAAGGAAGGCATTT | >MK480 | CACCCTTGCAAAGAGAGAGC | TGGTTGTGAACCCTATCCAAG |
| >MK59 | GAAGCCAGTGAGATTATCAGAT | TGGCCTTTCTCTTGTTGTTG | >MK481 | ACCCTTGCAAAGAGAGAGCA | TCAAGAAGATCCTGCCAAT |
| >MK60 | GT CAAGCTCACAAGCAACTACCA | TCCAAAAGCAGCAAACCTCA | >MK482 | TGTGCCCTTTCCTCAATCA | TCTCGATCCGCTTCAACTCT |
| >MK61 | TTATGGTCCCGATGTTAAGGA | TTTCCACTTCTTTGGCCTTG | >MK483 | GGCTTTTGGAGGAATGAACA | ATTGTTTACCCTTCGTTGC |
| >MK62 | GGGTCTGTTGACCCCTCAT | CAACTACAAGCTGGAAAAGGA | >MK484 | TCTTGCCACTGTAATCGGTGA | ATGCTTGCCATGTGATGTGT |
| >MK63 | ATCTTTCCAAGCAAGGACGA | GGACTCGCTGTCCATATACCA | >MK485 | CGGAGGCGCAATAATGTTT | GGTGGTGTGATGAGCTCTG |
| >MK64 | TGCAGCTAGGATGCAAAGAA | CCACCACTTGAACACCAACAA | >MK486 | TGATGCACTCAAGGCTCAAA | GATGATGTCCTTGTTCACAGA |
| >MK65 | GGATCCTCATCTTCATCACCA | CCGCTCCAATATTAACACC | >MK487 | CCTTACCACAACCTTCTCCTT | GTTGATGATTGCCTCGACCT |
| >MK66 | TTGAAAGAAGCCCTCCCTCT | AGAATAGAGGCCAAATCAGCA | >MK488 | TGGCGTATGAATTGGTCTCA | TTTGAAAAGGAGGCCAAGAA |
| >MK67 | CTAGCACCATGAAAGCCACA | TCAAAGGAGGAGAAGGTGGA | >MK489 | CAATTGGAAGCCTTGGAGAA | GGGAAGCATTGATGAGGA |
| >MK68 | TCCACCTTCTCCTCCTTTGA | CGCAATGATTCTGGTCCTTT | >MK490 | AGCTTGCCTGGAAGAACTGA | TTGCTGGAGGCATAGTTGGT |
| >MK69 | TGACGATGACGAAGATCCTG | CATCATCAGAAATCGAGCAATC | >MK491 | AAACCCGAATCACCTCTTC | TCGTAGCTTCGTTTCGTTCT |
| >MK70 | GGTGCCGGAATACAGAAAA | TGGTCCGGTCTTCTCTCAAC | >MK492 | CCTGCTCCGACTAATGCTTC | GTCGAATGTACCCAGCCAGT |
| >MK71 | GGATATTAGGACCGCAACCA | TTCGGAGCGGACAAGATACT | >MK493 | TCTCTTCTCCTCGCTTTTGC | CTTTCACAAACATCCCACCA |
| >MK72 | TCAAGGGCTGTACTACCTTGTT | CCGACGGTGCATAAATTTCAG | >MK494 | GCAGAGGAATTCAGGTCCA | CATTTGGCCAGAAGAAAAGC |
| >MK73 | T TCATGTCCAAGTCCAACAACA | AAACGCCTATGCAAATCAGG | >MK495 | AATTCACGCGTCTTTCAAGT | CGCTCCTCTATTCTCTTTGTG |
| >MK74 | CAGAAGGCACTCTGTTCTGATC | GTTGAGTAACCTCCACACATGA | >MK496 | TTTCTCGCCTCCTTCCTACA | CAGCTGGTTTCGAGAGAAACC |
| >MK75 | T TGGCTGTAGCTCAAGCAGAA | TGCTAGAGGAAGCAAAGCTGA | >MK497 | TTCCCTGGATTTAGCGATG | TGATGTTCCCACGTCTTCTG |
| >MK76 | CCACAAGCTCCTCCTACTCC | CGAGGCCGGTGTGTATAAAT | >MK498 | ACATGGCTCAGTGGGAAAAG | CACGTCAAGACAAGGTGGTG |

| | | | | | |
|--------|----------------------------|-------------------------------|--------|------------------------------|-------------------------------|
| >MK77 | ATGCCTCCCAAGCCTTATCT | TGAAGTCACACCCCACTTGT | >MK499 | AAGGGTTTACGCACTTGGA | CTGGTGAACCATCAACAAATG |
| >MK78 | TCAGAGCGAGACGATGACAC | TGGCATTGAATTGGAGTAGC | >MK500 | AGCTTTGTTTTCTAGCTCCAAC | CCATGAGTGTTTTGGGTGAT |
| >MK79 | CGGAGTCGATGAATTTTGCT | ATTTGGCTACCACGTTTTGC | >MK501 | GTCCAAATCCCATCTCCAAA | AGAAGGGCTGTTGCTGACAT |
| >MK80 | GCTTAACAGGGGTAGCACCA | GAGGACCAAATCCCATGCTA | >MK502 | AGCTTTGTTTTCTAGCTCCAAC | TTCAGCATCTGGATCTGGTG |
| >MK81 | TCCAAATTTCCCTTGAATGC | GGAACGAACCTTGCACCTAA | >MK503 | TGCTTGAATCGTCCTCTCT | CGGACAAGGGTTTCCATTTA |
| >MK82 | AAAAGAGGGAGGCAGCTACA | CTTCTTTGGAGCGGAAACTG | >MK504 | CATGGAATTCGGACTCTGGT | TGAAACACCTTGTGACGGACT |
| >MK83 | CCTTCTCATGATCACCTGGA | ATGCATCACGTGCTATGCTC | >MK505 | GATCAATCGAATTGGGGAGA | TCGGATTTCGAGAAAATCTGC |
| >MK84 | ATGCATCACGTGCTATGCTC | CCTTCTCATGATCACCTGGA | >MK506 | TGGATATTGGAATGGGATGC | TCGATTCTCACTATACCCTTCTC AC |
| >MK85 | CTTTCCCAAATTTACTCCCTTC | GGCAAAGTGTGGCTTCATCT | >MK507 | CGCTTGAGGTTTCGTCCTATC | GGCAGACACAAGGTTTTGGT |
| >MK86 | GAAGGGAAAACCAGATGGTAA A | CAGCAAGTTCTGCACCTCAA | >MK508 | GAGCCGGAAGAAGAAATGAA | AAACGGCGACATAGGATTTG |
| >MK87 | GAGTACTTGGGTCCCTGTCG | CTTTTAGGGTTTGGCCATT | >MK509 | CAGGTGACTTTTCTTTTCTCTGT G | GTGTGCATTCCCCTCAAAAT |
| >MK88 | GGAACAACAAAATGGCTGTG | TCTAGTTAAAGGTCCTAAGTTGT GC | >MK510 | TATGGGAATTGGCACAACAA | TGTTGCGGACCAGAGAAAAG |
| >MK89 | GTTGCTTGCAATGATGAAGG | CCTCCTCCTCCTCGGTAAAC | >MK511 | ATCATCTTCTCCGCCATCAG | CCACCATTGTTGTTGTTGTTG |
| >MK90 | ATGAAACCCACACGGAGTA | CCCCTTTCCATTAATTCCTCTC | >MK512 | CGACAACAAGGACAACAACA | TTTCCTGGTGGTGAAGATTTG |
| >MK91 | GCAGCACCAAATCATGTTCA | ACCACCACCACCAACGTTAT | >MK513 | GTGAACTTGGTGGAGGAGGA | GCCACTGAACCATCCAAATC |
| >MK92 | CCATCAATACAAACGCTGCT | ATTACCGCTGCAATTTTCGT | >MK514 | TACAGTGCGCATCCTGATCT | TGGAGTGAGAGTGAGAGTGAA AG |
| >MK93 | GAATAGCCTTCATCCCAGGAG | GGGGTAGCTAGGCACGTA | >MK515 | AGGGTTTCCACCAACAACA | GATGCCACGAGCCTTGAT |
| >MK94 | ACGTGCCTAGCTACCCCTAT | TCCACCATAGCAACATGCAG | >MK516 | GGAAATGAAAGGCCAAGTCA | GAGCCGGACCCAAGTATTTT |
| >MK95 | GGCTGCAACTACAAATCCAA | GGACGCGGTCATAGAAGAAG | >MK517 | TGTTTTGCATGAGAGAGAGAGA | GGATGATGAAGATCAGATGAAG G |
| >MK96 | GATGCGTTTAAATGGCCAAGT | GGCTGATTTCATAACACCCAAA | >MK518 | CATCCACA AAAAGCAGCAGAA | TCATTGTTGGAGGAATGCAG |
| >MK97 | CCTGTATTTGATGAGGCAGTGA | TGTTGTTGGCAGAATGGAAA | >MK519 | CGCGTTTTCTCTTGTCTCT | AGGAAGCAGAAATGGGGTTT |
| >MK98 | AACTCCCACACAACCACCAT | ATGGCAACCCAATTGAAAGT | >MK520 | TGGTTCATCAAGGTATCTCCA | TGACTGCTTTGTGGATGAATG |
| >MK99 | TGACATGCCCTCTCTTCTT | CAGTTTGGGCCCTTACCATA | >MK521 | TGATGCTCGTAAGGTGTTTCG | CCCCAGGTCAAAATCACAAC |
| >MK100 | GCAGCAGCAATGACAACAAT | CATCGCATTGAAACAACATCA | >MK522 | CAGGTGTGAACACTGGAGGA | GGTGCGACTGTGGTTCTTTTC |
| >MK101 | TACCTTTCCAGACCCCACT | AGTTGTGGCTGGTCTTCAA | >MK523 | TCACATTGCAGCCAATCATC | TTTTGATCGTTGCCTTTTCC |
| >MK102 | CTTTGGTGGTTGCACTCTCA | GGGAAAATCAAACAGGTAGCA | >MK524 | GCTTCCACTAAGGAGGCAAGT | AGCAACCCACCCAAAATGTA |

| | | | | | |
|--------|------------------------------|-------------------------|--------|--------------------------------|------------------------------|
| >MK103 | ACCTTATGGCATGAACCAG | CTATCCGCTGGACACCATT | >MK525 | CTCCATTACAGGGGCTATCA | CTGCAGAATTGACCCAGAT |
| >MK104 | AAATGGTGTCCAGCGGATAG | TTCCTACCACCAAACGAAGG | >MK526 | ATTGTTGCTGATGCTGCTGT | GTCAGTCGAGTCATTTTCATCAT C |
| >MK105 | CACCATGGCTGTCTTTCTCA | ATGTTGCTTGGTGTGTGCTT | >MK527 | TGGTGGATACTTTGCTTCCAA | GCAGGGGCCATATTATACCTT |
| >MK106 | TTGGAGCCCATTTATGTTCA | CTTCCCATAACTTGCCAGGA | >MK528 | CTTCCCAGCTTCCCTTTTCT | GGATCTAGGGCACAGATGGA |
| >MK107 | GCATTCCTGCATCAGTTTCTT | AGGTGAGTTAGCTGGGAGCA | >MK529 | TGATGATGCCTCTTGAATG | TTGGATGAGTTGTGGCAGAG |
| >MK108 | TGGATCATGGGTTAATTGGA | CAAAATCCCTCCACTTGTTG | >MK530 | TTTTCCAAGCTGCCTTATCG | ACACCATGAGTTTGGGTCGT |
| >MK109 | TCAGCTTTTACTACTCCCTTTTC C | CAGGGTCATCAAATCCAAGC | >MK531 | CCACACGCCTCTCTTAGCTC | CCGACTGTGAACAGCAACTC |
| >MK110 | GCACCGTTTGCTTCCCTTTT | ACGATACATGCCCATTTCTCC | >MK532 | AGCAACCTGGGCTCTTTTCT | AGCCCTAACTTCCCACGAAT |
| >MK111 | CACCCACACTATAACCCAGCA | TGTCCTTCACAGCTCACAAACA | >MK533 | GCGGTGCGTGTTCTTATCC | GTGAAATGTATCGGAGATTAG GAG |
| >MK112 | TGTTGTGAGCTGTGAAGGACA | CTCTCCGGCATCATCGTTAT | >MK534 | GCCGGGAAATTCTCTAAAT | GATTCCCAAAACAGCACCAT |
| >MK113 | CCCTCCCATTTCTCACTCAC | AAGAATTGCACCACCACACA | >MK535 | GCTATCTCAGCTGGCTCCAC | GCGCTGCTGTAATGAAACAC |
| >MK114 | GCAAGGCTACACTACTTCTGAC AC | AAGAGGCACACCGACAAAGT | >MK536 | TTCCATGTTGAGGTTTTGCT | CCCATTGCCTTGAACCTTGAT |
| >MK115 | TGCCTTTGGCTAGTGATGTC | ACCTTTGCTCCACCTCTTCA | >MK537 | CCACTTGTGATGCCAACTCTT | TTAAATTGCGCGACCTTCAT |
| >MK116 | TGCAGGTGCTTCACTTCATC | TTGGCTTCCATGTCAACAGA | >MK538 | CAGGACTGCCACCAATTCTT | CAGTGAAAATTTGCCGGTCT |
| >MK117 | TTCTTTCCCTTTCCCTCGTT | GGGTTTTGAGGGGTTTCTTC | >MK539 | CCCTCTATTCTTTCTTGCTCA | GGCTTTGGCTTTGGATTTTC |
| >MK118 | GATGGCAACAAACAAACCAA | GGAATCCAAGGCCCTAAAAG | >MK540 | GCTTATCGCGCTCTGTCTCT | ATCCAGAACTCGCCACTTG |
| >MK119 | ATTGCACGATGTCACAGCTC | GTAAAGCCAACCCAGTTGA | >MK541 | AATCCAGCAGCCACCATTAG | GCTGCAATGTGTGTTGTTGTT |
| >MK120 | TTTAGCCCCTGAAGTCGATG | ACCGGCACTGTCTCCAACTA | >MK542 | GCAGCAATCGTTACAGTCTCC | TGCTGCAATGTGTGTTGTTG |
| >MK121 | GTCGCTGAAATGGATGGTTT | TCAAGTGATGGAGGCACGTA | >MK543 | GATGGGGATGGGAGGAAA | GGTTCGAGTTCATTTCCAGA |
| >MK122 | CCCAGATCAGGAAAACCTCA | ACTCAGAACCGCTTGCTCTCC | >MK544 | ATGAACTCGAACCCGAACAG | GCTTGTCTATCTGGATGTGGA |
| >MK123 | TTCTCATTTTCTCCTCTCCTC | CTTGGGTCCAACCTTGAAAA | >MK545 | CCCATGCAAATGAAGGTTTT | ACCTGTCACCTGGTTTCGTTTC |
| >MK124 | GGCCTACGAGTTGCTCTCTG | CCGATTTACATAGCCCTCCA | >MK546 | TGGAACAATAACCCCTGAAGCA | CCCATTCAAAAAGGGAGCAAT |
| >MK125 | GCGATCTCTGCCCTATTTGT | CATCAAGACACCACTACTGCCTA | >MK547 | TGGAGTAGAGAGCATTAAAGAAG CTG | ATACTTCCATCGCGGACAAT |
| >MK126 | TTTGCAGAAAGGGGAAAAGA | CAGAGTCTTGAGGCGTGAAA | >MK548 | CAATTGGCGCTGTTATCAAAA | CCTCCATTGCAAATGCATAA |
| >MK127 | AAATGCAAAGGAATGCCAAC | TACTTCTGGGCATGGGTAGC | >MK549 | GCATGCACCCTAGGTTTTGA | TGGCGAGGTTAAAGTGTGTG |
| >MK128 | TGCTGCAAAATGGTCTGTTA | CCCAACAGCCTAAACTATTTTCC | >MK550 | TCTGCATCTGGTGAAGCAAG | TGGATTCCCAGGTTGATTTTC |

| | | | | | |
|--------|----------------------------|------------------------|--------|----------------------------|-------------------------|
| >MK129 | TGAATGTAAAAGGGGAAGAAG G | AGTGTCTGGGGAAAATGTGC | >MK551 | TCTGCATCTGGTGAAGCAAG | TTCAGCTTCTGGGTCTGGAT |
| >MK130 | CTGGGATCCATGTGCTTCTT | ACCCTGCAGTTCATTTTCGTC | >MK552 | CAAATGTGGGGCTCTAGGTG | TAGCCAAAACATGCCTCATCA |
| >MK131 | AGCTCAATCGAGGTGAATGG | AATGGGATGTCCCTGAGATG | >MK553 | TGGTGTTGGGCACTACAAAC | GCAACGACTCTTGGGGATTA |
| >MK132 | GGCTGACCTTCTCTGCATCT | CCGGAAAGCAAAGAGGAAAT | >MK554 | GTTGCTTTGCTGGCATGATA | CCCAATCCCTAAAACTGAACC |
| >MK133 | GCCAACACCAAAGGACAGTT | GGGAGGGAAACAACAAACAA | >MK555 | AGAGGCGGAGTCAGGATTTT | CCTGTGACCAGTTTTGAGGA |
| >MK134 | CTGTAATGGCTGGGCATTTT | AATTCGATCCGTTGACCTTG | >MK556 | TGGCATGTTTTAATGGCAAG | CGTCGAAAACCATAACATGA |
| >MK135 | CTCCCCAAACCCCATTTGTAT | ATGGGTGGCACACTTGTTTT | >MK557 | GGCAAGAAAACAAAGGGAAC | TGGAGCCATAATTACCTTCTTG |
| >MK136 | GGGGTGAAAGAAAAGCCAGT | CATTCGTGTTTTGGTTGCTG | >MK558 | AAGCTTTTCAAGTCACGGAAG | TTTCCCTTGCTTCTTCTCCA |
| >MK137 | GCGGTGTTACGACTTCGTTT | CCTCCATAACCCCTTTCTCC | >MK559 | GCAGACAGCAGCACAAAATC | TGGTCATTTCCATAGCCTCA |
| >MK138 | GCCGGATATTAGCGGTGTTA | GCTGCCAGGCTTGATATCTTC | >MK560 | GGGATTGTCATGAGGCATTT | TGTGCAAGAGAAAGCCAAGA |
| >MK139 | TGAAAGACGTGGAGCTGTTG | CAATGTGCTTGGCAGAAATG | >MK561 | AACACCAAACCAGTCCCAAG | TTTGCAGAATGGACCAGGAT |
| >MK140 | AAGTACAAGCTGCCCAATTAGC | AAATGCCCTAAATCTCCAA | >MK562 | TGGAAACAAGAGCTTCCCTTT | TTCTGGGCAGTGTGTTGTA |
| >MK141 | TCGTCGTCATCGTCTTCATC | CACGGGATGGTATGTGAATG | >MK563 | AGGTGGAGGAGTTTGTGGTG | AAGCCTCTTGTGGACAGGAAT |
| >MK142 | CATTGATGGAGCTCGCTTTT | CATCTTCTTCGTCTTCCCTCCA | >MK564 | ACCCATTAACCCCATTTTCA | GCAATACCCAATTCTCCAAA |
| >MK143 | CGGTTTTGGAGGAGAAATCA | TCTGTCCACGGAACCTCTCT | >MK565 | TTTCCCTTTGCTGTTGCTCT | CCCCGGTATCGGATTACACT |
| >MK144 | AATTCGGATGTTTCGCAATC | AGCAACCACTTCTCCAGCAT | >MK566 | TGAAGGTGACAGGATGCTTG | TTGCTTCTGCAACTACACTTGA |
| >MK145 | GGATTTCACTTCCCACTTGC | TGAAGGGTGTGCTTAGGG | >MK567 | TCTAGCCCTGAGATTTGGACA | CTTCTCAACTTTCTTGTTCG |
| >MK146 | GCTGATAAATCACGGCTTCC | CAGGGATCTGCAGGGACTAA | >MK568 | TCAACCGTGCCCTAATTCTC | ACGGCCTTTTCTTACACCT |
| >MK147 | CCAACTTGCTCTTCCCTCTT | GAACCGTAGCGAGTGAGACC | >MK569 | ATCGGAGATTCTGCATGCTT | AACGTCATCGTTTTCCCTA |
| >MK148 | AAAGCGAAGGAACAATCGAA | CCTCTCCCCATTGAAACTGA | >MK570 | TTCGTTTTGGTGCTCTCACA | TAGGGTTTTCCAGCCATCAAG |
| >MK149 | TTCCCGTCCAATTTAGCATC | TCCTCTCCCCATTGAAACAG | >MK571 | TCATTTACCTCGCTCATGGA | TCCTTGGCTCACTAATTCTTCTG |
| >MK150 | TCTCCCATTAAACCAAACA | ATCGTCACCTCCTCCTCCTT | >MK572 | TCTTGAACCCACCAACAAAA | GATGGTGCTGGGAGAGTAGC |
| >MK151 | GTGGATCACATTTCGCATCAC | GGCAAATAGCAATCACACGA | >MK573 | CGTCAACGAGAAAGAAGAAAG A | TGTACATGGCCCTTTGGAAT |
| >MK152 | TCCACAACTACCCTCCAATTCT | TCCAGATGCTTGAGTTGCAC | >MK574 | TTCATGCCAACCAGTCTCTCT | TTGGGAAGCGATCCATTATT |
| >MK153 | ACGGTGTAGCACAGCATGTC | TGAGTCAGAAAACACGCCTTA | >MK575 | GTGAATGTGGTGGCATTGTT | TCAAGCTCAAATGGCTTCCT |
| >MK154 | CAACTTCATGTTTCTCACTCTCC | ACCAGCAGCAGCAGAAGAAG | >MK576 | GCAAACACAACATGGTGGAG | GCCAATTTCAATTTCTGTCTG |
| >MK155 | AAAAGGCAAGGGTAGTGTGC | CAAGGCCGCAGTGTATGT | >MK577 | CCGGCAATAAGCATGAAACT | TGTTTACGTGCTGGACAACC |

| | | | | | |
|--------|-------------------------------|------------------------|--------|-------------------------------|------------------------|
| >MK156 | TGGGAAATAGGTGGTCACAAA | TCCTCCCATCGAAGAATCAC | >MK578 | TTTTACCACGTCCATCAAG | TCGCCTTTAAACTCCTCTTCC |
| >MK157 | GTGGTGCTAATGACTATGATTC CA | TGATTCCCAAAACCAACCAT | >MK579 | GAGTGCCGGATTTAGTGGA | CCAGTCCTTTGGTATCCATGTT |
| >MK158 | TTTCAACCCACACATACCC | TGAAGTTGTGGAGAATTGAGGA | >MK580 | CCTTTTCAGTTGCTGCTTCC | TGTTCTCCTCCTGCGTTCTT |
| >MK159 | CTGTTGATGTGCCAGTTAATCC | AGTGTTCCTGCACTTTCTG | >MK581 | AAGAACGCAGGAGGAGAACA | TCCAAGCCCTAAACCACTTG |
| >MK160 | ATCTCAGACGAATCGCCATC | AGGATTCCCAAAATGCCTCT | >MK582 | TCCCTTGCAATATACACACAGC | CACCAATCCTAAGCTCCACAA |
| >MK161 | TCATCAACTTCAAGGGGACA | CTTGATCCCATCCATCCTTG | >MK583 | GGGTTTACACATTTACAAAGTTA CC | GCAGAGGCCATAACTGAGGA |
| >MK162 | TCTTTCATGGGCATTCCTCT | GACCGTTTATTTGGCTCTCG | >MK584 | CGTCCTTTACCACCGCTTTA | TCGGCATTGCATGAAAGATA |
| >MK163 | GGGAGGAACCTTGAGTGGTAGG | TGGTTTGATTTGCATGGTTG | >MK585 | TCGTCTCATGGCTTCTTGTTG | GGAGGAGAAAGGTTGGTGGT |
| >MK164 | TTTTAATCGAGGTGCCAAGG | TGGTTTGATTTGCATGGTTG | >MK586 | CCCTTCCCTTGGGTTATGAG | TCGGAACAGGGTAAAAGGAA |
| >MK165 | GCAGCACCAGAAGAGTGTGA | GGCGGAGTTATAATGGCTGA | >MK587 | ATGTCTCCACATTGCAACA | CCTTCGCCTTTTCTTCTCCT |
| >MK166 | CTGACCCTGACAACCCTGAT | TTTTGCCAGCATTTCAAGTTG | >MK588 | AATTGCAACTTTGCCACTCC | CCACTAGTCCCACTCGCTTC |
| >MK167 | CCAACAAAGCCAAAACATGA | GGTGTGAATAAAGGGATCCAA | >MK589 | GGGTGCATCCTTCCTAGAGT | CCACTAGTCCCACTCGCTTC |
| >MK168 | ATTTGGGCATCAACACCTTC | TGGGAACCTACTTAGCCATGC | >MK590 | TGCTATTGTATGTTACGCAGT | TGATCAGCGGCTTCTAATTG |
| >MK169 | TGGTTTTCTTTAGCTAGCATGG | GCCTCACAGACAACCCAAAA | >MK591 | CCCTTCCTAAATCCTGCAAA | TGCTGGTAGCTTCCTCGAAT |
| >MK170 | TTTTGTGAACGCAGAAAGG | CGACGCATCACATTCATCTC | >MK592 | ATTTTCTGGGGTTGTTGCTG | TTGACGAAAACAATCGCCTA |
| >MK171 | GGATGTGAATGAGGGTTTTGA | GGGATCTCCCTCGTCTATCC | >MK593 | AAAGGCACAAACAGGAATGG | TGTCAGTCTGCTGGCTATGG |
| >MK172 | TCCACACTCATGTTTGAAGTAG AAC | GCACCTTCATCTCCAATTCC | >MK594 | CATCTGGGAAATCAGACTTAGC TT | GCTAGAACCAAAGGCATCCA |
| >MK173 | TCAACCTCGAAAATCGTCAT | GAATTTCCGGACACCCATTTG | >MK595 | CCATTCCACAATTCTCGTCA | CATCGTCCCTGTTGTTCTCA |
| >MK174 | TGCCCTTAAAGAAGGAATGG | GCCTCTGAAGCTACCACCAC | >MK596 | AACAAGAGCCTCGGAGTTTG | TCCTGAACGACGACTCATTTC |
| >MK175 | CTCTTCAATTGGCCTCCTTG | ACGCCAACTCCGTATTGAAA | >MK597 | TTGTCCGAAACAGAAGCACA | GTACATTTGTGCGCGACATT |
| >MK176 | AAGGATTTTGCCTTGGAAT | TCCTGCCATCCTTTGGTAAG | >MK598 | AAAAGGGCACTCCTCCAAAG | CCTCCATTGAGTTCCATGA |
| >MK177 | CACCCTTCACTTGATGGATTT | CCGGATGGTGTGAAAAGTGT | >MK599 | ATACGATGATGCGAAACGTG | GTATGCCACCAGGAAAAGGA |
| >MK178 | CCATCCGGTCATTCAAATCT | CACCATGCTAATGCTCCTCA | >MK600 | TGCTGGAAGTTGCACTCAAT | TGGTCGTCATCATCATCAAA |
| >MK179 | TGACCATTTACATTTGCTAGG | CCCAAGTTGCTAAAAGCAG | >MK601 | CAACAGGAGGATTTGCCAGT | ATCATCAAGTCCGTGCACAA |
| >MK180 | CCTAGCAATGGAAGAAGAACG | AATGTCCCTAGCCAAACACG | >MK602 | CCCCATATGATGAAAACCAA | GGCCTTGTCTCTCTGGTGA |
| >MK181 | TATGGGACGGAGAAGGAGAA | ATTCCCAATCCCAACAATCA | >MK603 | GGAGGACATGTTTGGTCCTG | TTGTTCCCTGACCCTTCATC |

| | | | | | |
|--------|-------------------------------|------------------------|--------|------------------------------|-----------------------------|
| >MK182 | CTTCTAATGTAGCTATCGCAA CTG | AAAGGCTGTTGGTTTCCTTG | >MK604 | ACCCCGCGTTTTCTTTTC | CGGAATACGGATTGGGTCTA |
| >MK183 | CCCTTTGAACCAATCACAGC | AACCAAACCACCCTTCAAGAT | >MK605 | TCAGCCATGGACTTAACCAA | GGATCTGAGTTGGGGTTTCA |
| >MK184 | TGCATGCTCTGCTTGTATTG | GCCCCTGGAGATAAAAGAGC | >MK606 | CGTTCGTATCTCTCTTTCTCTCT G | TGATTTGTGGGTTTGTGAGG |
| >MK185 | GCTATGGTGTCTGTTTGTCTA | CCTCGAAGGTGGCAACATAA | >MK607 | GCCAAAGTTGAAGGGCTTTT | CGATGAAGTACCCAAACTCAA |
| >MK186 | CCCAATCAAGAACCTCACCT | GAATGTGGAAGCAACATGGA | >MK608 | TTGCTTGCTAGCTCATACGC | CTCAGGCTCCAATCCTTCAC |
| >MK187 | TGTGTCCTGTGGCAACATTT | TCCAAACTCACCAACCAATTC | >MK609 | GCTCTCGTTGCCTATTTCCA | CCGTCCATTTTCCACTTCAC |
| >MK188 | GACAAAACCTTGAGCCATAGTCC TT | AAATGGTTGATGGGGAGTCA | >MK610 | CAAGCCTCAAACAACCTCTCC | CGAGAAGGCCATAGAGCAAC |
| >MK189 | TCAAACAACCCCAATTCACA | AAATGGTTGATGGGGAGTCA | >MK611 | TCTCTCTCACAAAAGCCAAA | ATGGCTTCCACTTTCCACAA |
| >MK190 | TCTTCTCCCACTGCTCT | ACCCAGCCTAAAAGCCTCAT | >MK612 | TCTTTGACTCGTGGATCG | CAAGCGATAATTCTACGCCTCT A |
| >MK191 | CAAACCTCCAAAATCAGCACTC | TTTTGCAGCTCCATAAACC | >MK613 | CCGAAGGTCTTCAGCATCTC | GGGGCATCTCTTTTAGTGGA |
| >MK192 | TTTGAACAAGAACCGGCATT | CAAAGACACGAATCACAAGCA | >MK614 | ATGCTCTCGGTGAAAGGAAA | CCTCAACGCATCTTCAAACA |
| >MK193 | GGTTGGCGTCAATTCAAATC | TCCACGTCTCTCTAGCACCA | >MK615 | AAGCTCCAATTCATCGGCTA | GTCCACTGGGAACAAGTCT |
| >MK194 | AAGCACTTTCTACCAATGAACT CC | CGGTGGAATATTGTTGCTCT | >MK616 | TCAACCGCCTTTGAAAAGAT | GGTCGGGTACAGATGAAGGA |
| >MK195 | TTCTGGTTTGACCCACAAAG | TTGAAGGCCCTGAATCTAA | >MK617 | TTTGTGAATGGGGAATTTG | ATTGTCCAACCAAAGCGGTA |
| >MK196 | AGTCTCAGGGTCAGGGTCAG | TGGGACAACGAGTTGAATAGG | >MK618 | GCTGCTAAACTCAAGCAGAA | GCTCAGCTTTCAGAAGAAACCA |
| >MK197 | CCAGCAACTCTTAGGGGAAA | CATTATTTGTGCCTCCACCA | >MK619 | ATTGGACCCTGCAATCAGTC | ATTGTTATGTGCCGCGATTT |
| >MK198 | AATCCAATTGACGGTCGTG | GAACCCAGATTGAAGCAA | >MK620 | GGGGTTAAGGAGTGATCAATTT C | CTTGAACAAGCATCAAGTGGT |
| >MK199 | TTAGGGCTCGAAGAGGACAA | ATCAGCAACTAGGGGAGCAA | >MK621 | CGTGGCATGCTCTAACAAAA | ACGGGATTACAATGGGTTGA |
| >MK200 | AGGAAAATGGCCTGTGGATT | TTCGTCAAAGCGAATACTGG | >MK622 | TGGGGTTGGTTTATTTTCGT | TTGATGACTCGCTGGAATTG |
| >MK201 | CAGCTCGAGAAATTGGTATTTG | TGGCTAGTCCTCAGAACACCTT | >MK623 | AAATTGGGCATATTCAGCA | TGTCCCCAAAATTGGTTGAT |
| >MK202 | AAGTATTTGGCGTCCCAGAG | TGGCTAGTCCTCAGAACACCT | >MK624 | TGGGTTGTGAGAGGAATAAGC | TGCAGATTTCTCACCTGCAT |
| >MK203 | CCATCAGGCTTTTGAGGAAA | GTCATCCTGCCCTATGCTA | >MK625 | CGCGCTATTACAAACAGGTG | GCAAGTTGTTGTTGGTGGTG |
| >MK204 | GCAAGCATTTTCGATAGTGTGA | ATCCTGCCACCTTAAGCAGA | >MK626 | TGACATCCACCAGCTTCTTG | CATTCACTCCACCCTGTGC |
| >MK205 | TCGAGGTCCATACTCGTGAA | ATCCTGCCACCTTAAGCAGA | >MK627 | CATTTTCCCAACAGCGAAAT | TTCCTGAACCACCAACATGA |
| >MK206 | GTCTCCCACTCCCTCCCTCT | CTTCCCCATTGATCCCTACA | >MK628 | CGTTTAGCCGTCCACTCTCT | TGTTGTTGCTGTGGTTGGTT |
| >MK207 | AAGATTTGCCGTTGTTCTG | TGAAAGCAAATTGTACCACAGG | >MK629 | CGTTTAGCCGTCCACTCTCT | GAGGCTTCCCAATGTCAAAA |

| | | | | | |
|--------|-------------------------------|-----------------------|--------|------------------------------|------------------------------|
| >MK208 | ATGACGACGTTTCGGATCA | CTCGTCGATCCTCGGTAAC | >MK630 | GGAGCCATCGACAGCAGTAT | CCTATCATACGCACGAGCAG |
| >MK209 | TGTCCCCAGTAGAGAAGAGAG G | TCAATTCGAACACCTCCACA | >MK631 | TCAGGATTCCTCAGCAGAT | TAATGCGGAATGTTTTGTGC |
| >MK210 | TCGGATTCGTGCATAAAGGT | CAGCAATGCAATATCCCACA | >MK632 | GGAGCCATCGACAGCAGTA | TTGCTTTCCACAATCCCATA |
| >MK211 | GGAAAGGTCAGCGTTTGGTTA | AGATTTTCGATGGTGGAGGTG | >MK633 | TGCACTTGTTTTAGACCCTTCC | TGTCACTCTCAATTGCCAACA |
| >MK212 | CCCCTCTCATCACTCTTTCC | AGCTTCTTGTACGGCGGTTT | >MK634 | GTGGCCTACGAAGGTGAATG | CGGTCTTTCTGTTGCATCCT |
| >MK213 | TGCCTGCTTCAGCTACACCT | CACCACCACCATTACCTTCA | >MK635 | TGGGTTTGGATTAAGGCATAG | AAACAGTGCCGATGAACTCC |
| >MK214 | AGTCCACAAGGAAGCTGCAA | GCTGCCATATTGCTTCGATT | >MK636 | AACGTGGACTACCCTGCATC | AAGAGGAGGCCAAAGAAAGG |
| >MK215 | AGGTGCAACTAGGCCAGCTA | CAACCCCAAAATTCCTTCA | >MK637 | TCTTCAATCATGGGTGATCCT | AAGGAGGAGAAGGTGGTGGT |
| >MK216 | TCTGGATTCGGGTCGTGTAT | TCTTCTCCGGCAACTCATCT | >MK638 | ATTTCTCACAAACGGCCAAAG | TGCTATTTCCCAAGCTCTTC |
| >MK217 | TGGAGAGGCTGGAGAAAAGA | TTCCCCTAAACACCTCCTC | >MK639 | CGATTTGGGAATGTTGCTTT | CGTACAATGGATGATTCTCTTC AC |
| >MK218 | GCCTTAGCCAAAGTCCAAGA | CTGGATGCAATGGGAAGAGT | >MK640 | AGAAGAGGCTCATTGCTGCT | CCCCTTGTGAAGATGCTGAT |
| >MK219 | AGGTGAGGTTTGAAGGAGGA | TCCATGTTCAATTGACCCAAA | >MK641 | TCTGTCACTTTTCCCATGC | TGAGTTCAAGAGCAGCCAAG |
| >MK220 | GTCCATTTAATCTTGTCTGAGA TCC | CTGAGCCACCATAGCCATTT | >MK642 | TGCTCGTAAGCAGGGTCTTT | TTCATCACACGGAAAAGCTG |
| >MK221 | GATGGTAATGCTGCTGCTGA | CCACGGTTTTCAAAAGACGAT | >MK643 | AGAACGAAGCCGTTGTTGAT | GAAATCGTTGATCCGCTCTC |
| >MK222 | GTGGAAGTGGGAAGCAAAC | TGCCACCAAGCAAAATATCA | >MK644 | GAATCAGTGCAACGAGCAAA | TGACGATGCAGAAGCGATAG |
| >MK223 | TTGGCGAACATTCTGGAAAT | CGCACGGAACATTAGTGAGA | >MK645 | CCATTGACAGCCCATTATCC | TGATGTGAAAAGAGTTGATGAGG |
| >MK224 | CCTGAAACGGACGAGGTTTA | TCAATGTCTTGGCAAACGAA | >MK646 | CCTTATCCCTGCCTCACAAA | TTTCCATCCATCCCAGAAAA |
| >MK225 | GCCCTCCTTCTGTTCTTTT | GTTGAGCGGGTGTAAAGGAG | >MK647 | CTCGGAACTGTTAGCACACG | TGTATCCAGGGACAAGGGTTT |
| >MK226 | AGGAAAATGCAACGATGACC | CCCGTCTCCACCTTCACTAA | >MK648 | TGTATCCAGGGACAAGGGTTT | CTCGGAACTGTTAGCACACG |
| >MK227 | TGTGGCTTCTCTTGAAATCG | TGAAACCACCTCTCATGGAA | >MK649 | CGAAAAGCTCCTCAAAGCAAC | CGGAGATGAAGGGAAGAGTG |
| >MK228 | TCAATCCCAACAGCAACAAA | AAGCCAAAGTGGTAGCCTTG | >MK650 | GCACATGTTTTCTGAGATTGC | ACGTATATTTGCACGCATGG |
| >MK229 | AAGCACCTTGGCTCATGAAT | GCTGTTGCTGACATGTGGTT | >MK651 | AGAGAATGGTACTGGCACTGG | GCTTGTGTTTTGGGAGCATT |
| >MK230 | CATGAGATGGCCAAGTTTGA | GGCGACGCTCTCAACTACTC | >MK652 | TCAACAAGATTTGCCACAA | TCCACTCCTCATCATCACCA |
| >MK231 | TTCTCCCCAACTCTCCTTTTC | ACCAGCAGCAGCTCTCAGT | >MK653 | CAGCCGCTAAAAGGAGAGAA | CTCATTCGAACGCAGTGAAA |
| >MK232 | ATGTTCCGATTTTCCGTGAG | AGCAACCACCTCAACAGCTT | >MK654 | TCTTCTTAGACACAACCTTACC TG | AACTCCGTTGGGTTTTCTT |
| >MK233 | GGCCAACATCACAGAACAATC | TCCAAGACATGGGAAGAAGG | >MK655 | TCACGACAAGATAACGGATCA | CTGTGGCTGAAAATAAACCTA |

| | | | | | |
|--------|-------------------------------|-------------------------------|--------|------------------------------|-----------------------------|
| >MK234 | AGCCAGTGCCAGAGAAGAGA | TCAATCATCTCGCGTACCAA | >MK656 | GACGCAGAACTCATCTTGTTC | TTTCCTTTCATCAACCACCA |
| >MK235 | GAAACATCCCTGTTGGTGGT | ACTCCCCACTAAAGCCTCGT | >MK657 | TTTTCTCCATGGTGTGGTT | TCCACTTGCATGCTCTAATACC |
| >MK236 | GGGACACCTTGCTAGAGCTG | TGGTGGTGTCTATGAAAGTGATG | >MK658 | TCGTGAAAGAAAGGATGCAA | AATATCAGAGCCACCTCCACT |
| >MK237 | CTCTCTTAGGCGATTGTATTTGT TG | AGAGGAAGCGAGTGAAATAC | >MK659 | AGACTACAATTCCGTGCTTGG | AATATCAGAGCCACCTCCACT |
| >MK238 | TTCAGACAAAGCAGCCAAAA | GCTTGCTTCATTAGATGCACTT | >MK660 | CTCCTCTCTTCCCCACTTCC | TGCCATTTCCGTCTCTGTAA |
| >MK239 | CATGGCAATGTCAGGATCAG | CTCGCACAAAATCATGAAA | >MK661 | AAACGTTCCGGATTTTCACA | GAAAGGGCCAGATTCCTCTC |
| >MK240 | ACTTGATCCGCTATGCTTGC | TTCCTCCTCCTCCATATCCA | >MK662 | TCTTTGCTGTTCCGCACTAA | ATGGGGCCCTAGAAGAAGAA |
| >MK241 | GGGACGCCTAAAGCAATTTT | GCTTGCTTCCCGAAAAGTAGT | >MK663 | GCCTGCAAAATCTCCATCAAC | CCACTTCCATGACTGTTCAAGA T |
| >MK242 | TTTGCAACCTCATCAACTGG | TGGCTGCTCTTCTTGGATCT | >MK664 | TGGGCTTCAACTTTATCATGG | TGGCTAAAACCTCCCTCCATATCT |
| >MK243 | ATCCAAGAAGAGCAGCCAGA | ACTTGCAAGGGTGAAGAAGA | >MK665 | CGGAGCTGGAGGAGGAAATA | ATGAGGCTGAGCTCGTTGTT |
| >MK244 | TGTAGGAGGTGGCAAGTGTG | TACCTTCACGTGCGTATCCA | >MK666 | GGTTTCGAGTTATGGCAGCA | CCCCACACCAACTGAGAAAT |
| >MK245 | GAATTATCTCCGCCCTGATG | TTGCCCTCCTAGACTTCACC | >MK667 | TCCTGCTGCTAAAAGACGTG | GGGTCTTCTCTTGGCCTTTT |
| >MK246 | AGCTCTTCCCCATTTTACCC | TCTTTGAGTTCCTCACACG | >MK668 | GAGCATGCAACACTCTTTTCC | CAAAAGAAACCGACCTTTCG |
| >MK247 | CATGAGTCAATCCCAACAACA | AAATCGAAAACCAGGCATGA | >MK669 | CCAGATTCCGGTAGTTTCTCC | AATTCCACCAGTTTCTCCA |
| >MK248 | ACGCCTACAAAAGGCAGAAA | TCGATTTGTTGTTGGGAAGA | >MK670 | CATCTTCCCTAATTATCCTTCA GC | TGGGAAATTCATTGGTTGGT |
| >MK249 | AACAGATGGAAGCTGCAAAGA | CCCTAGCTAATCAGAACTAACCA AA | >MK671 | ATCCTCCTCCTCCTCCTCCT | AAGTCATTCAAATCTGGCCTGT |
| >MK250 | ACCACCAATTGCAAGTCTCC | ATTAAGTACGCGCCAACG | >MK672 | CGGAAACAGTCGTAGGGGTA | TTTGCATCAAGCAGAAGCAG |
| >MK251 | TCAAAGACGGTTTTAGGGAGA A | GCGATGTTCTCAGCTCAAT | >MK673 | CCAAGTGGAGGAGGAGTGAA | AAATCGCATCTTCCGTTGTT |
| >MK252 | TGTCAAAGTGGCTCACCTG | TCCTCCTCCAATCTTGTCCA | >MK674 | CCCCTTCTCTGTCTTACCT | CAACAACAGTGTCCCAAAGAA |
| >MK253 | TTTTCCCCATCTCCGTCATA | CCGTTGTCTTGCCTTGTGTA | >MK675 | TGAAATATGCTGCTCCGAAA | CTTCACTTCGGCTTCATTCC |
| >MK254 | GCATGTATGATGGTGGTGG | AACCTTGTCAAATCCCATGC | >MK676 | TCCCTGCAGAACAAGAACCT | GAATTTCTGCTGCCATTTCC |
| >MK255 | TCCAACGTTTCAATCCCAGT | TTTTCATCCAATAGCCTCCA | >MK677 | TGGTGGACACAGAGCAAGTC | AAATCGAGCACTCGTGAACC |
| >MK256 | TTTCATGGAGCGCAGAGATA | GCCATACACTCTTCCGGTGT | >MK678 | CCAGGAAGGACAGCAAATGA | CGTTTGAGAAGTTCCGAGGT |
| >MK257 | CTCTACCCTCCCCAACTCC | CATCTTCCGCAACCAGTTTT | >MK679 | CCAATTTGCCACGCTTTAAT | CTGGGGTTTTACAGCAAAGG |
| >MK258 | GCCCTCTAAAGCTTCCATT | GAAGCCACGGAAGAGAGAAA | >MK680 | CTCATTCCCACCAATTCTCC | AGCTGATTGCCGAAGCTAGT |
| >MK259 | TACATGCACAGCAGGAGGAG | TGTCCCACAGAACACAGTCAG | >MK681 | GGTGGAACCTTGTGGAGAAA | CCTCCTCTGTTTCTTCGTCTC |

| | | | | | |
|--------|-------------------------------|-------------------------------|--------|-----------------------------|-------------------------------|
| >MK260 | GGGAATTGCCATAATCTCCA | CAGAGAGGGAGGAGGGAAAGT | >MK682 | TAAACCCGTTGTTGATGTGG | CACCATATTGGTCAGGGACA |
| >MK261 | CATGGTTGGGGACTTGACTT | CCCTAGATTGAAACATTGACCA | >MK683 | TGGAAGAGGAAGACGAAGAAG | TTGCCTCCAACAGTCCTTTC |
| >MK262 | CTTCTCAACTCATCTCATCTTTC TC | GTCGGAGGTCCATGTTCATCT | >MK684 | TACTCTCCGTCCACCGTCTC | CCCGACCTTGAACTACACCT |
| >MK263 | GATGGACGGTGCAAATTTCT | ATGTCTCTCTCCCCAAAAGC | >MK685 | GCTGACTTGGAAATCACAGC | TTTGGTCGTTGATCGATTGT |
| >MK264 | ATTGGGCATTTCGTCAAGTTC | CTCTCTCCGGTTTGTATGAGG | >MK686 | CGGGCAATCTAGAGCAGTATG | TGCTAACCAAGGAAACGGAAT |
| >MK265 | ACACTCTCCAGACCGCACTT | GGGTACAACCATTTCATCTTTC | >MK687 | GCCTCCCAAGCCTTACCTAT | TCGTTCCATGGCATAACAGAG |
| >MK266 | GCGCTGGGTTCTTTATGTTC | CCCGTGTTCAATTGTGAATG | >MK688 | TGCTAAGAGTAGTAACCCCAAC C | TCCGGCTCAAATAAAAATGCT |
| >MK267 | TCCATCTCTTGGCGCTTTAT | CGACGACGAACTCATTTTCC | >MK689 | AGGCGAAAACCAACTTTCCT | TCGTCGTCTTACCATTTTG |
| >MK268 | CCATCATCTTCGGCTCTCTC | TTCGAATTGTTTAGCGTCCA | >MK690 | CTTTGAATTTGAGGGGAAGG | TCCTGCAGTGTCTGTGGAAG |
| >MK269 | CTTGGTGGCCATATGAATGTT | ACCCTAATTGGAACCCCAAG | >MK691 | TCAAATGCAGCAGTGGTCTC | GAAAAGGTGGACAAGGGTGA |
| >MK270 | GTGTGCTCCCAATCAGAAG | GAAAGAAGAGAGGAAGAGTGAA ACA | >MK692 | GCATGCCCATCATGTTGTAG | GGGTGCTTCCATGAACTTGT |
| >MK271 | TAGCACCCCTTATGGTGGAC | CTGCAGTATCAAACCTTCACTCA | >MK693 | CCTTCCTTTCTCCACCCTTC | CTGTGATTGGCAAGCGAGTA |
| >MK272 | TTCTGCGGAAAAGGTAGCAT | CCCCTTGGTCATCCTGTAAA | >MK694 | CGGCTTTTCTTTCGTCATCA | CTTTCGCGTTCGTTTCACTT |
| >MK273 | ACACAAACTCCGTCCCACTC | GATTTCGGACAACGAACGAAG | >MK695 | ATGGATGCTTCGCCATCTTA | TGTGGCATGTCTTTGAAGGT |
| >MK274 | TGCACCATTTTGAATAGCTC | CCCCTGCCTACCTCTAATCC | >MK696 | GGGATGCCCACTCTTTGATA | CTTTTGTTCATCGCGCAGT |
| >MK275 | TTCGTGAGGCCTTATTCCAC | ATACGAGCTTGCCACTGCTT | >MK697 | CATGGTCCGATGAAGAAGAAA | TTTTCGATGACTCGGTTTCAC |
| >MK276 | CCTTCTTGGCACAAGCTCTTT | GGTTGTGGGGTTTTAGCAGA | >MK698 | TGAGAGCTCATGGTATTGGAGA | AAGCATCATCAGAGCCACAA |
| >MK277 | GAGGATGCAGGAGGAACAAA | AGAGGCGGAATTTGAGGACT | >MK699 | CCGCGAAGAAAATGAGGAT | TGGACCAATACCAACTTGTAAC C |
| >MK278 | AAATGGCTTCGGTGAAGATG | TTTCCTTGAATCCCCAACAA | >MK700 | CTCAAGCGCATTTTCCTCAT | AAGCATGGATGTTTCGTGTG |
| >MK279 | CCTAAACAGCACGCTGAGAA | CAAAAACGGCTCCAACCTTTA | >MK701 | CACTCAAGCAACTCCCCATT | CATAAGCCTGAGCCAACTCC |
| >MK280 | GGGCCCATCAGATTATAGCC | GATGAAAGCCCCTCACAAAA | >MK702 | CACTCAAGCAACTCCCCATT | ACTTGATGCCTGAAGCTCGT |
| >MK281 | GCCAACTCTTAAAAGGCAAC | TGACTGATGTTGGTGACAAATG | >MK703 | TGCCCAACCAAAATTTCTTA | TTTGGGAACTCAACTCAACT |
| >MK282 | GATTAAAGGCACGGTGAGGA | GGGCGATTTTGGGTATGAAT | >MK704 | TGTTTGCCACTGTGAATGGT | GCCACACATGCATTAGCAAC |
| >MK283 | CAACATGTGTCCCTCGAAGA | TATGCCTTGGGGAATACTGG | >MK705 | TGAGACACCTGAGGATGCAG | CTCTTTGAACTCGGCTATCA |
| >MK284 | TGATACAACCTTGCTGATCC | GTGCCTTTGGTTGTTGGAGT | >MK706 | GATGTTGCTGCTGGAATGAC | CATTGAACTACTACTCTCACCA CCA |
| >MK285 | TCCCTTGTTGTTTTGCGATT | TGGTTCAGTGTGATGTTTCCA | >MK707 | CTTGCAAATTCACAGGCAAA | GAGCACAATTCCCACCAGAC |

| | | | | | |
|--------|------------------------|------------------------|--------|------------------------------|------------------------------|
| >MK286 | CCCACCTATTTGGAAGACGA | TTTCGGATTGAACACTGCAT | >MK708 | AATGGTGGTTTTCCCAATGA | AAATGTGGCTCTGCATGTTG |
| >MK287 | CAAAGGATGCAAGGCCTAAA | CACCACCATTTCCACCTGAT | >MK709 | ATGGAGGTGTGGAGTGGTTC | TTTGGGGCATGAAGAAGTTT |
| >MK288 | TGCAAATAACGACGGATTCA | GATGCAAGGCTACAAAATGC | >MK710 | CGTCAATTCTCCTCCATCCA | TTTCTCCCTTCCCTTTTGCT |
| >MK289 | TAGCATCCCAAATCCTCAGC | GACTCACAAAGAGGGCATT | >MK711 | TCAACGATTGCTGCTCACTT | CACTTGATGGCTCCTTTGGT |
| >MK290 | CAGCTCAGGTGGAGAAAAGG | TGTCAATGCTTCCCTTCCAT | >MK712 | AACGACGTCTCAAGCAACAA | TCCACTCGTAATTTCTCCAA |
| >MK291 | CCGTGAGGTCTCTGTTTCATT | CAACTCTTTCCCATCCTTCG | >MK713 | GACCCATTTGTGCCAGTTCT | CAGAGACCGTGGGGAAAGTA |
| >MK292 | ATAACGCAAAGGGCACAATC | GGAGGAACATCCAGGTGCTA | >MK714 | GAACTCCCTCATCACCCGTA | AATACGCGAACCACACCAAT |
| >MK293 | TGATAATGCAAAGGGCACAA | GGAGGAACATCCAGGTGCTA | >MK715 | GGACCTTAATTTCCCCAAAA | CCCCAAACAACCTCACACCT |
| >MK294 | CTTGTGGGAACGCCATC | GCTCCTTTTATCCCTCATTCA | >MK716 | CTCAGAATGCTTCAACACTTAC | TGTTGGATGGATGGAGGTTT |
| >MK295 | GCTGCGTTCAAGGAGAAGAG | GGTACATCAGTCCACGAGCA | >MK717 | CCTTTCATATCATACTTTCCCA CT | TCTCGAAACTTCTTCTTCCA |
| >MK296 | GCTTGGGGTTTTGTGTTTTG | CGTCATTTGGATCTGTGTCG | >MK718 | CGGGTTGGGTGAATTTGTA | GAGAAGACTCATGTTTGAGGGT TT |
| >MK297 | GAGGAATTGAGGAGAGGAGAA | GATGGCTGGTGGACTTTGTT | >MK719 | ATCCTCCATTCCGGTGTCAA | TCGTTCAATCCTCCAGATCC |
| >MK298 | AACAAAGTCCACCAGCCATC | TCCAATCAACCATGTGGGTA | >MK720 | CCCTTAAAGCCTGGTAGTGTG | GGAGGAAGCAAACAAAGAGC |
| >MK299 | GTCTTTGGATGGCTGCATTT | TGGTTGAGTAGGTCGCATGA | >MK721 | AAGCCTCGGATTGAATAGCA | TGTTCAACCAAAACCCCTTA |
| >MK300 | CCGGTCTGCTACTACCTCA | ATCCAACCACCACATCAGT | >MK722 | GCGGAAATGGTGAAGGAGT | GGGAAGAAGAGTGGGAGGAA |
| >MK301 | TTGCCCAACTTAGAGATTCA | TTGCATCCTTGAGGGAAAAG | >MK723 | CAAGGAGAGGGTAAACCGAAG | TAGCTCCGAGGCAAGTTTGT |
| >MK302 | CGCTGTTGAATGCTCGATAG | TCACAATCGACAAGGATGGA | >MK724 | TGGGTATTGATGGAAATGGAG | CCTATGATCCGCCGTTAAAA |
| >MK303 | AACTCACAGCACACACACCA | GATGATGAGAGCCGGTTGAT | >MK725 | ACGCTGTGGAATCGGTACTC | CTCCTCTGCTGCTTCAGCTT |
| >MK304 | TTCTTCATGGCTCTCACTGC | CTTGGATCGGAGCTGTTTTC | >MK726 | GGCTTACTGTCACCCTTTCTTT | GCACCCAAACAAGCCTCTAA |
| >MK305 | CGGAAGCAAAACCAAAACAT | TTTGATCCCGGTACACATT | >MK727 | TTGCTTCTCTTCTCTCTTTTC | GAAACTCCAATTCGCCATA |
| >MK306 | GGTCTGACCACCACCTCTA | TACTTCAGGGCGTTCAGCTT | >MK728 | AAATTGCACCCACAAAATG | CAAAGGCGTATCAAAGCTGA |
| >MK307 | CAGCCTGCAGAGTGCATAGA | CACCCAACCTCCTTTGAGAA | >MK729 | CCTTTTGAAAGCGTTCGACT | CGCCGTGAAGAGCAATATG |
| >MK308 | TAGGAGAGTTGGACGGTTGG | CTGCCAGTTTCTTCTCTCTG | >MK730 | TTGGATACCAAGGCTAGTTGAG A | TGTGTCCCAATTGAGTTTTCTG |
| >MK309 | TTGGAAGAAACGGGACAATC | GATAGCAGCAGCCCATTACC | >MK731 | CGCCTTTGTCGATTGCTTAT | CTGCTCCTCGGTGATTTCTC |
| >MK310 | TGGAATAATGAAAGACCCTTGG | ATTTGATCATCCCTCCATGC | >MK732 | GGTGTCAAGGAGACCACGTT | TTCAGCTTGCCAGTGTTCAG |
| >MK311 | CATCATCAACATTGCCAGT | GCGGGAATAGCAGATCAGAG | >MK733 | TTCACATGGGATGCTGTTTT | CCATAATTTCCAGCTCCAA |
| >MK312 | CACTTTACCCTCCCCAGACC | TGCAATGAATAGGAAGAACGTG | >MK734 | TACAAATATGCAGGGCACGA | CCTCCATTCGAGTTCATGA |

| | | | | | |
|--------|-------------------------------|------------------------------|--------|------------------------|-------------------------------|
| >MK313 | ATGGCACCTAAGGGTGTAGC | ACATCAATCCCAAGCCAAC | >MK735 | TCGCTGGAATCTTGATCCTC | GAAAATGCCCTTCTTTTCC |
| >MK314 | CATGAGAAAGTGTGCCGAGA | TTCAAATAGGGTGCTCGTGA | >MK736 | CAACAAAGAACCAGCCTTGA | CATTGTCAAGTGGTTGAAGATAG AG |
| >MK315 | TCCCCTTAACAAGCTTCTCG | TCTGCTTTCTGCTAGCTTGTTG | >MK737 | CAAAGGGGATCTACAATTCCA | TCACATGGAGCAAATCCAAA |
| >MK316 | TAAGCCATACCGCATTGTTG | CCTGTGGGATTGCACTAGGT | >MK738 | CGTTCTTTTGGGATGGAAAA | CCAAAACCCAATCAGTCCAG |
| >MK317 | CGATTGCGAAGTCTCCTTTT | CGCCTTCTTCAACAATCCAT | >MK739 | TGATTGCACACTGCATACCC | ACGTCCATCTCACCTGGTTC |
| >MK318 | ATGGTTGCCGTTGATGTGTA | CAGAAGTCGGAAACCCAAAG | >MK740 | GGTGGGTATCAAGATAATTGG | CCCCTTTGAGTTCAAGACTCC |
| >MK319 | CAATAGCATCCAACCACTCC | ATCCTGTTGAGAAGCCGTGT | >MK741 | CTTTACCCTCCCAGACCTCA | TCAAAACCCCTCTTCACCAG |
| >MK320 | CCTCAAGCAAAGAATTGCAG | CTTCTCAGAGCCATTAACATCTT G | >MK742 | AATTCCAGCAAGAGGCAAGA | AAAGGAGCGGACGGAGTAAT |
| >MK321 | TTTAGCAGGTGGGGTTGTTT | ACCGGTGGTAGTCACCAAAC | >MK743 | CATACCGAACGGTGCATGT | TGAGCATGGAATGGTGAATG |
| >MK322 | CCCTTGGGGTATGATCCTTT | AGAGCCTAGCCAACCTCCACA | >MK744 | TGTGAGCCTTGAGATTGCAG | TCTCCCAGGGTATCTTCACG |
| >MK323 | GTGAGGGCACCAGCTGTATT | CAGCAGCACCCAAATGTAAG | >MK745 | TCCTGAAACGCCATAATCAA | TTCGAATCTCAACATCCTGCT |
| >MK324 | TGGTGGTGAATTAATGGAAGC | TTGGAGGAGGAAATGAAGGA | >MK746 | CGCCTAATTTTGAATCTCCA | GTAATCTCAACATCCTGCT |
| >MK325 | TGACTCCGAAAGACTTGAGA | CCACCAATTGATCGTCTCCT | >MK747 | GCTTGTGATTGGACCACCTT | GCCATCCAGGCAAAGATTA |
| >MK326 | GCCTGGTTTTCTATATTACACCT TC | GAGGGAGGGATCATAGTGAGG | >MK748 | CGAAGAATCCAAGGTGAAGC | CTTGGCATAGGGGAATAGCA |
| >MK327 | AGTGCTAGCTTTGTTGGTGA | CTTGAGGTGAAGACATGCAATC | >MK749 | TTTCTCTTTATCGCACCACA | TGCAGCATTTTACCGCACTA |
| >MK328 | AGATTGCATGTCTTCACCTCAA | TGTTCTCTCTGTCCATCC | >MK750 | CATTCCGGTGTACCATCAAA | ACTTGTCCATCGCATCCTTG |
| >MK329 | TGATTGAGTTCCTCAGGTTACA GA | ACAACCAACTGGTGCCAAAT | >MK751 | AAGGTTGGTTCGGACAACCTG | CCAACCTCTGCATAAGCACA |
| >MK330 | GAGTTTGGGGACACTGCCTA | GAACCTGATGGTTGGGGAGA | >MK752 | TATTTACATGCAGCGCAAGC | GCAGCACAATCTCCCTTTG |
| >MK331 | AATGACAAGTGGAGTAGAGAG CA | TGGTGTGCAACTTCTGATGG | >MK753 | AGCGCAATAACTTTCTCTCCA | GGTGTTTTGTGATCATTCCAAG |
| >MK332 | TTGTCATGTGGCTCTGAAAAG | CAGCACAAACGGAGACAAGA | >MK754 | CGCAGACAAAAGGAAGACTG | TCCGAAGGTAAAGGGGAAGT |
| >MK333 | AAAACAAGACCATGTGGATGC | GGCCTTTGCTTCTTTTCTCC | >MK755 | GCCAAGAACGGTGTGGTAT | GGCTGAAGAGAAAGGCTTGA |
| >MK334 | TCCACCTCTTGAGGCATTTT | CTGTAGCCCTTTCCCTTGCT | >MK756 | AATGGTGGATTCTGTCTTGA | ATTGCCCTTTCTTTCCCTTG |
| >MK335 | TGGCATCATCTTTCCAACAA | AGTCGGCTTTGAATCTCTGG | >MK757 | GCAGCACAATCTCCCTTTG | TATTTACATGCAGCGCAAGC |
| >MK336 | CAAGCATTGGCTCTTAGGATG | GCCTCCTTTGTAGGCTCTTG | >MK758 | AATTGTGATGGCGTTGTGAA | TTGTTTATGCAAGCTCCAAG |
| >MK337 | AAGCCCACTTTTCTTTTCT | CAAAAGAGCCCGTTTACAA | >MK759 | GCCTTATAGAGTGTGGGAGAAA | TCTTGACTTTGGCATCAACA |
| >MK338 | TCAAGGAAGAAAGCCAAACA | TGTTATCGTCGTCATGAACTTG | >MK760 | GCCTTATAGAGTGTGGGAGAAA | GCAGCAGCCTCTTCACAAGT |

| | | | | | |
|--------|-----------------------------|-------------------------------|--------|------------------------------|-----------------------------|
| >MK339 | CACATTTAAAGAAGCACCTTGG | TCAAGAATCCCAAGATTTTCG | >MK761 | ACCCCTCCTCAATCTCTTGG | CCAATCATCGGATTCTGACC |
| >MK340 | GTCATTGCTGGCAGAGTCAA | TCACGACAAAGCATGGAAGT | >MK762 | CACCACCGTTCATCTCACAA | GGCTCTGACCTGACCACAAT |
| >MK341 | GCCAAAGCAAAGCACCTCTA | AAAACAGAGACCCATTGGTGA | >MK763 | AGGGAGGAAGGAATTGGAGA | CTCGACAATCCTGCACTTGA |
| >MK342 | TGGTGGTATCAGTGGATGATG | AAACAACCCCAGAAGATGGA | >MK764 | GATCACCACAGCCCAAATTC | CGATTCTTCGCCTTCTCTTG |
| >MK343 | CTCCAATTTCTTTCTCTTTCTC C | GGTCAAAATGGAATGGTGGT | >MK765 | CGGAATACGTTTTGCAGGAT | GAGGCTAGAGGTGTGGCTTG |
| >MK344 | CACGATGAGAACTGCTGAGG | TGGCGAATCCAACACACTTA | >MK766 | GGGCCACTATGCTGAACAAC | AGCTGGACTGAGAACCAGGA |
| >MK345 | GATAACCGAGGCGTATGGAG | GTGGCACGTGTGAGAGAAAA | >MK767 | TCAGGGGTGTAATGCTGTG | TTCTTGTGACCACCAACACC |
| >MK346 | TTCCTCCCAACAACGGTTTA | TTCCAGAAACCCAACCTCTCC | >MK768 | CTAATGGCAATCGCTCAACA | CCCCGGTAAGTCCATATCCT |
| >MK347 | GGAAAACGATACACCGGAAG | GGTCCTTTAACACGGTCTGC | >MK769 | TTTCTCTCCTTGGCTTGAA | TGGATGGTGCAAAATTAGAGC |
| >MK348 | TCTTCAGGACCCTGTTCCAC | TGCTCTTCGGTCACTCACAT | >MK770 | GTTACTGTGTGATTTGTGTGTG TG | TGATGGGGTTTGTTCAGAAA |
| >MK349 | TTAGTGTGGGTTGGGGTCT | TCGGCATTACGTCAAACAA | >MK771 | GGGTTTGACACTGGGAAAAG | GGGTGCATGGACTAAGAGGA |
| >MK350 | CCTGACCCGGTTACACAGAT | CATCTGCTGCTGCTGTTGTT | >MK772 | GGTATGTGGGTTACGGAAA | CAGCGGCAATTCAAAAAT |
| >MK351 | CAGGGATCATGCAACTGAGA | CACTATACCCTGCCCAGACC | >MK773 | AAGGAGGGATAGAAAGACAAA GG | GGGTAGCTATACTTGGAGGAGG A |
| >MK352 | AAAACAAGTCGAGCGTAGCC | GTAGAGTCCGTGGTGGTGGT | >MK774 | TGAGGAGCCAACAACAGCTA | TTTCGCCATTCACTTTCC |
| >MK353 | AGGAGGAGGGGAAGAGGATT | CCAGCTCCACCATTTGTCTC | >MK775 | GGAAAGAATGAATGGCGAAA | AGCAGCTAAGCCAAGATCCA |
| >MK354 | CCACCCTACCACCTTTGCTA | GGGTTGGTTGAGGTTGTTGT | >MK776 | GGGACAGAAGAAGGACATCG | CCTCCTTACTTGGCTGATGC |
| >MK355 | TCTTCATTCCTTCTCCCTAGCA | GTCCGATCAATTCGAACAAA | >MK777 | AGTACCGCCGAGAAGAAAG | TTGTCCAACAATCCGATCAA |
| >MK356 | CCACATACTTTGGCAATCACA | TGAGCAGAGGGAGAAGAAGG | >MK778 | GCATCAGTACAGCTGAGCAAA | GTAGCGAAAGGTTGGGATGA |
| >MK357 | AGCCAGATCGAGGAAGTTGA | CCGGCATTCTTGATCGTACT | >MK779 | GTGCCGAGCTGACCAAAT | CCTCGGATGATCTCCTGTGT |
| >MK358 | AAGGCTTTTGGGGAGACAAT | CCGTGTCTAAAACATGAAATGC | >MK780 | GAGCAACAACAATGGAAGCA | CAGCTTTTGGCTGATTTTGA |
| >MK359 | TTGATATTAACCATGGCAGCAG | TTGATGGGATTGCACAGAAA | >MK781 | CCCTCACTTTCTTTGCGATT | ACCATTGCTCCGGTAGACAG |
| >MK360 | CGGAACCCAATCTTCAAAAC | GCCTGGTTCAACACCAGATT | >MK782 | AGACACTTTCAACCTTTTGCAG | TGAAATAACGGGAATGGTGA |
| >MK361 | GCATTCTGCCTCGATATTC | CGTAGGCCAAAGAGTTCCAC | >MK783 | GGGTAGCGAAGGATCATTGT | CCAAATGAATGACTGGCCTA |
| >MK362 | TCACCATGGATATCGGGAAT | TTCTCTCTCCTCCCTCTCC | >MK784 | GCAGCACAAATACCCATCAA | TCACATTAGCACCCCTCTCC |
| >MK363 | ACCCCACTCCACCCTATACA | AACCGCTTCTGAAATCTTCC | >MK785 | ACGATGAACGCATCAACAAA | GCGCAATGAGACAGTCGTAG |
| >MK364 | CTCAACCTGGAATTCTTGCTG | TTCTCAAAGAGAGTATTAGGCCA AC | >MK786 | CCGGAAGTGTGACAAGGTTT | GTACTCAGCCACAGCAACGA |

| | | | | | |
|--------|-----------------------------|-----------------------|--------|-----------------------------|------------------------------|
| >MK365 | TAGGTGTGGTAATGCGTGGA | TCCCAATGTCAAACCATGAA | >MK787 | CTTGCATTCTGGGCACACTA | GTACACGGGCTATGCCAAGT |
| >MK366 | ATGGATTGGCTGAAACAGGA | CGATGGCTTTGAAAAGTGAA | >MK788 | AAAAGAACCAATCCCCTTCC | GCTCCCCACAGAATTCACAT |
| >MK367 | ACGACATGATGATGGTTGGA | TGCAAGTTCGAGTCACCAAG | >MK789 | GCGCGATCAACAATATTTC | TCTTGGCGAATAACCCCTTA |
| >MK368 | TGCAAGACATAAAGGGATTGTC | GCACTTCTAAGCATGGTAGCC | >MK790 | TTGGGGTTGCCTTAACATTT | GTCCAGGATCTCAGGAACGA |
| >MK369 | TCAGGTGAATTTGGGTAGGAA | CAGGTTGTGAACAAGCAGATG | >MK791 | TGGCAGTAATGCTGGATGAA | TCTGCATTGTCTCCAAAATGTC |
| >MK370 | ACCCAAAAGGGAAGAATTGG | CTTTCGTCTCCTCCTCCTC | >MK792 | GTGGCTTGACCTCCATCATT | CGGCTTAATCCTTGGTGAGA |
| >MK371 | ACCCAAAAGGGAAGAATTGG | TCTCTCCCTCCTCACTTTCGT | >MK793 | GATGTTCCACCAACGATGAA | CCTCTGAGTTGGTAGCATTGG |
| >MK372 | CCACTCCAAATGGGGAAATA | TTCCCTTGCAATTCTTGACC | >MK794 | AGCTTGTCTGTTTGAGCAGGT | GCAGCAAAATATCTGAAGAACT GG |
| >MK373 | TCAACAAGCTCAAGCCCTAAA | GCTTGGCTGCTACATTTGCT | >MK795 | AGCATATCCGTGTCGTGATG | AGGGGCAGTGAGAACTGAGA |
| >MK374 | AGCTGTCACCTTTGGCGAGAT | GGCCTCAGTTTTGGATGTGT | >MK796 | AAGGGCCATCTATCCCAAAG | CACCATCAAACTCTGCCTTC |
| >MK375 | CGGAATTTTCGCTCAGATTT | CCCCTGCACATCAACTGAC | >MK797 | TGAATTATCGCGTTCCCAAT | TGCTGCTGGAATTCTCCATA |
| >MK376 | TGTGAAGAAGGAGGGAGTTCA | CCCATTTTGCCAACTCAAAAC | >MK798 | CTGGAGCATTGAAGCAGACA | CTTCCAAATCCTCCCTCTCC |
| >MK377 | GTGGCAGCAGCAACAACATA | GTCGTTAGCCGGGTTGTAAA | >MK799 | TTCCACCTTCATCTTTCTCCA | GACCATCACCCCAAGCTAAA |
| >MK378 | GCAGCAAGGGCTTTCACTTA | CTCTTGTGCAGCCTTCATGT | >MK800 | ATGCATTGCGGAAGAAGAAT | GCAGTAAGGTTGACAAGAATTG G |
| >MK379 | CTGCCAGTGAGCTTGTTTTG | TGCACCATGAGAACCTGAAG | >MK801 | AGGAACAGCCATTAGCTCCA | GCATCCCCAATTTTCATCAAC |
| >MK380 | TCTCCCAAAGCTACGCCTAA | GGCCAGTGTTGCAAATATCA | >MK802 | TGGGACCACTCCATTTTCTC | TATACCAACTCCGGCACCTC |
| >MK381 | AATATTGATCACCCGGCAAC | CCGCTATTATACGCACCTTTT | >MK803 | CAAAGAGAAGCAAGAAGACAA GG | GACTTAGCGAGGCGTTTGAC |
| >MK382 | CCTTTTGCTTTGAACCCATC | CTGCTTGAAGTTCTGCTGGA | >MK804 | CCTTCGTCTTCGTCTTCGTC | ATGGTGGTGGTGATTGTGAA |
| >MK383 | CTCCAGCAGAACTTCAAGCA | TAGTTGTCGGTGTGGGAGTG | >MK805 | ATTAGCAGCAGCCAAGCAAT | CCCGACACCTAATCCAACAT |
| >MK384 | TGGAGAAGATGGAAAAGAGAAG C | GGCCGTGGTGGTTAGATTAG | >MK806 | AGGAGTGGTGGAGGAGGAAT | TGGATCCTCTCTTTGCCTTT |
| >MK385 | TACCCCTTCTCCTTCCACCT | TGGTTCATCCGATTCATTT | >MK807 | GTCAAGCACTAGGCGGACAC | GGGCCATAACTTCAAATCATCA |
| >MK386 | TGCTTCAACCTCCACCTTCT | TCATCATTCGCTGCAACTTC | >MK808 | GCCCGTCTTGCGAATATTTA | CGATTTGTGGTTTGGTTCCG |
| >MK387 | TTGGAAACATGGAGAAGTTGC | GCTTCAGATGGCCAATTGTT | >MK809 | CTTCACTCCAAATGGGGAAA | CTTGATTCCACCCGTATCGT |
| >MK388 | CCGAAAAGGAAAAGACACCA | ACCCCAAAAGGGATCTGAAA | >MK810 | GATTGGCGGGAGACAAATTA | AACCCAACGTAACCTACAACCT TT |
| >MK389 | TCCCCTCAGTCTTGATTGAAA | AACATGGTGTCTCCCCATC | >MK811 | TTCCAAGCATTTAGCCTCGT | AGGCACTTTGCGTTCAGATT |
| >MK390 | CCCCACGTAAATGAATCACC | TTGCTGGGTTGATCGAATCT | >MK812 | GCTGTTTTGGAAGGGCATAA | TCATGGAACCCAAATTTTCC |

| | | | | | |
|--------|-------------------------------|-----------------------|--------|-------------------------|------------------------|
| >MK391 | GGAGGAGGAGGAGGAAGAGA | GGGGTTTCGAAGAATTTGGT | >MK813 | TACCTCAATCCCAAGTCCA | GCAAAGGGGAGGAAGAAAAG |
| >MK392 | CACAACGATGAGGAGCAAAA | TTCTTAGCTTGGTTCCCACAA | >MK814 | CACTCACAAGCCACTCTCCA | AATTATTCGCGCAATTCGTC |
| >MK393 | GCACTTGTGCTGAGGGTACA | CAAGCCTCAAACTCGAAGAA | >MK815 | CCAAGAATCAACTCCATCAACA | TTCCCATCCAAATCCCATA |
| >MK394 | GGAATTTCTAGCAGCGGACA | GACTGGGAACGTGGGATAGA | >MK816 | TTGGGATTGATGATCCTGTG | TTCTTGGCCGCTTTCTCTTA |
| >MK395 | TGGATGGATGAGCATGAAAA | TACACCCATCAACCAAGCAA | >MK817 | GCTGCTCTTCTCACTCCTTTC | TCCCCTCTCCAAGATTTTCC |
| >MK396 | TGTCTTTGCAAGAATCTCTACT GTG | TACACCCATCAACCAAGCAA | >MK818 | GGGTGGTCATCAAAATGCTC | CTAGGTGGGAAGTGGCAAAA |
| >MK397 | CACCCCTTACCAACCTCACA | GGTCCCCTGTACCTCTCCA | >MK819 | TTACTGCTCATGCAGCCAAC | AAGAAGTCATCGTTGCACACA |
| >MK398 | AACATTTGGGACCACTACGG | GTGCAATCAAAATCCGGTAA | >MK820 | CGATGACCTGAAGCTACGTG | TTGTGCTGGATCAGTTTGGGA |
| >MK399 | GCAGCTTTGCCATGGTTC | GGCAATGTGGAGTCAAAGA | >MK821 | TGAAAAGAAGGCGTAATGGA | TACCCCAAGTCTCTGCCTA |
| >MK400 | CAGATGCAAGTTCCCAAGAA | CTGAGTTGGTGGTGTCTCG | >MK822 | GGTGTGATGATGCGAAGAA | ATCAGTCAAGCAACGAGCAG |
| >MK401 | TTCCAATCAAGCCCTCAAC | CAGAGCAGCAGTGAAATTGG | >MK823 | GAAGGCTGAGGATTGTGACC | CCCCTCTTCTGCATTTTCA |
| >MK402 | GCCTCTCCTCAGATCCCTTT | CGCATGTAGATCCCAATCCT | >MK824 | TTCTCCACCTCTCTCCTTTT | GGTCCACCACCACTACCATC |
| >MK403 | GCTGTGGAAGGCTCACAGTT | CCTCTGTTTGGGTGTTGCTT | >MK825 | ATCCAAAAGCCACAACCAAT | GCATTTGCATTTCCATTTCC |
| >MK404 | AATCCCTTCCCTCTCTTCT | AAGGAAAGTCATGGCGAGAA | >MK826 | TAATGATGGTGGGGTTGGTT | TCGCCGTTGAAGAAATTATG |
| >MK405 | TGTCCTTAAGCCGTGTGTTG | GCTGAATGCGAACATAGCTG | >MK827 | GCAGAAATCATCATCGAGCA | TATCGAATGGCACTCCACAG |
| >MK406 | GGACTGATGGGAACAGAAGC | CAAGCAGCATCCATGTGAAA | >MK828 | TGGCTCCAGATAACAACAAGT | TCGGATTAGACCAACCATGAG |
| >MK407 | CGTCTCTCGAAAGGTGAAG | ACCGATCATCACACCAAACA | >MK829 | TTTCGACTAATTCGGGTTTCG | GCAACAATCGCCTTTTCATT |
| >MK408 | GGCGTAGATTACCGCTACT | CAACTCCGCACAGCTAATGA | >MK830 | CCCCTCTCTACAATTAGGTCATC | TTCTCTAGCCATTAGGGGTTTT |
| >MK409 | CTTTCACCGGTAGGGCAATA | GACAGTCACGAAAATTTGCTA | >MK831 | AAAGGGCATGCAACTAATGG | GGTAGCCCCAATTTGAGAGA |
| >MK410 | TGCTACGGTTTGCAAGTCAA | GCATCACATAGCGAACATGAA | >MK832 | ATTCCGGTGATGGGTTGTT | ACAATCTGGTTTCGCCTTAC |
| >MK411 | TCTCTTTGTAACGCATGTGGA | GCAGGTACAGGTGCAGGAGT | >MK833 | CAGAGCATTGGGAATTTGCT | TCGAATGATCCAAAGCTTCC |
| >MK412 | ACCTGTACCTGCACCTGCTC | TAAACAAAATTGCCGCTTCC | >MK834 | ATGGTAGGGAGAAGCAAGCA | CCTGCAGTAGCCCAACTCTC |
| >MK413 | GTTTCCGAGGCATAGGAGGT | CCTCGTGTAGCGGAAAGGTA | >MK835 | TCTCTTACAAACCCTGCAA | GTAGGGCCAGCCATAATTGA |
| >MK414 | AAATGGCTTCTTCAGGTGGA | GCCGGTGATGAAGATGAGTT | >MK836 | TGACAAGTTTCGGATGTGCT | TGCTTTCCAGATACGTGCAG |
| >MK415 | GGCACAAAATGGGGTAATTG | TGTCCACGAAACAGGTGCTA | >MK837 | GGCTAACAATGTGGGACAAGT | TCCTTTTGGTAATTTCTGTCC |
| >MK416 | TTCCAAATTCTGTGTGAACG | GAGCATGACTTGCAACTTGAG | >MK838 | CTGCCAGTGGAGTCTTCTTGT | AAGGGGTTGTGCAGAGACAG |
| >MK417 | GCTGGAGAAGAAGACGGAGA | CGGTGGTGGAAAGAAATTGA | >MK839 | AAGCATAAACTCGGGAGATCA | CCACCTTCCATTATTCTTCTC |

| | | | | | |
|--------|-----------------------------|----------------------|--------|----------------------|-----------------------|
| >MK418 | TGACGACAACGACGAAGAAG | TTTCTGCAATTTGGGTCTCC | >MK840 | ACACACACCCTCCCCAAA | TGATCTCCATAGCCCATCATC |
| >MK419 | CTCCTTGAATCTTTTACACAC AC | TTTGAAGCCCAGCATTCTTT | >MK841 | GCATCACTTTGTGCTCCTCA | CAGCCAAGAGACGAGGGTAG |
| >MK420 | CCATTGCCATAACCTTCCAC | TCATCACGAACGGAGCAATA | >MK842 | AGGATGAAAGGCGAGGAAAT | TCTTCATGATTTCGACCCACA |
| >MK421 | AGAGGAGAGGGAATGCAAGG | AGCACAAAGCTGAGCATCAA | >MK843 | CCCCTTACCCTACATTAGCC | CTCTTTCTTTGCCCAAGTGC |
| >MK422 | TGAACTGGATTTGGAAGGAGA | TTTCTGTCTGAAGGGATTGA | >MK844 | CCATCAAAAACCTCTTCCTC | TAGATCGCATTAGCCACCT |
| | | | >MK845 | GGGACGAGGGAAAGTGTGT | GCAAGCCACCTTTATTTACCA |

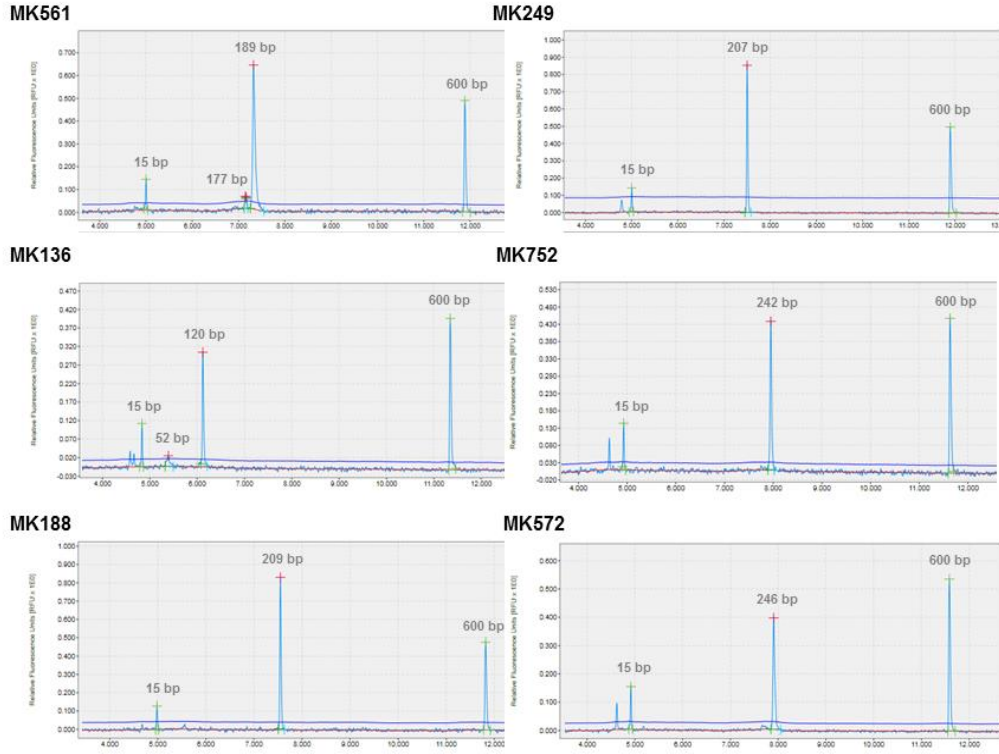
4.5. Markör Dizayn ve Haritalama Süreçlerinin Deneysel Doğrulanması

Tür içi ve türler arası polimorfizm potansiyeli, GMATA programı “Mapping” modülüne hem patates (*Solanum tuberosum*) hem de yakın akraba tür domates (*Solanum lycopersicum*) kromozom sekansları yüklenerek belirlenen markörler *Solanum lycopersicum* ve *Solanum tuberosum*'un kromozomları üzerinde primer sekanslarının amplifikasyon profillerinin elde edilmesinde kullanılmıştır. *Solanum tuberosum* cDNA sekanslarında geliştirilen yeni 845 SSR markör havuzundan seçilen ve tek-kopya (single-copy) lokusları temsil eden bazı markörler türler arası amplifikasyon ve polimorfizm için deneysel olarak da test edilmek amacıyla genotipleme, gelişmiş hassasiyet ve yüksek çözünürlüklü görüntüleme sağlayan kapiler elektroforez sisteminde yürütülmüştür. Kapiler elektroforez yürütmesi sonucunda PCR amplikon büyüklükleri belirlenmiştir. Test edilen tüm markörler, e-Mapping sonuçları ile öngörülene benzer, beklenen boyutlarda amplikonları başarıyla üreterek (e-Mapping sonuçlarında ± 50 bp), in silico markör dizayn ve transfer edilebilirlik analizlerinde yüksek verimin alınabileceğini kanıtlamıştır. Bu çalışma kapsamında geliştirilen markörler içinden temsili bir markör setinin amplifikasyon profillerini gösteren kapiler elektroferogramlar Şekil 4.1'de verilmiştir. Benzer çalışmalarda SSR markörlerinin amplifikasyon başarısı, yüksek verimle tasarlanan in silico SSR geliştirme yaklaşımları sonucunda araştırma grupları tarafından değerlendirilmiş ve rapor edilmiştir.

Havuç genomunda yapılan bir çalışmada kromozom bağlantılı SSR markörleri geliştirmek için in silico yaklaşımı kullanılmıştır. Toplam 55.386 markör geliştirilmiş ve protein kodlama dizilerine karşılık gelen markör lokusları belirlenmiştir. In silico haritalama analizi, bunların 51.160'ının tek lokus markörleri olduğunu ve 4.226'nın birden fazla lokusu amplifiye ettiğini öngörmüştür (Uncu ve Uncu, 2019).

Capsicum chinense genomunda yapılan bir çalışmada, trinükleotid SSR'lar tespit edilmiştir ve tekrar lokusları 53.749 PCR markörüne dönüştürülmüştür. *C. annuum* genomuna markör aktarılabilirliği, 17.992 aktarılabilir markör tanımlayarak, türler arası haritalama parametreleriyle analiz edilmiştir. Polimorfizm seviyesi yüksek lokusların seçilmesi ile iki biber türü arasında, 12 biber kromozomu üzerine dağıtılan polimorfizm potansiyeli yüksek 4994 SSR markör geliştirilmiştir. Aday polimorfik markörlerden seçilen bir markör seti ile yapılan laboratuvar çalışmaları, hem biber türünden %100'lük bir amplifikasyon oranıyla sonuçlanmış hem de 31 markör polimorfik aleller üretmiştir.

Böylece, *in silico* markör geliştirme ve türler arası haritalama analizleri, primer dizayn doğruluğunda ve beklenen alellerin hesaplanmasında oldukça verimli olduğunu kanıtlanmıştır (Uncu, 2018).



Şekil 4. 1. Yüksek çözünürlüklü kapiler elektroforez ile görüntülenen bazı polimorfik SSR markörleri

4.6. Polimorfik Fonksiyonel Markörlerin Protein Anotasyonu ve 3D Modellemeleri

Solanum tuberosum kromozomlarında haritalanan ve polimorfizm potansiyeli türler arası ‘Mapping’ analizi ile belirlenen markörlerin peptit sekansı düzeyinde anotasyonları ve 3D (üç boyutlu) modelleme analizleri Swiss Prot server’ı kullanılarak gerçekleştirilmiştir, proteinin işlevi, post-translasyonel modifikasyonlar, domainler (kalsiyum bağlama bölgeleri, ATP bağlama bölgeleri, zing finger, homebox, SH2 ve SH3 alanları vb.) ve 3D yapıda polimorfizme dayalı olası değişimler belirlenmiştir. 3D yapı oluşturma analizleri, polimorfik bulunan markör lokuslarının alel alternatiflerinin tamamı için gerçekleştirilmiştir.

Swiss Prot Server ‘homoloji-temelli modelleme’ yapan bir yazılım sunucusudur. İlgilenilen bir proteinin yapısal bir modelini oluşturmak için evrimsel olarak ilişkili yapıların taranmasına dayanır. Bir homoloji modeli oluşturmak dört adımda gerçekleşir:

1. Yapısal şablonların belirlenmesi
2. Hedef dizi ve şablon yapılarının hizalanması
3. Model oluşturma
4. Model kalitesinin değerlendirilmesi

Kısaca, hedef proteinle evrimsel olarak ilişkili proteinleri belirlemek için hassas dizi arama araçlarıyla deneysel olarak belirlenmiş protein yapılarından oluşan bir kitaplık aranır. Homoloji kalıplarının modelleri üzerinden sorgu peptid sekansının modeli üretilir. Bir veya daha fazla şablon tanımlanırsa, hedefin hizalanması bilgisi, şablonların 3D koordinatları ile birlikte şablon dizileri, ilgilenilen protein için yapısal bir model oluşturmak amacıyla kullanılır (Waterhouse ve ark., 2018).

Solanum lycopersicum kromozomlarına aktarılabilişliđi ve polimorfizm potansiyeli belirlenen 99 markör için homolojiye dayalı protein modelleme işlemleri yapılmıştır. 99 markörün denk düştüğü *Solanum tuberosum* cDNA sekansları 6-frame translasyonları gerçekleştirilerek protein sekanslarına çevrilmiştir. Protein sekansları SWISS-MODEL sunucusuna yüklenerek hedefin hizalanma bilgisi, şablonların 3D koordinatları ile birlikte şablon dizileri, ilgilenilen protein için yapısal bir protein modeli oluşturulmuştur. Modelleme sonucunda hedef proteinle evrimsel olarak ilişkili şablon proteinler ve işlevleri belirlenmiştir. Hesaplanan modellerin dizi kimliđi, dizi benzerliđi model kalite bilgileri (QMEAN deđerleri) hedefin öngörülen ikincil yapısı ile şablon arasındaki hizalama bilgileri de elde edilmiştir. En uygun şablonları seçmek için, bu bilgiler kullanılmıştır. Markör lokuslar için modellenen protein bilgileri Çizelge 4.11.’de verilmiştir.

Çizelge 4. 11. Markör lokuslar tarafından kodlanan peptit sekanslarının 3D modelleme (üç boyutlu modelleme) sonuçları.

| Markör ID | Şablon | Tanım | QMEAN-Z | QMQE | QMEANglobal |
|-----------|----------|-------------------------------------------------|---------|------|-------------|
| >MK136 | 3mb3.1.A | PH-interacting protein | -1,57 | 0,06 | 0,56 |
| >MK157 | 1h88.1.C | MYB Proto-oncogene protein | -1,66 | 0,13 | 0,66 |
| >MK247 | 3ulx.1.A | Stress-induced transcription factor NAC1 | -3,75 | 0,20 | 0,65 |
| >MK283 | 5w3x.1.B | Disease resistance protein RRS1 | -3,93 | 0,09 | 0,46 |
| >MK402 | 7p8k.1.A | Disease resistance protein RRS1 | -1,69 | 0,09 | 0,62 |
| >MK409 | 5gnj.3.A | Transcription factor MYC2 | 0,35 | 0,02 | 0,60 |
| >MK438 | 5xd6.1.A | Protein kinase superfamily protein | -1,84 | 0,23 | 0,65 |
| >MK439 | 5xd6.1.A | Protein kinase superfamily protein | -1,84 | 0,23 | 0,65 |
| >MK480 | 3p86.1.A | Serine/threonine-protein kinase CTR1 | -0,64 | 0,23 | 0,74 |
| >MK481 | 4im0.1.A | Serine/threonine-protein kinase TBK1 | -5,31 | 0,18 | 0,46 |
| >MK495 | 6l5j.1.A | Rootletin | 0,92 | 0,04 | 0,57 |
| >MK57 | 2ayd.1.A | WRKY transcription factor 1 | 0,45 | 0,11 | 0,68 |
| >MK690 | 5b3g.1.A | Protein Scarecrow | -2,54 | 0,24 | 0,61 |
| >MK693 | 5gnj.1.A | Transcription factor MYC2 | 1,48 | 0,03 | 0,54 |
| >MK698 | 5gnj.1.A | Transcription factor MYC2 | 1,48 | 0,03 | 0,54 |
| >MK699 | 5gnj.1.A | Transcription factor MYC2 | 1,48 | 0,03 | 0,54 |
| >MK90 | 2ayd.1.A | WRKY transcription factor 1 | -1,17 | 0,12 | 0,67 |
| >MK15 | 5zd4.1.A | Maltose-binding periplasmic protein | 2,07 | 0,08 | 0,62 |
| >MK345 | 6j9b.1.A | B3 domain-containing transcription factor FUS3 | -1,36 | 0,30 | 0,77 |
| >MK353 | 2ebi.1.A | DNA binding protein GT-1 | -4,55 | 0,06 | 0,37 |
| >MK375 | 6sc5.1.A | E3 ubiquitin-protein ligase RNF31 | -5,34 | 0,07 | 0,23 |
| >MK473 | 6a5p.1.I | DNA-directed RNA polymerase subunit | -1,66 | 0,01 | 0,42 |
| >MK474 | 6kks.1.A | Transcription factor WER | -0,56 | 0,15 | 0,77 |
| >MK487 | 4l0r.1.A | Centrosomal protein of 57 kDa | -1,21 | 0,12 | 0,39 |
| >MK52 | 3g4s.1.0 | 50S ribosomal protein L37Ae | -3,29 | 0,01 | 0,36 |
| >MK643 | 2yqi.1.A | High mobility group protein B3 | -0,20 | 0,01 | 0,52 |
| >MK650 | 5oqo.1.B | Condensin complex subunit 2 | -2,08 | 0,01 | 0,19 |
| >MK653 | 1ryt.1.A | Rubrythrin | -0,96 | 0,00 | 0,41 |
| >MK669 | 6fx5.1.A | Microphthalmia-associated transcription factor | -2,59 | 0,00 | 0,32 |
| >MK437 | 5eyb.1.A | DNA-binding protein reb1 | -3,48 | 0,16 | 0,52 |
| >MK529 | 4atk.1.A | Microphthalmia-associated transcription factor | 0,16 | 0,04 | 0,52 |
| >MK548 | 2juh.1.A | Telomere binding protein TBP1 | -2,48 | 0,04 | 0,49 |
| >MK570 | 6g1l.1.A | Microphthalmia-associated transcription factor | -1,00 | 0,03 | 0,46 |
| >MK708 | 5gnj.1.A | Transcription factor MYC2 | 1,22 | 0,04 | 0,65 |
| >MK755 | 4m59.1.A | Chloroplast pentatricopeptide repeat protein 10 | -3,58 | 0,26 | 0,59 |
| >MK216 | 3zqc.3.A | MYB3 | -1,40 | 0,17 | 0,64 |
| >MK218 | 6ryl.1.B | Protein Wuschel | 0,33 | 0,14 | 0,73 |
| >MK324 | 2gcc.1.A | ATERF1 | -0,71 | 0,05 | 0,61 |

| | | | | | |
|--------|-----------|--------------------------------------------------------------------------|-------|------|------|
| >MK344 | 5tos.1.A | Serine/threonine-protein kinase BIK1 | -3,73 | 0,28 | 0,62 |
| >MK447 | 4m59.1.A | Chloroplast pentatricopeptide repeat protein 10 | -3,58 | 0,26 | 0,59 |
| >MK518 | 6g6l.1.A | Myc proto-oncogene protein | 0,68 | 0,00 | 0,58 |
| >MK552 | 4m59.1.A | Chloroplast pentatricopeptide repeat protein 10 | -7,14 | 0,32 | 0,47 |
| >MK556 | 6qdv.1.U | Cell division cycle 5-like protein | -1,22 | 0,21 | 0,68 |
| >MK557 | 6qdv.1.U | Cell division cycle 5-like protein | -1,22 | 0,21 | 0,68 |
| >MK558 | 6qdv.1.U | Cell division cycle 5-like protein | -1,22 | 0,21 | 0,68 |
| >MK656 | 7o3w.1.D | Protein sll0617 | -0,99 | 0,08 | 0,56 |
| >MK657 | 2gcc.1.A | ATERF1 | -1,52 | 0,10 | 0,50 |
| >MK659 | 4ox0.2.C | Developmental protein sepallata 3 | 2,15 | 0,11 | 0,75 |
| >MK666 | 1ryt.1.A | Rubrerythrin | -1,08 | 0,00 | 0,38 |
| >MK166 | 3ulx.1.A | Stress-induced transcription factor NAC1 | -6,11 | 0,14 | 0,52 |
| >MK170 | 5air.1.B | Low-density lipoprotein receptor-related protein 6 | -1,73 | 0,40 | 0,74 |
| >MK261 | 6e7i.1.A | Polypeptide N- acetylgalactosaminyltransferase 2 | -4,86 | 0,11 | 0,59 |
| >MK428 | 4hc7.1.A | Trans-acting T-cell-specific transcription factor GATA-3 | -1,24 | 0,03 | 0,60 |
| >MK449 | 115y.1.A | C4-dicarboxylate transport transcriptional regulatory protein dctd | -0,95 | 0,09 | 0,61 |
| >MK645 | 4uut.1.A | Homeotic Protein Ultrabithorax | -2,68 | 0,06 | 0,53 |
| >MK706 | 5t0q.1.A | Transcription factor MYC3 | -6,69 | 0,21 | 0,44 |
| >MK764 | 6zfv.1.A | Endothelial transcription factor GATA-2 | -5,67 | 0,02 | 0,39 |
| >MK770 | 5xd6.2.A | Protein kinase superfamily protein | -1,10 | 0,45 | 0,82 |
| >MK812 | 7mq8.25.A | WD repeat-containing protein 36 | -7,01 | 0,46 | 0,46 |
| >MK17 | 1wj2.1.A | Probable WRKY transcription factor 4 | -4,29 | 0,06 | 0,57 |
| >MK178 | 6wmi.1.A | Zinc finger protein 410 | -4,50 | 0,10 | 0,41 |
| >MK737 | 6y36.1.B | CCAAT-binding factor complex subunit HapC | -0,62 | 0,17 | 0,67 |
| >MK75 | 5ly0.1.A | LOB family transfactor Ramosa2.1 | 0,05 | 0,37 | 0,75 |
| >MK752 | 5ly0.1.A | LOB family transfactor Ramosa2.1 | 0,59 | 0,12 | 0,66 |
| >MK757 | 5ly0.1.A | LOB family transfactor Ramosa2.1 | 0,59 | 0,12 | 0,66 |
| >MK76 | 5ly0.1.A | LOB family transfactor Ramosa2.1 | -0,30 | 0,28 | 0,71 |
| >MK185 | 2zoq.1.A | Mitogen-activated protein kinase 3 | -2,44 | 0,25 | 0,69 |
| >MK303 | 3ulx.1.A | Stress-induced transcription factor NAC1 | -4,18 | 0,22 | 0,63 |
| >MK775 | 2ayd.1.A | WRKY transcription factor 1/ | -2,22 | 0,12 | 0,48 |
| >MK354 | 2ebi.1.A | DNA binding protein GT-1 | -2,10 | 0,04 | 0,39 |
| >MK417 | 5i9f.1.A | Pentatricopeptide repeat protein dPPR-U10 | -2,23 | 0,31 | 0,69 |
| >MK508 | 1wj6.1.A | KIAA0049 protein | -2,16 | 0,05 | 0,59 |

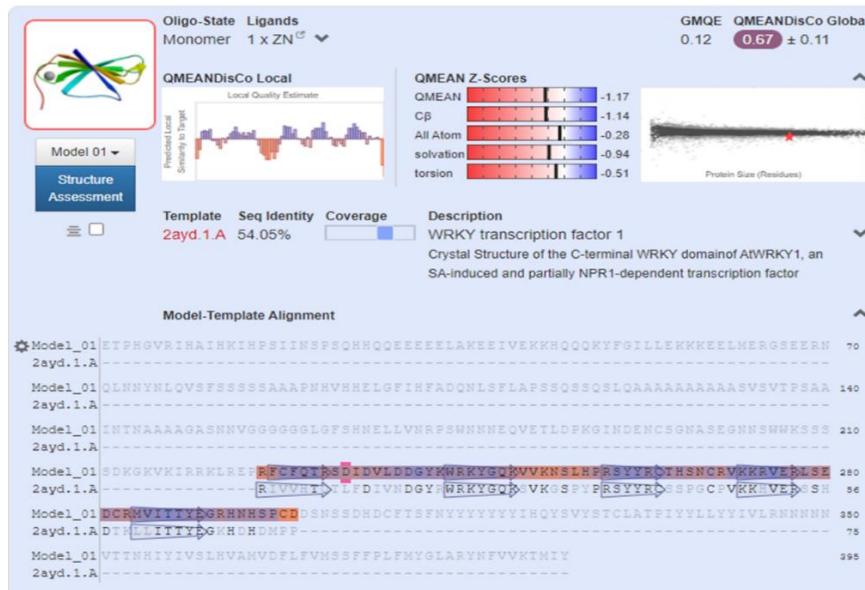
| | | | | | |
|--------|----------|----------------------------------------------------------------------------------------|-------|------|------|
| >MK545 | 2mou.1.A | StAR-related lipid transfer protein 6 | -6,28 | 0,10 | 0,50 |
| >MK628 | 4b8m.2.A | Aurora Kinase B-A | -0,93 | 0,53 | 0,77 |
| >MK268 | 4ywc.1.A | Transcription factor MYC3 | -4,25 | 0,09 | 0,44 |
| >MK304 | 6g1l.1.A | Microphthalmia-associated transcription factor | 0,20 | 0,04 | 0,51 |
| >MK387 | 6s6q.2.A | LRR receptor-like serine | -2,55 | 0,45 | 0,71 |
| >MK575 | 5xd6.1.A | Protein kinase superfamily protein | -2,68 | 0,22 | 0,63 |
| >MK576 | 2e1o.1.A | Homeobox protein PRH | -0,96 | 0,10 | 0,60 |
| >MK677 | 5wjq.1.C | Zinc finger protein 568 | -4,59 | 0,10 | 0,40 |
| >MK818 | 5vmz.1.A | Transcriptional regulator Kaiso | -1,88 | 0,04 | 0,43 |
| >MK819 | 5vmz.1.A | Transcriptional regulator Kaiso | -1,88 | 0,04 | 0,43 |
| >MK10 | 5eyb.1.A | DNA-binding protein reb1 | -3,19 | 0,20 | 0,49 |
| >MK231 | 5aq8.1.A | OFF7_DB12V4 | -4,34 | 0,10 | 0,56 |
| >MK292 | 1yw0.1.A | tryptophan 2,3-dioxygenase | -0,65 | 0,02 | 0,51 |
| >MK376 | 3a01.1.A | Homeodomain-containing protein | -1,35 | 0,16 | 0,63 |
| >MK407 | 5o26.1.A | Serine/threonine-protein kinase WNK3 | -2,88 | 0,40 | 0,67 |
| >MK784 | 5tos.1.A | Serine/threonine-protein kinase BIK1 | -3,52 | 0,34 | 0,65 |
| >MK423 | 5gat.1.C | Nitrogen Regulatory protein area | -2,87 | 0,02 | 0,34 |
| >MK810 | 3zqc.1.A | MYB3 | -3,92 | 0,15 | 0,58 |
| >MK197 | 2ayd.1.A | WRKY transcription factor 1 | -1,35 | 0,09 | 0,66 |
| >MK208 | 2dmn.1.A | Homeobox protein TGIF2LX | -1,42 | 0,06 | 0,60 |
| >MK209 | 6zmw.1.u | Eukaryotic translation initiation factor 4 gamma 1 | -2,75 | 0,26 | 0,57 |
| >MK33 | 7cvo.1.E | Chimera of Nuclear transcription factor Y subunit C-4 and Zinc finger protein constans | 2,02 | 0,04 | 0,68 |
| >MK426 | 1ul4.1.A | Squamosa Promoter Binding protein-like 4 | -3,84 | 0,07 | 0,54 |
| >MK540 | 5ux6.2.A | GDP-fucose protein O-fucosyltransferase 1 | -6,81 | 0,20 | 0,43 |
| >MK58 | 4m59.1.A | Chloroplast pentatricopeptide repeat protein 10 | -2,71 | 0,20 | 0,61 |
| >MK73 | 6j4k.1.A | Protein phosphate starvation response 1 | 1,11 | 0,07 | 0,67 |
| >MK730 | 1h88.1.C | MYB Proto-Oncogene Protein | -1,67 | 0,15 | 0,69 |

Model kalitesi, ortalama kuvvet potansiyelleri olarak ifade edilen çeşitli istatistiksel tanımlayıcıları kullanan yerel bileşik puanlama fonksiyonu QMEAN ile değerlendirilir, modelin geometrik özellikleri (çift atomik mesafeler, burulma açıları, çözücü erişilebilirliği) deneyselden elde edilen istatistiksel dağılımlarla karşılaştırılır. Doğal yapıya beklenen benzerliği açıklayan 0 ile 1 arasında bir güvenilirlik puanı atanır. Daha yüksek değerler (1'e yaklaşan), daha yüksek model güvenilirliğini gösterir. Global QMEAN tahminleri, elde edilen değerleri bir dizi yüksek çözünürlüklü X-ışını yapısından hesaplanan puanlarla ilişkilendiren bir Z- skoru olarak verir. QMEAN tahminini, hedef şablon hizalamasından elde edilen global kalite tahmin puanı GMQE ile

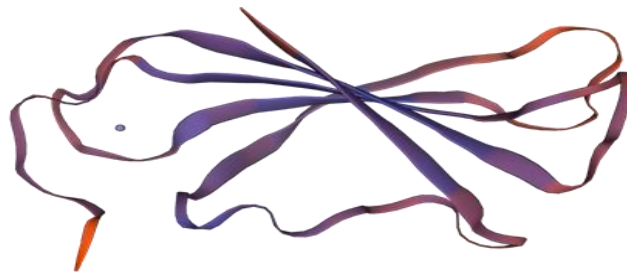
birleştiren birleşik bir kalite tahmini sağlar. Ortaya çıkan GMQE tekrar sıfır ile bir arasında bir sayı olarak ifade edilir, burada daha yüksek sayılar daha yüksek güvenilirliği gösterir. Hedef dizi için tanımlanan uygun şablonlar, tahmin edilen global kalite tahmin puanına (GMQE) göre sıralanan bir tablo biçiminde listelenir. Her şablon biyolojik olarak ilgili ligandları, öngörülen oligomerik yapı korumasını ve hedef şablon hizalamasını listeler. Modeller etkileşimli olarak görüntülenebilir, başlangıçta modelin iyi veya kötü modellenmiş bölgelerini vurgulamak için QMEAN skoru üzerinden atanan model kalite tahminleriyle renklendirilir (Waterhouse ve ark., 2018).

Her kromozomdan seçilen bazı markörlerin protein modeli sayısal parametreleri, model-şablon hizalaması ve üç boyutlu protein model şekilleri aşağıda verilmiştir (Şekil (4.2-4.13)).

a.



b.

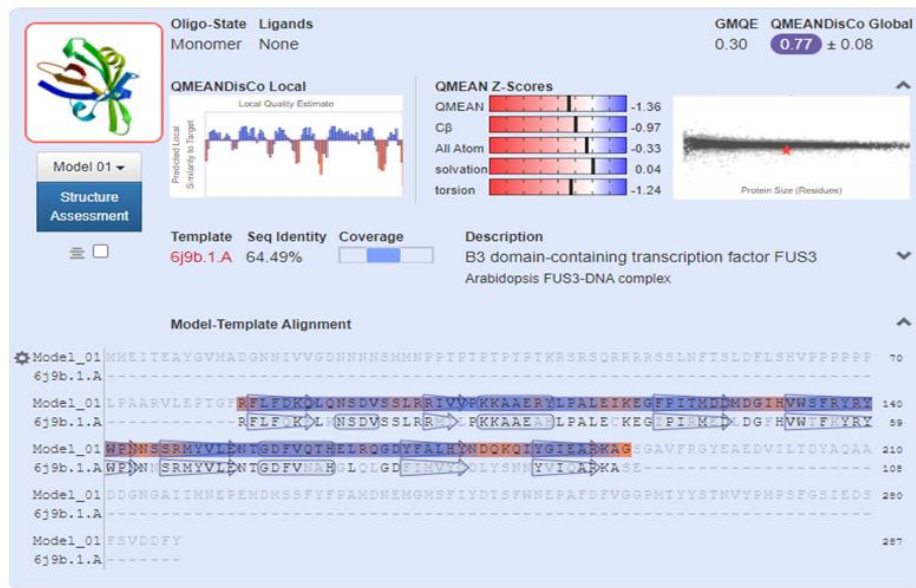


Şekil 4. 2. Kromozom 1 üzerinde haritalanan >MK90 markörünün üç boyutlu protein modeli

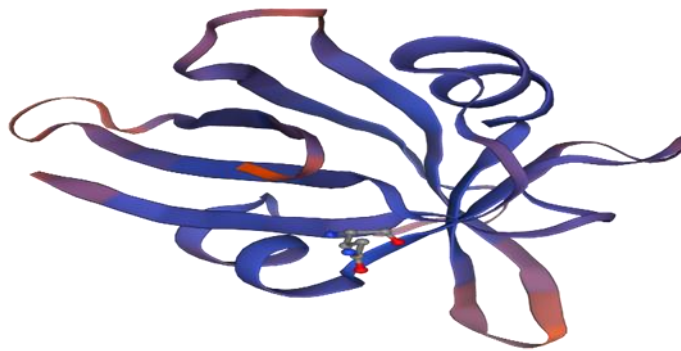
a. Protein modelinin sayısal parametreleri.

b. Üç boyutlu model

a.

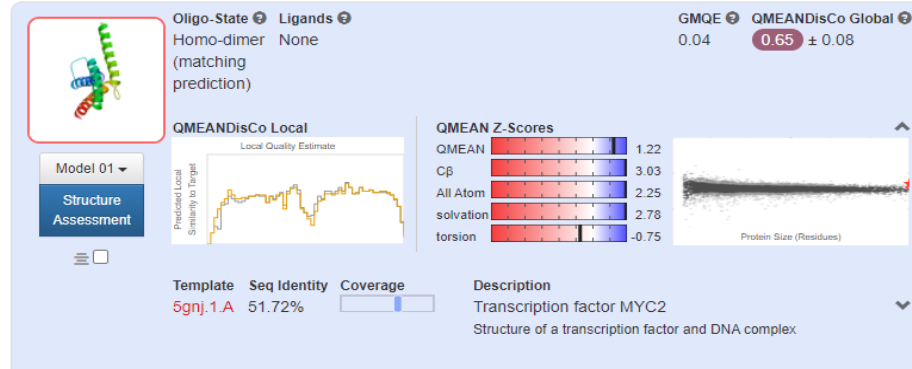


b.



Şekil 4. 3. Kromozom 2 üzerinde haritalanan >MK345 markörünün üç boyutlu protein modeli.
 a. Protein modelinin sayısal parametreleri.
 b. Üç boyutlu model

a.



b.

| | | |
|------------|------------------------------------------------------------------------------|-----|
| Model_01:A | EIGQSSSNLKGFDKNKEWNSDKRKF SINDEFEDVSM DGT LNYDSDEFIDISNKTEDGKIKIGENSSNAAST | 350 |
| Model_01:B | EIGQSSSNLKGFDKNKEWNSDKRKF SINDEFEDVSM DGT LNYDSDEFIDISNKTEDGKIKIGENSSNAAST | 350 |
| Sgnj.1.A | ----- | |
| Model_01:A | VSGADQKGKKKGLPAAK N LMAERRRRKKLNDRLYMLRSVVF K I K MDRASILGDAIKYLKELQQDINELHN | 420 |
| Model_01:B | VSGADQKGKKKGLPAAK N LMAERRRRKKLNDRLYMLRSVVF K I K MDRASILGDAIKYLKELQQDINELHN | 420 |
| Sgnj.1.A | -----PLNHVEAERTRREKLNQRFYALRAVVPNVSKMDQASILGDAIAYINELRSKVVKTIES | 62 |
| Model_01:A | ELSTPANSSLSPATSFYPLTPTASALPSRIKEELVPSSFPSPSSPTGQPARIEVRVREGRAVNIHMIC | 490 |
| Model_01:B | ELSTPANSSLSPATSFYPLTPTASALPSRIKEELVPSSFPSPSSPTGQPARIEVRVREGRAVNIHMIC | 490 |
| Sgnj.1.A | ----- | 63 |
| Model_01:A | SRKPGVLLSTMKTLDLSGLDIQQAVISCFNGFVLDVFRAEQSNEGQDIHPDQIKAVLMDSAGFNGGVIAA | 560 |
| Model_01:B | SRKPGVLLSTMKTLDLSGLDIQQAVISCFNGFVLDVFRAEQSNEGQDIHPDQIKAVLMDSAGFNGGVIAA | 560 |
| Sgnj.1.A | ----- | |

c.



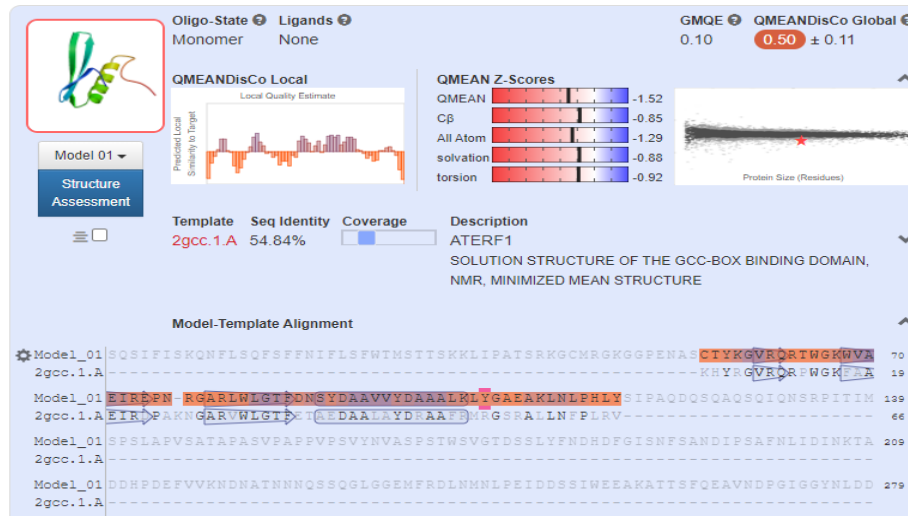
Şekil 4. 4. Kromozom 3 üzerinde haritalanan >MK708 markörünün üç boyutlu protein modeli.

a. Protein modelinin sayısal parametreleri

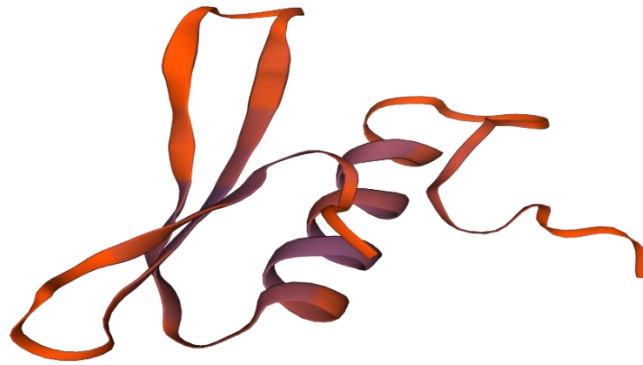
b. Model-Şablon dizi hizalaması

c. Üç boyutlu model

a.



b.

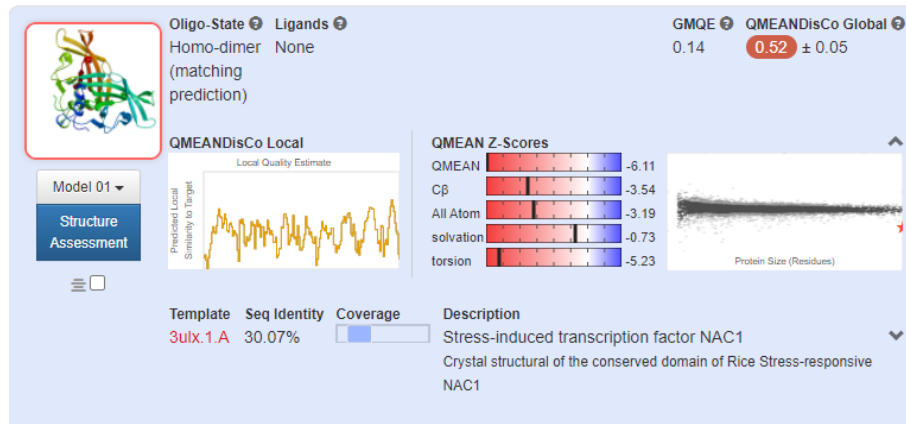


Şekil 4. 5. Kromozom 4 üzerinde haritalanan >MK657 markörünün üç boyutlu protein modeli.

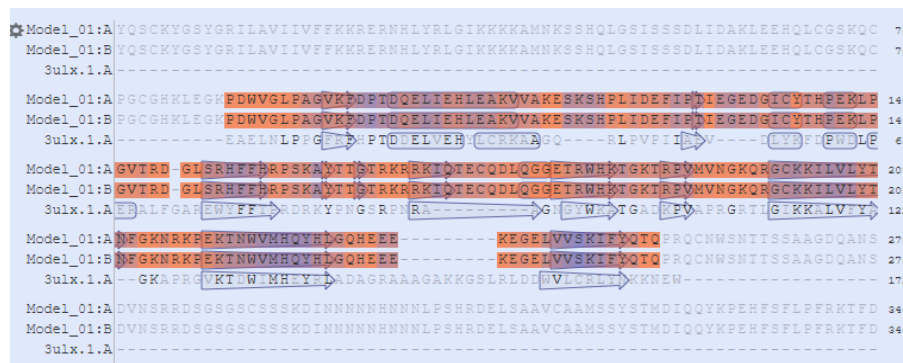
a. Protein modelinin sayısal parametreleri.

b. Üç boyutlu model

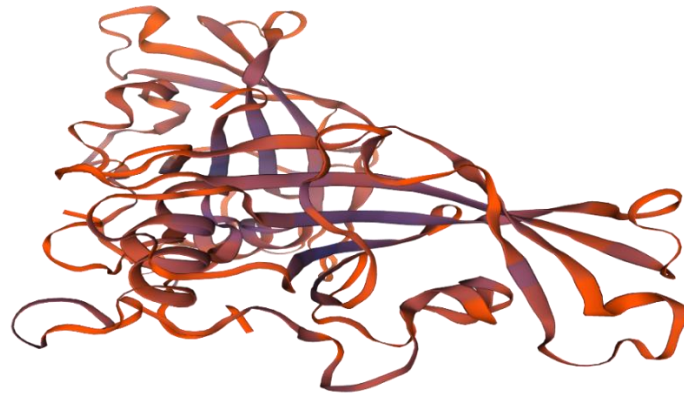
a.



b.



c.



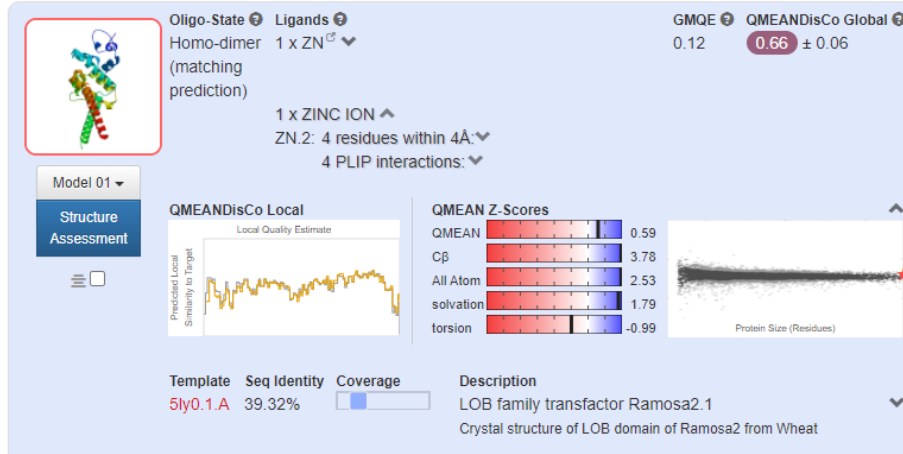
Şekil 4. 6. Kromozom 5 üzerinde haritalanan >MK166 markörünün üç boyutlu protein modeli

a. Protein modelinin sayısal parametreleri

b. Model-Şablon dizi hizalaması

c. Üç boyutlu model

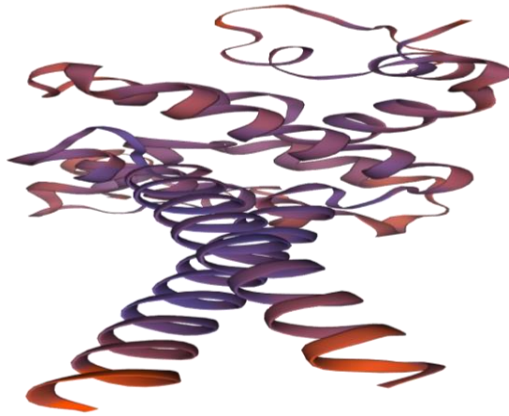
a.



b.

| | | |
|------------|------------------------------------------------------------------------|-----|
| Model_01:A | NSLISFFSIPAGDAAQISLQTYNYNMQNMSTFTTPNNNNNNIPIHQTPHTFITDHNLNHNKYTKRVH | 70 |
| Model_01:B | NSLISFFSIPAGDAAQISLQTYNYNMQNMSTFTTPNNNNNNIPIHQTPHTFITDHNLNHNKYTKRVH | 70 |
| 5ly0.1.A | ----- | |
| Model_01:A | NNNDKNVLLRVSGAGCAACAKYQRRKCAPDCVLAIFYFPHDRQRQFLNAHKLFGVSNITKIIRHLDPQFK | 140 |
| Model_01:B | NNNDKNVLLRVSGAGCAACAKYQRRKCAPDCVLAIFYFPHDRQRQFLNAHKLFGVSNITKIIRHLDPQFK | 140 |
| 5ly0.1.A | -----APTTFGAGAPCAACKFIRRKCLPGCVFAPYFPEEIQKFAVNHK)FG(SNVTKLLNL)FPHQR | 64 |
| Model_01:A | DEAMRTIIFQSDVRANDPVGGCYRIIRDQLQRHIDYCKAELDIVIHQLAYCRAQVAASEQQQILTDHAT | 210 |
| Model_01:B | DEAMRTIIFQSDVRANDPVGGCYRIIRDQLQRHIDYCKAELDIVIHQLAYCRAQVAASEQQQILTDHAT | 210 |
| 5ly0.1.A | EDAVSSLAYEARV)DPVYG(CVGAISVLRQVHRLQKELDAAHTELLRY)CGE----- | 118 |
| Model_01:A | EESFNDNCCDGVVNCADALVNNNNWASSYDQQPIDQYNPHDSFYNHPRCDQQAEDDQQYDGIQDGLDLT | 280 |
| Model_01:B | EESFNDNCCDGVVNCADALVNNNNWASSYDQQPIDQYNPHDSFYNHPRCDQQAEDDQQYDGIQDGLDLT | 280 |
| 5ly0.1.A | ----- | |

c.



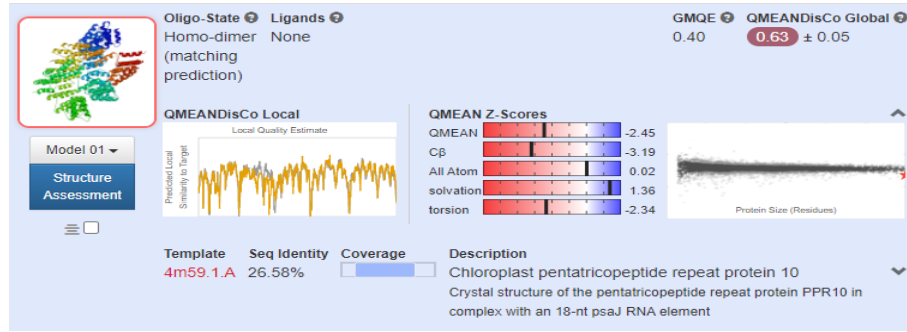
Şekil 4. 7. Kromozom 6 üzerinde haritalanan >MK752 markörünün üç boyutlu protein modeli.

a. Protein modelinin sayısal parametreleri

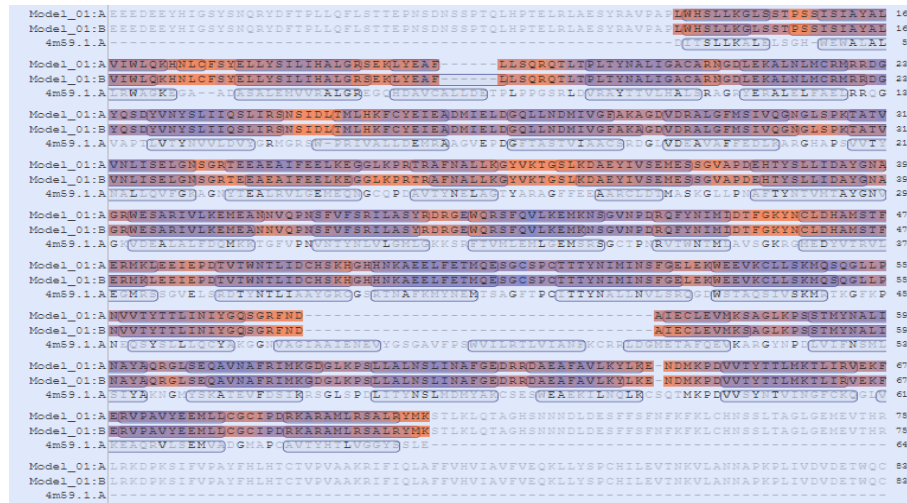
b. Model-Şablon dizi hizalaması

c. Üç boyutlu model

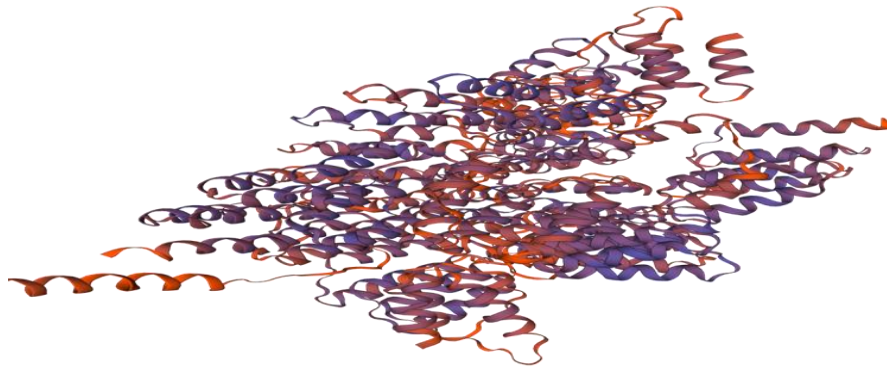
a.



b.



c.



Şekil 4. 9. Kromozom 8 üzerinde haritalanan >MK417 markörünün üç boyutlu protein modeli.

Protein modelinin sayısal parametreleri

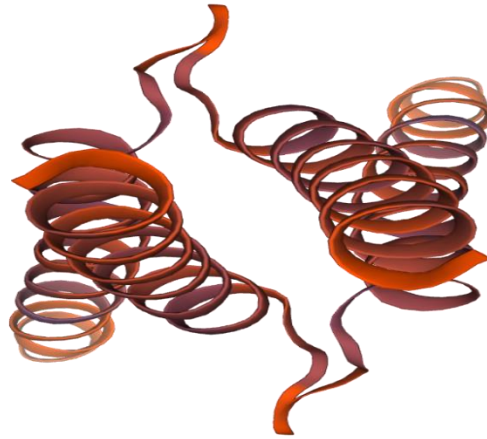
b. Model-Şablon dizi hizalaması

c. Üç boyutlu model

a.



b.

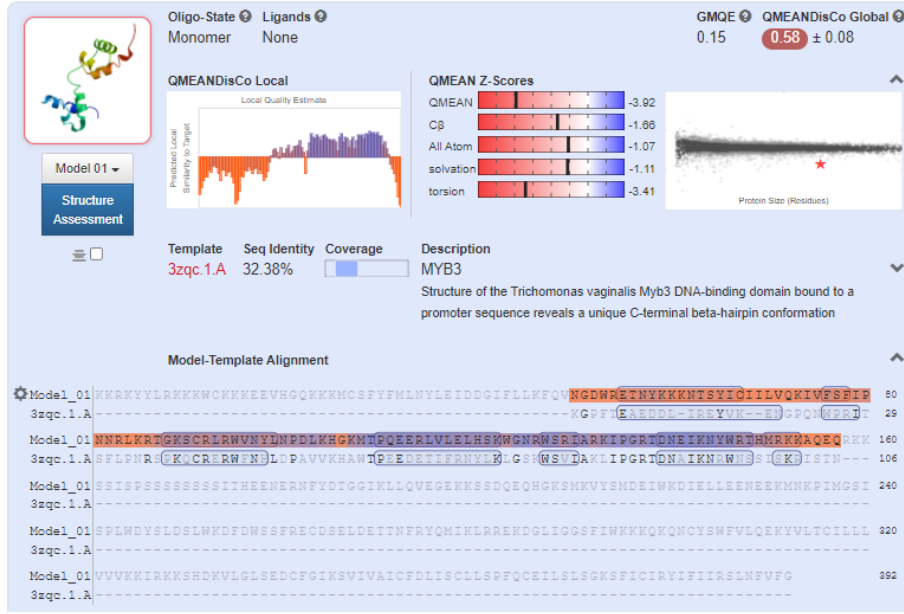


Şekil 4. 10. Kromozom 9 üzerinde haritalanan >MK304 markörünün üç boyutlu protein modeli.

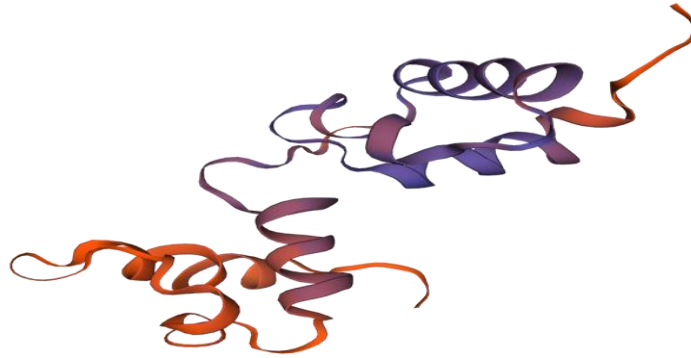
a. Protein modelinin sayısal parametreleri ve Model-Şablon dizi hizalaması

b. Üç boyutlu model

a.



b.

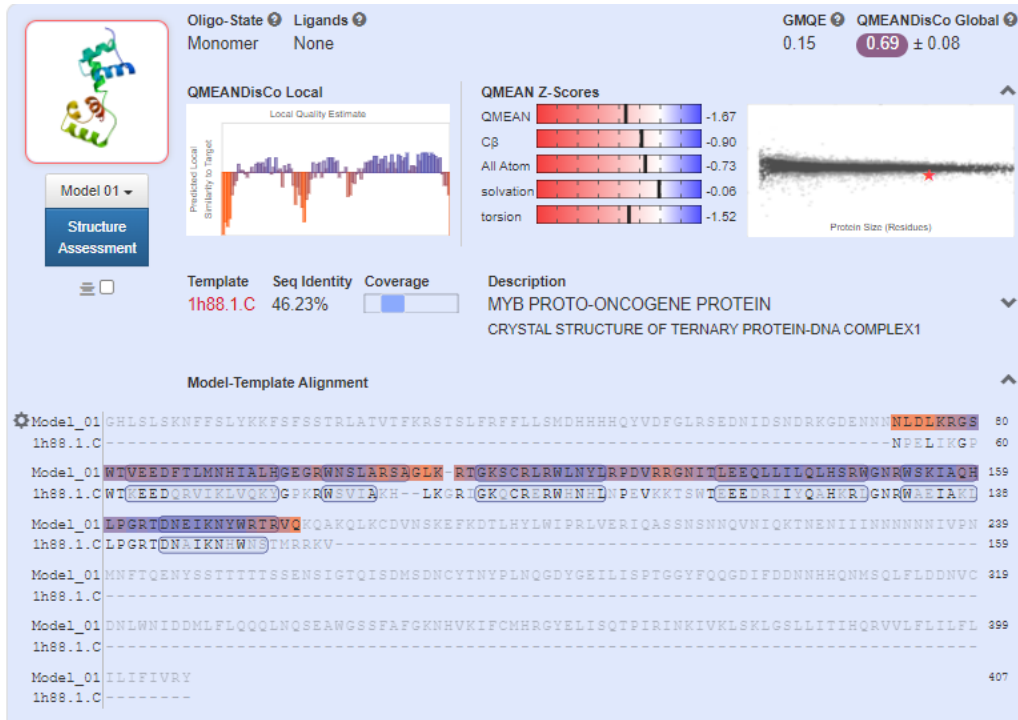


Şekil 4. 12. Kromozom 11 üzerinde haritalanan >MK810 markörünün üç boyutlu protein modeli.

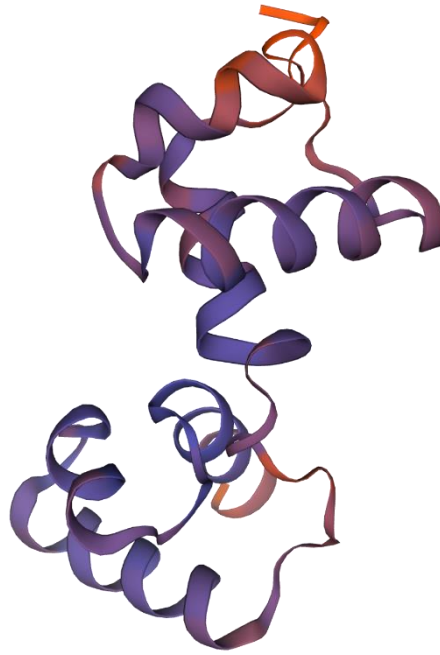
a. Protein modelinin sayısal parametreleri ve Model-Şablon dizi hizalaması

b. Üç boyutlu model

a.



b.



Şekil 4. 13. Kromozom 12 üzerinde haritalanan >MK810 markörünün üç boyutlu protein modeli.
a. Protein modelinin sayısal parametreleri ve Model-Şablon dizi hizalaması
b. Üç boyutlu model

Kromozomlar üzerinde haritalanan ve polimorfizm potansiyeli e-Mapping analizleri ile gösterilen polimorfik markörlerin, üç boyutlu protein modelleri, sayısal parametreleri, model şablon hizalama değerleri incelendiğinde, dizayn edilen markör lokusların kodladığı proteinlerin üç boyutlu model şablonlarının neredeyse tamamının transkripsiyon faktörü kodlayan genlere (WRKY, MYB, NAC vb.) denk düştüğü gözlenmiştir.

Domates kromozomlarında haritalandığında türler-arası polimorfizm potansiyeli görülen markörlerden bazılarında motif tekrar sayısı modifikasyonları yapılarak tekrar modellenmiş ve model değerlerindeki değişimler Çizelge 4.9’da verilmiştir.

Çizelge 4. 12. Seçilen markörlerin sekanslarında yapılan modifikasyon sonrası modelleme değerlerindeki değişimler.

| Alel | Markör | Şablon | Tanım | QMEAN-Z | GMQE | QMEANDisCo |
|--------|--------|----------|----------------------------------------------------------------------|---------|------|------------|
| Alel-1 | MK-90 | 2ayd.1.A | WRKY transcription factor 1 | -1,17 | 0,12 | 0,67 |
| Alel-2 | | | | -1,25 | 0,12 | 0,68 |
| Alel-1 | MK-345 | 6j9b.1.A | B3 domain-containing transcription factor FUS3 | -1,36 | 0,30 | 0,77 |
| Alel-2 | | | | -1,59 | 0,27 | 0,67 |
| Alel-1 | MK-708 | 5gnj.1.A | "Transcription factor MYC2 | -1,22 | 0,03 | 0,65 |
| Alel-2 | | | | -1,26 | 0,03 | 0,64 |
| Alel-1 | MK-657 | 1gcc.1.C | Ethylene Responsive Element Binding Factor 1 | -1,03 | 0,11 | 0,58 |
| Alel-2 | | | | -1,26 | 0,10 | 0,60 |
| Alel-1 | MK-166 | 1ut7.1.A | No apical meristem protein structure of the conserved domain of ANAC | -2,65 | 0,15 | 0,54 |
| Alel-2 | | | | -2,41 | 0,15 | 0,54 |
| Alel-1 | MK-752 | 5ly0.1.A | LOB family transfactor Ramosa2.1 | 0,59 | 0,12 | 0,66 |
| Alel-2 | | | | 0,54 | 0,12 | 0,66 |
| Alel-1 | MK-775 | 2ayd.1.A | WRKY transcription factor 1 | -2,22 | 0,12 | 0,48 |
| Alel-2 | | | | -2,18 | 0,13 | 0,49 |
| Alel-1 | MK-417 | 4m59.1A | LOB family transfactor Ramosa2.1 | -2,45 | 0,40 | 0,63 |
| Alel-2 | | | | -3,77 | 0,28 | 0,62 |
| Alel-1 | MK-304 | 5gnj.3.A | Transcription factor MYC2 | 0,26 | 0,04 | 0,53 |
| Alel-2 | | | | 0,26 | 0,04 | 0,53 |
| Alel-1 | MK-376 | 3a01.1.A | Homeodomain-containing protein | -1,35 | 0,16 | 0,63 |
| Alel-2 | | | | -3,16 | 0,16 | 0,69 |

| | | | | | | |
|---------|-----|----------|-------------------------------------------------------|-------|------|------|
| Al-el-1 | MK- | 6kks.1.A | MYB-3 | -0,13 | 0,12 | 0,69 |
| Al-el-2 | 810 | | | -2,75 | 0,16 | 0,62 |
| Al-el-1 | MK- | 1h88.1.C | MYB PROTO- | -1,67 | 0,15 | 0,69 |
| Al-el-2 | 730 | | ONCOGENE PROTEIN | -1,67 | 0,15 | 0,69 |
| Al-el-1 | MK- | 7p8k.1.A | | -1,69 | 0,09 | 0,62 |
| Al-el-2 | 402 | | Disease resistance protein RRS1 | -5,5 | 0,04 | 0,27 |
| Al-el-1 | MK- | 5v3m.1.C | Zinc finger protein 568 | -4,68 | 0,08 | 0,39 |
| Al-el-2 | 699 | | | -1,06 | 0,01 | 0,46 |
| Al-el-1 | MK- | 7d8t.1.C | | 0,31 | 0,04 | 0,52 |
| Al-el-2 | 529 | | Microphthalmia- associated transcription factor | 0,42 | 0,04 | 0,43 |
| Al-el-1 | MK- | 5xd6.1.A | | -2,58 | 0,27 | 0,63 |
| Al-el-2 | 344 | | Protein kinase superfamily protein | -3,12 | 0,26 | 0,61 |
| Al-el-1 | MK- | 6cth.1.A | Protein kinase | -2,58 | 0,28 | 0,63 |
| Al-el-2 | 344 | | superfamily protein | -3,12 | 0,26 | 0,63 |

Yapılan modifikasyonlar sonucunda protein işlev tanımında bir değişim gözlenmemiştir, ancak QMEAN-Z, QMQE ve QMEANglobaldisco değerlerinde gözlenen değişimler post-translasyonel modifikasyonlar, kalsiyum bağlama bölgeleri, ATP bağlama bölgeleri, zing finger, homebox, SH2 ve SH3 alanları vb. ve 3D yapıda polimorfizme bağlı değişimlere neden olabildiği modelleme analiz sonuçlarına dayalı olarak ortaya konmuştur. Bu durum, çalışma kapsamında geliştirilen markörlerin tekrar sayısı değişimi kaynaklı alellerinin transkripsiyon faktörlerinde fonksiyonel modifikasyonlara sebep olabileceğini ve dolayısı ile geliştirilen markörlerin fenotip-genotip ilişkilerinin oluşturulmasında kullanılacak fonksiyonel markörler olma potansiyelini göstermektedir.

5. SONUÇLAR VE ÖNERİLER

5.1. Sonuçlar

Patates (*Solanum tuberosum*) tüm dünyada özellikle de gelişmekte olan ülkelerde temel enerji kaynağı olan, nişasta açısından zengin ekonomik değeri yüksek bir bitkidir. Patates geniş tarımsal arazilerde yüksek verimle yetiştirilebilir. Ancak ıslah çalışmaları, tetraploid genomu ve yüksek heterozigosite seviyeleri nedeniyle geride kalmıştır. Otopoliploidi tarafından oluşturulan birçok genetik karmaşıklık nedeniyle, patates genetik çalışmalarının çoğu, çift ebeveynli popülasyonlar kullanılarak diploid düzeyde gerçekleştirilmiştir.

Patates genom sekansı ve ilgili genetik haritaların varlığına rağmen, birçok önemli tarımsal ve ekonomik açıdan önemli patates özelliğini etkileyen genler henüz keşfedilmemiştir (Sharma ve ark., 2013; Patates Genom Dizileme Konsorsiyumu 2011).

Önemli bir faktör de birçok özelliğin karmaşık genetik konumunun yanı sıra QTL çalışmaları için oldukça heterozigot ebeveyn hatlarının kullanılması nedeniyle birçok QTL özelliğinin genetik konumlarının kesin olmayan bir şekilde belirlenmesidir. Ayrıca, biparental çaprazlamalarda faydalı özellik alellerine sıkı bir şekilde bağlı olduğu bulunan markörler, genellikle daha geniş germplazm havuzlarına sınırlı aktarılabilirliğe sahiptir (Sharma ve ark., 2018).

SSR markörleri koodominant yapıları, genomlardaki bollukları, yüksek oranda tekrarlanabilirlikler, hiperpolimorfizmleri ve ilgili türlere yüksek düzeyde aktarılabilirlik gibi avantajlara sahip olmaları nedeniyle ıslah çalışmaları için oldukça önemlidir. SSR markörleri genetik haritalama, ebeveyn tayini genetik çeşitlilik analizleri, popülasyonların araştırılması, gen kaynaklarının korunması çalışmaları gibi çalışmalarda kullanılabilir. QTL çalışmalarında daha güvenilir sonuçlar vermesi ve istenilen gen bölgelerinin belirlenip seçilmesini sağlaması açısından tüm araştırmaların yanı sıra Patates gibi otopoliploid genomların araştırılması için de oldukça önemlidir.

Sekanslama teknolojilerinin hız kazanması ve daha ekonomik hale gelmesi, genetik bağlantı haritalarının oluşturulması, QTL analizi MAS gibi genetik araştırmalar için geliştirilmiş algoritmalar ve yazılımların geliştirilmesi, genel olarak hem moleküler çalışmalara hem de patates gibi otopoliploid mahsul türlerinin ıslah çalışmalarına son zamanlarda önemli ilerlemeler sağlamıştır. Özellikle biyoinformatik gelişmeler, açık-

erişim veri tabanlarından elde edilen verilerle, aday genlerin içinde veya yakınında bulunan fonksiyonel markör geliştirmeyi mümkün kılmıştır.

Bu çalışmada cDNA sekanslarından oluşan patates (*Solanum tuberosum*) transkriptomunda transkripsiyon faktörü homologlarını belirlemek üzere, Arabidopsis ve pirinç transkripsiyon faktörü gen setleri veritabanı olarak kullanılarak BLASTx analizleri gerçekleştirilmiştir. Patateste transkripsiyon faktörü kodlayan gen setlerinin homologları belirlenmiştir. Yapmış olduğumuz her iki Blastx analizinde eşleşmeler için elde edilen e-değerleri “0” veya 0’a oldukça yakın değerlerdir ve bu da Blastx analizi ile bulunan eşleşmelerin “tam eşleşme” olduğunu, istatistiksel olarak “rastgele” olma olasılığının 0 veya 0’a yakın olduğunu göstermiştir.

Blastx analizleri ile Arabidopsis transkripsiyon faktörü peptid sekansları ve patates cDNA sekanslarının eşleştirilmesi sonucunda 5089 benzersiz cDNA’nın 1984 Arabidopsis transkripsiyon faktörü homoloğu olduğu bulunmuştur.

Pirinç transkripsiyon faktörü peptid sekanslarının, patates cDNA sekansları ile homoloji analizleri sonucunda 4379 benzersiz cDNA’nın 1791 benzersiz pirinç transkripsiyon faktörünün homoloğu olduğu bulunmuştur.

Her iki homoloji analizi sonucunda ortaya çıkan toplam 9468 cDNA sekansı, Arabidopsis ve pirinç ortologlarından ötürü iki veri seti arasında yinelenen cDNA sekansları filtrelenerek, her bir cDNA sekansı birer kez temsil edilecek şekilde 6052 adet cDNA sekansına sadeleştirilmiştir. Her cDNA’ya denk düşen TF erişim numaralarının hangi transkripsiyon faktörü ailesinin üyesi olduğu plantTFDB’den alınan peptid dosyaları ile belirlenmiştir. Transkripsiyon faktörlerinin patates transkriptomunda dağılımına baktığımızda, en fazla temsil edilen transkripsiyon faktörü gen ailesinin 1392 gen ile WRKY TF ailesi olduğu tespit edilmiştir (Grafik 4.1.).

TF kodladığını belirlediğimiz sekanslar, GMATA yazılımı kullanılarak analiz edilmiş, toplam 1435 SSR lokusu tespit edilip, bu lokusları çoğaltacak 845 SSR markörü geliştirilmiştir (Çizelge 4.10). Markörlerin türler-arası transfer edilebilirlik ve polimorfizm potansiyellerini araştırmak için geliştirilen markörler hem patates hem de yakın akraba tür olan domates kromozomlarında haritalanmıştır. Haritalama analizlerinde kullanılan parametreler dahilinde patates kromozomlarında 535 domates kromozomlarında ise 99 markör haritalanmıştır (Çizelge 4.8.)

Solanum tuberosum transkriptomunda geliştirilen yeni 845 SSR markör havuzundan seçilen bazı markörler türler-arası amplifikasyon ve polimorfizm için deneysel olarak da test edilmiş, kapiler elektroforez sistemi ile elde edilen analiz

sonuçları, primerlerin beklenen boyutlarda amplikonları başarıyla ürettiğini (e-Mapping sonuçlarından ± 50 bp), in silico markör dizayn ve transfer edilebilirlik analizlerinde yüksek verimin alınabileceğini kanıtlanmıştır (Şekil 4.1.).

Solanum tuberosum kromozomlarında haritalanan ve polimorfizm potansiyeli türler arası 'Mapping' analizi ile *Solanum lycopersicum*' da belirlenen markörlerin peptid sekansı düzeyinde anotasyonları ve 3D (üç boyutlu) modelleme analizleri Swiss Prot server'ı kullanılarak gerçekleştirilmiştir. Türler-arası transfer edilebilirliği gösterilen 99 markör için protein modelleme analizi yapılmış tamamının transkripsiyon faktörü ailesinden olduğu, transkripsiyon mekanizmasıyla evrimsel olarak ilişkili şablon proteinler ve işlevleri ve QMEAN değerleri belirlenmiştir (Çizelge 4.11).

Polimorfik olduğu belirlenen 18 markörün peptid sekanslarında modifikasyonlar yapılarak tekrar protein modelleme analizi yapılmış ve modellere ait 3D yapıda ve QMEAN değerlerindeki değişimler karşılaştırılmıştır. Bu karşılaştırma sonucunda beklendiği üzere protein anotasyonları değişmemekle birlikte (homoloji şablonları aynı kalmıştır), QMEAN değerlerindeki değişimin hedefin hizalanması bilgisi, şablonların 3D koordinatları ile birlikte şablon dizileri, ilgilenilen protein için yapısal modellerde değişimler olabileceği gözlenmiştir (Çizelge 4.12). Her kromozomdan seçilen markör lokuslar için yapılan protein modelleme analizleri sonucunda elde edilen üç boyutlu protein modelleri, protein modelinin sayısal parametreleri ve model-şablon dizi hizalama bilgileri elde edilmiştir (Şekil 4.2.- Şekil 4.3.).

5.2. Öneriler

Yapılan çalışmaya genel olarak bakıldığında 6052 cDNA sekansından transkripsiyon faktörü kodlayan toplam 845 SSR markörü geliştirilmiş ve türler-arası aktarılabilir markörler de belirlenmiştir. Markörler, Transkripsiyon faktörü kodlayan gen bölgelerindeki polimorfizmi genotiplememizi sağlayacağı için bitki ıslahında seçim markörü olarak kullanılabilir. Genler arası bir bölgeden geliştirilen SSR'lar ile kıyaslandığında kodlayan bölgelerden geliştirilen SSR'lar fenotipik olarak canlının nasıl davrandığını, nasıl yaşadığını, neye nasıl tepki verdiğini değiştirme olasılığı daha yüksektir bu sebeple fonksiyonel markörlerdir. Genetik çeşitlilik, ıslah, ya da belli bir özellik ile ilgili özelliği kontrol eden bir gen seti için geliştirdiğimiz markör seti daha yüksek olasılıkla ilgili bir markör tespit etmemizi sağlayacaktır

Transkripsiyon faktörü gen aileleri bakımından karakterize etmek ve bu genler üzerinden regülatör fonksiyona sahip olma potansiyeli yüksek DNA markörleri geliştirmek, patatesin polimorfizm seviyesi ve çeşitliliği hakkında bilgiler sağlamak, patates yetiştirme programlarını hızlandırmak için önemli ve güvenilir bir kaynak olacaktır. Patates bitkisinde karakterize edilen transkripsiyon faktörü genlerinde geliştirilen SSR markörleri ile yapılacak çalışmalar, bitkilerde çeşitli biyotik ve abiyotik streslere yanıt verebilecek ve değişen bir ortamda optimum büyümeyi sürdürebilecek daha verimli genotipleri seçmeyi kolaylaştıracaktır ve ıslah çalışmaları için oldukça önemli olacaktır. TF'ler tarafından kontrol edildiği bilinen sayısız hücresel süreç üzerinde de çalışmalara yarar sağlayacağı düşünülmektedir.

6. KAYNAKLAR

- Afgan, E., Baker, D., Batut, B., Van Den Beek, M., Bouvier, D., Čech, M., ... & Blankenberg, D. 2018, The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic acids research*, 46(W1), W537-W544.
- Agarwal, M., Shrivastava, N., & Padh, H. 2008, Advances in molecular marker techniques and their applications in plant sciences. *Plant cell reports*, 27(4), 617-631
- Aggarwal, P., Gupta, M. Das, Joseph, A. P., Chatterjee, N., Sinivasan, N. ve Nath, U. 2010, Identification of specific DNA binding residues in the TCP family of transcription factors in arabidopsis. *Plant Cell*, 22(4), 1174–1189. doi:10.1105/tpc.109.066647
- 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.
- Alzohairy, A. M., Gyulai, G., Ohm, H., Szabó, Z., Ragheb, S. M., Ali, M. A., ... & Bahieldin, A. 2015, Nuclear and Organelle Specific PCR Markers. *LAP Lambert Academic Publishing*, 53-78.
- Ambawat, S., Sharma, P., Yadav, N. R. ve Yadav, R. C. 2013, MYB transcription factor genes as regulators for plant responses: An overview. *Physiology and Molecular Biology of Plants*, 19(3), 307–321. doi:10.1007/s12298-013-0179-1
- Barone, A. 2004, Molecular marker-assisted selection for potato breeding. *American Journal of Potato Research*, 81(2), 111-117.
- Biswas, M. K., Nath, U. K., Howlader, J., Bagchi, M., Natarajan, S., Kayum, M. A., ... & Nou, I. S. 2018, Exploration and exploitation of novel SSR markers for candidate transcription factor genes in *Lilium* species. *Genes*, 9(2), 97.
- Botstein D, White RL, Skolnick M, et al. 1980, Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet* 32(3):314–331
- Calliope, S. R., Lobo, M. O., & Sammán, N. C. 2018, Biodiversity of Andean potatoes: Morphological, nutritional and functional characterization. *Food chemistry*, 238, 42-50.
- Cao, Y., Li, K., Li, Y., Zhao, X. ve Wang, L. 2020, MYB transcription factors as regulators of secondary metabolism in plants. *Biology*, 9(3), 1–16. doi:10.3390/biology9030061
- Chen, F., Hu, Y., Vannozzi, A., Wu, K., Cai, H., Qin, Y., ... Zhang, L. 2017, The WRKY Transcription Factor Family in Model Plants and Crops. *Critical Reviews in Plant Sciences*, 36(5–6), 311–335. doi:10.1080/07352689.2018.1441103

- Cheng, J., Zhao, Z., Li, B., Qin, C., Wu, Z., Trejo-Saavedra, D. L., ... & Hu, K. 2016, A comprehensive characterization of simple sequence repeats in pepper genomes provides valuable resources for marker development in *Capsicum*. *Scientific reports*, 6(1), 1-12.
- D'hoop, B. B., Keizer, P. L., Paulo, M. J., Visser, R. G., van Eeuwijk, F. A., & van Eck, H. J. 2014, Identification of agronomically important QTL in tetraploid potato cultivars using a marker–trait association analysis. *Theoretical and applied genetics*, 127(3), 731-748.
- Danisman, S. 2016, TCP transcription factors at the interface between environmental challenges and the plant's growth responses. *Frontiers in Plant Science*, 7(DECEMBER2016), 1–13. doi:10.3389/fpls.2016.01930
- Dennis, D. J., Han, S. ve Schuurmans, C. 2019, bHLH transcription factors in neural development, disease, and reprogramming. *Brain Research*, 1705, 48–65. doi:10.1016/j.brainres.2018.03.013
- Finatto, T., Viana, V. E., Woyann, L. G., Busanello, C., da Maia, L. C. ve de Oliveira, A. C. 2018, Can WRKY transcription factors help plants to overcome environmental challenges? *Genetics and Molecular Biology*, 41(3), 533–544. doi:10.1590/1678-4685-gmb-2017-0232
- Gebhardt, C. 1994, RFLP mapping in potato of qualitative and quantitative genetic loci conferring resistance to potato pathogens. *American Potato Journal*, 71(5), 339-345.
- Gebhardt, C. H., Ritter, E., Barone, A., Debener, T., Walkemeier, B., Schachtschabel, U., ... & Salamini, F. 1991, RFLP maps of potato and their alignment with the homoeologous tomato genome. *Theoretical and Applied Genetics*, 83(1), 49-57.
- Gebhardt, C., Blomendahl, C., Schachtschabel, U., Debener, T., Salamini, F., & Ritter, E. 1989a, Identification of 2n breeding lines and 4n varieties of potato (*Solanum tuberosum*, ssp. *tuberosum*) with RFLP-fingerprints. *Theoretical and applied genetics*, 78(1), 16-22.
- Gebhardt, C., Ritter, E., & Salamini, F. 2001, RFLP map of the potato. In *DNA-based markers in plants* (pp. 319-336). Springer, Dordrecht.
- Gebhardt, C., Ritter, E., Debener, T., Schachtschabel, U., Walkemeier, B., Uhrig, H., & Salamini, F. 1989b, RFLP analysis and linkage mapping in *Solanum tuberosum*. *Theoretical and Applied Genetics*, 78(1), 65-75.
- Guedes Corrêa, L. G., Riaño-Pachón, D. M., Guerra Schrago, C., Vicentini dos Santos, R., Mueller-Roeber, B. ve Vincentz, M. 2008, The role of bZIP transcription factors in green plant evolution: Adaptive features emerging from four founder genes. *PLoS ONE*, 3(8). doi:10.1371/journal.pone.0002944
- Guo, M., Liu, J. H., Ma, X., Luo, D. X., Gong, Z. H. ve Lu, M. H. 2016, The plant heat stress transcription factors (HSFS): Structure, regulation, and function in response

- to abiotic stresses. *Frontiers in Plant Science*, 7(FEB2016). doi:10.3389/fpls.2016.00114
- Hardigan, M. A., Laimbeer, F. P. E., Newton, L., Crisovan, E., Hamilton, J. P., Vaillancourt, B., ... & Buell, C. R. 2017, Genome diversity of tuber-bearing *Solanum* uncovers complex evolutionary history and targets of domestication in the cultivated potato. *Proceedings of the National Academy of Sciences*, 114(46), E9999-E10008
- Hartung, F., & Schiemann, J. 2014, Precise plant breeding using new genome editing techniques: opportunities, safety and regulation in the EU. *The Plant Journal*, 78(5), 742-752
- Heim, M. A., Jakoby, M., Werber, M., Martin, C., Weisshaar, B. ve Bailey, P. C. 2003, The basic helix-loop-helix transcription factor family in plants: A genome-wide study of protein structure and functional diversity. *Molecular Biology and Evolution*, 20(5), 735–747. doi:10.1093/molbev/msg088
- Heyman, J., Canher, B., Bisht, A., Christiaens, F. ve De Veylder, L. 2018, Emerging role of the plant ERF transcription factors in coordinating wound defense responses and repair. *Journal of Cell Science*, 131(2). doi:10.1242/jcs.208215
- Hirsch, C. D., Hamilton, J. P., Childs, K. L., Cepela, J., Crisovan, E., Vaillancourt, B., ... & Buell, C. R. 2014, Spud DB: A resource for mining sequences, genotypes, and phenotypes to accelerate potato breeding. *The Plant Genome*, 7(1), plantgenome2013-12.
- Hong, J. C. 2016, *General Aspects of Plant Transcription Factor Families. Plant Transcription Factors: Evolutionary, Structural and Functional Aspects*. Elsevier Inc. doi:10.1016/B978-0-12-800854-6.00003-8
- Jakoby, M., Weisshaar, B., Dröge-Laser, W., Vicente-Carbajosa, J., Tiedemann, J., Kroj, T. ve Parcy, F. 2002, bZIP transcription factors in *Arabidopsis*. *Trends in Plant Science*, 7(3), 106–111. doi:10.1016/S1360-1385(01)02223-3
- Kalia, R. K., Rai, M. K., Kalia, S., Singh, R., & Dhawan, A. K. 2011, Microsatellite markörs: an overview of the recent progress in plants. *Euphytica*, 177(3), 309-334.
- Kardolus, J. P., van Eck, H. J., & van den Berg, R. G. 1998, The potential of AFLPs in biosystematics: a first application in *Solanum* taxonomy (Solanaceae). *Plant Systematics and Evolution*, 210(1), 87-103.
- Katiyar, A., Smita, S., Lenka, S. K., Rajwanshi, R., Chinnusamy, V. ve Bansal, K. C. 2012, Genome-wide classification and expression analysis of MYB transcription factor families in rice and *Arabidopsis*. *BMC Genomics*, 13(1). doi:10.1186/1471-2164-13-544
- Krishnamurthy, S. ve Hampsey, M. 2009, Eukaryotic transcription initiation. *Current Biology*, 19(4), 153–156. doi:10.1016/j.cub.2008.11.052

- Kujur, A. L. I. C. E., Bajaj, D. E. E. P. A. K., Saxena, M. S., Tripathi, S. H. A. I. L. E. S. H., Upadhyaya, H. D., Gowda, C. L. L., ... & Parida, S. K. 2013, Functionally relevant microsatellite markers from chickpea transcription factor genes for efficient genotyping applications and trait association mapping. *DNA research*, 20(4), 355-374.
- Li, C., Ng, C. K. Y. ve Fan, L. M. 2015, MYB transcription factors; active players in abiotic stress signaling. *Environmental and Experimental Botany*, 114, 80–91. doi:10.1016/j.envexpbot.2014.06.014
- Licausi, F., Ohme-Takagi, M. ve Perata, P. 2013, APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: mediators of stress responses and developmental programs. *The New phytologist*, 199(3), 639–649. doi:10.1111/nph.12291
- Litrico, I., & Violle, C. 2015, Diversity in plant breeding: a new conceptual framework. *Trends in plant science*, 20(10), 604-613.
- Liu, M. M., Wang, M. M., Yang, J., Wen, J., Guo, P. C., Wu, Y. W., ... Du, H. 2019, Evolutionary and comparative expression analyses of TCP transcription factor gene family in land plants. *International Journal of Molecular Sciences*, 20(14), 1–24. doi:10.3390/ijms20143591
- Love, S. L., & Pavek, J. J. 2008, Positioning the potato as a primary food source of vitamin C. *American Journal of Potato Research*, 85(4), 277-285.
- Machida-Hirano, R., & Niino, T. 2017, Potato genetic resources. In *The potato genome*, Springer, Cham. 11-30 Springer, Cham.
- McGinnis, S., & Madden, T. L. 2004, BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic acids research*, 32(suppl_2), W20-W25.
- Metzgar, D., Bytof, J., & Wills, C. 2000, Selection against frameshift mutations limits microsatellite expansion in coding DNA. *Genome research*, 10(1), 72-80.
- Mitsuda, N., & Ohme-Takagi, M. 2009, Functional analysis of transcription factors in Arabidopsis. *Plant and Cell Physiology*, 50(7), 1232-1248
- Mizoi, J., Shinozaki, K. ve Yamaguchi-Shinozaki, K. 2012, AP2/ERF family transcription factors in plant abiotic stress responses. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1819(2), 86–96. doi:10.1016/j.bbagr.2011.08.004
- Mohanta, T. K., Park, Y. H. ve Bae, H. 2016, Novel Genomic and Evolutionary Insight of WRKY Transcription Factors in Plant Lineage. *Scientific Reports*, 6(October), 1–22. doi:10.1038/srep37309
- Monte, M. N., Rey Burusco, M. F., Carboni, M. F., Castellote, M. A., Sucar, S., Norero, N. S., ... & Feingold, S. E. 2018, Genetic diversity in Argentine Andean potatoes

- by means of functional markers. *American Journal of Potato Research*, 95(3), 286-30
- Morgante, M., Hanafey, M., & Powell, W. (2002). Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature genetics*, 30(2), 194-200
- Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., & Erlich, H. 1986, January Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. In *Cold Spring Harbor symposia on quantitative biology* (Vol. 51, pp. 263-273). Cold Spring Harbor Laboratory Press.
- Murre, C., McCaw, P. S., Vaessin, H., Caudy, M., Jan, L. Y., Jan, Y. N., ... Baltimore, D. 1989, Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell*, 58(3), 537–544. doi:10.1016/0092-8674(89)90434-0
- Nadeem, M. A., Nawaz, M. A., Shahid, M. Q., Doğan, Y., Comertpay, G., Yıldız, M., ... & Baloch, F. S. 2018, DNA molecular markers in plant breeding: current status and recent advancements in genomic selection and genome editing. *Biotechnology & Biotechnological Equipment*, 32(2), 261-285.
- Nakashima, K., Takasaki, H., Mizoi, J., Shinozaki, K. ve Yamaguchi-Shinozaki, K. 2012, NAC transcription factors in plant abiotic stress responses. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1819(2), 97–103. doi:10.1016/j.bbagr.2011.10.005
- Navarre, D. A., Goyer, A., & Shakya, R. 2009, Nutritional value of potatoes: vitamin, phytonutrient, and mineral content. In *Advances in potato chemistry and technology* Academic Press, 395-424.
- Nei, M. 1973, Analysis of gene diversity in subdivided populations. *Proceedings of the national academy of sciences*, 70(12), 3321-3323.
- Ohama, N., Kusakabe, K., Mizoi, J., Zhao, H., Kidokoro, S., Koizumi, S., ... Yamaguchi-Shinozaki, K. 2016, The transcriptional cascade in the heat stress response of arabidopsis is strictly regulated at the level of transcription factor expression. *Plant Cell*, 28(1), 181–201. doi:10.1105/tpc.15.00435
- Ohama, N., Sato, H., Shinozaki, K. ve Yamaguchi-Shinozaki, K. 2017, Transcriptional Regulatory Network of Plant Heat Stress Response. *Trends in Plant Science*, 22(1), 53–65. doi:10.1016/j.tplants.2016.08.015
- Ooka, H., Satoh, K., Doi, K., Nagata, T., Otomo, Y., Murakami, K., ... Kikuchi, S. 2003, Comprehensive Analysis of NAC Family Genes in *Oryza sativa* and *Arabidopsis thaliana*. *DNA Research*, 10(6), 239–247. doi:10.1093/dnares/10.6.239
- Ortiz, O., & Mares, V. 2017, The historical, social, and economic importance of the potato crop. In *The Potato Genome*, Springer, Cham pp. 1-10

- Ovchinnikova A, et al. 2011, Taxonomy of cultivated potatoes (*Solanum* section *Petota*: *Solanaceae*). *Bot J Linn Soc* 165:107–155
- Parmar, R., Seth, R., & SHARMA, R. K. 2020, Genome-wide identification and characterization of functionally relevant microsatellites from putative transcription factor genes of tea for efficient genotyping.
- Paz-Ares, J., Ghosal, D., Wienand, U., Peterson, P. A. ve Saedler, H. 1987, The regulatory *c1* locus of *Zea mays* encodes a protein with homology to myb proto-oncogene products and with structural similarities to transcriptional activators. *The EMBO journal*, 6(12), 3553–3558. doi:10.1002/j.1460-2075.1987.tb02684.x
- Pirooznia, M., Perkins, E. J., & Deng, Y. 2008, Batch Blast Extractor: an automated blastx parser application. *BMC genomics*, 9(2), 1-5.
- Potato Genome Sequencing Consortium, 2011, Genome sequence and analysis of the tuber crop potato. *Nature*, 475.
- Puranik, S., Sahu, P. P., Srivastava, P. S. ve Prasad, M. 2012, NAC proteins: Regulation and role in stress tolerance. *Trends in Plant Science*, 17(6), 369–381. doi:10.1016/j.tplants.2012.02.004
- Rajendrakumar, P. 2015, Molecular Markör Development Using Bioinformatic Tools. In *Sorghum Molecular Breeding* (pp. 179-195). Springer, New Delhi.
- Riechmann, J. L., Heard, J., Martin, G., Reuber, L., Jiang, C. Z., Keddie, J., ... Yu, G. L. 2000, Arabidopsis transcription factors: Genome-wide comparative analysis among eukaryotes. *Science*, 290(5499), 2105–2110. doi:10.1126/science.290.5499.2105
- Senan, S., Kizhakayil, D., SASIKUMAR, B., & SHEEJA, T. E. 2014, Methods for development of microsatellite markers: an overview. *Notulae Scientia Biologicae*, 6(1), 1-13.
- Scharf, K. D., Berberich, T., Ebersberger, I. ve Nover, L. 2012, The plant heat stress transcription factor (Hsf) family: Structure, function and evolution. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1819(2), 104–119. doi:10.1016/j.bbagr.2011.10.002
- Schwechheimer, C., Zourelidou, M. ve Bevan, M. W. 1998, Plant transcription factor studies. *Annual Review of Plant Biology*, 49, 127–150. doi:10.1146/annurev.arplant.49.1.127
- Shandilya, J. ve Roberts, S. G. E. 2012, The transcription cycle in eukaryotes: From productive initiation to RNA polymerase II recycling. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1819(5), 391–400. doi:10.1016/j.bbagr.2012.01.010

- Sharma, H., Bhandawat, A., Kumar, P., Rahim, M. S., Parveen, A., Kumar, P., ... & Roy, J. 2020, Development and characterization of bZIP transcription factor based SSRs in wheat. *Gene*, 756, 144912.
- Sharma, S. K., & Bryan, G. J. 2017, Genome sequence-based marker development and genotyping in potato. In *The Potato Genome*, Springer, Cham. pp. 307-326
- Sharma, S. K., Bolser, D., de Boer, J., Sønderkær, M., Amoros, W., Carboni, M. F., ... & Bryan, G. J. 2013, Construction of reference chromosome-scale pseudomolecules for potato: integrating the potato genome with genetic and physical maps. *G3: Genes, Genomes, Genetics*, 3(11), 2031-2047.
- Sharma, S. K., MacKenzie, K., McLean, K., Dale, F., Daniels, S., & Bryan, G. J. 2018, Linkage disequilibrium and evaluation of genome-wide association mapping models in tetraploid potato. *G3: Genes, Genomes, Genetics*, 8(10), 3185-3202.
- Spooner DM, et al. 2007, Extensive simple sequence repeat genotyping of potato landraces supports a major reevaluation of their gene pool structure and classification. *Proc Natl Acad Sci USA* 104:19398–19403
- Spooner, D. M., Ghislain, M., Simon, R., Jansky, S. H., & Gavrilenko, T. 2014, Systematics, diversity, genetics, and evolution of wild and cultivated potatoes. *The botanical review*, 80(4), 283-383.
- Spooner, D., Jansky, S., Clausen, A., Herrera, M. D. R., & Ghislain, M. 2012, The enigma of *Solanum maglia* in the origin of the Chilean cultivated potato, *Solanum tuberosum chilotanum* group1. *Economic botany*, 66(1), 12-21.
- Taheri, S., Lee Abdullah, T., Yusop, M. R., Hanafi, M. M., Sahebi, M., Azizi, P., & Shamshiri, R. R. 2018, Mining and development of novel SSR markörs using next generation sequencing (NGS) data in plants. *Molecules*, 23(2), 399
- Tanksley, S. D., Ganal, M. W., Prince, J. P., De Vicente, M. C., Bonierbale, M. W., Broun, P., ... & Martin, G. 1992, High density molecular linkage maps of the tomato and potato genomes. *Genetics*, 132(4), 1141-1160.
- Uncu, A. O., & Uncu, A. T. 2020, High-throughput simple sequence repeat (SSR) mining saturates the carrot (*Daucus carota* L.) genome with chromosome-anchored markers. *Biotechnology & Biotechnological Equipment*, 34(1), 1-9.
- Uncu, A. T. 2019, Genome-wide identification of simple sequence repeat (SSR) markers in *Capsicum chinense* Jacq. with high potential for use in pepper introgression breeding. *Biologia*, 74(2), 119-126.
- Vieira, M. L. C., Santini, L., Diniz, A. L., & Munhoz, C. D. F. 2016, Microsatellite markers: what they mean and why they are so useful. *Genetics and molecular biology*, 39(3), 312-328.

- Vinson, C., Acharya, A. ve Taparowsky, E. J. 2006, Deciphering B-ZIP transcription factor interactions in vitro and in vivo. *Biochimica et Biophysica Acta - Gene Structure and Expression*, 1759(1–2), 4–12. doi:10.1016/j.bbaexp.2005.12.005
- Von Koskull-Döring, P., Scharf, K. D., & Nover, L. 2007, The diversity of plant heat stress transcription factors. *Trends in plant science*, 12(10), 452-457.
- Yamasaki, K., Kigawa, T., Seki, M., Shinozaki, K. ve Yokoyama, S. 2013, DNA-binding domains of plant-specific transcription factors: Structure, function, and evolution. *Trends in Plant Science*, 18(5), 267–276. doi:10.1016/j.tplants.2012.09.001
- Zhang, Z., Liu, W., Qi, X., Liu, Z., Xie, W., & Wang, Y. 2015, Genome-wide identification, expression profiling, and SSR marker development of the bZIP transcription factor family in *Medicago truncatula*. *Biochemical Systematics and Ecology*, 61, 218-228.
- Zhao, Z., Guo, C., Sutharzan, S., Li, P., Echt, C. S., Zhang, J., & Liang, C. 2014, Genome-wide analysis of tandem repeats in plants and green algae. *G3: Genes, Genomes, Genetics*, 4(1), 67-78.
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., ... & Schwede, T. 2018, SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research*, 46(W1), W296-W303.
- Wang, Y., & Liu, A. 2020, Genomic characterization and expression analysis of basic helix-loop-helix (bHLH) family genes in traditional Chinese herb *Dendrobium officinale*. *Plants*, 9(8), 1044.
- Wang, X., & Wang, L. 2016, GMATA: an integrated software package for genome-scale SSR mining, marker development and viewing. *Frontiers in plant science*, 7, 1350.
- Welner, D. H., Deeba, F., Leggio, L. L., & Skriver, K. (2016). NAC transcription factors: from structure to function in stress-associated networks. In *Plant transcription factors* (pp. 199-212). Academic Press.
- Wetterstrand KA. DNA Sekanslama Maliyetleri: NHGRI Genom Sekanslama Programından(GSP)elde edilen veriler: www.genome.gov/sequencingcostsdata . Erişildi [23.07.2020].
- Williams, C. E., Wielgus, S. M., Haberlach, G. T., Guenther, C., Kim-Lee, H., & Helgeson, J. P. 1993, RFLP analysis of chromosomal segregation in progeny from an interspecific hexaploid somatic hybrid between *Solanum brevidens* and *Solanum tuberosum*. *Genetics*, 135(4), 1167-1173.