



T.C.
NECMETTİN ERBAKAN
ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ



PATLİCAN (*Solanum melongena* L.)
BİTKİSİNDE REGÜLATÖR RNA'LARIN ve
TRANSPOZON LOKUSLARIN
KARAKTERİZASYONU

Dilay MALYEMEZ

YÜKSEK LİSANS TEZİ
Moleküler Biyoloji ve Genetik 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.

Dilay MALYEMEZ
Tarih:

ÖZET

YÜKSEK LİSANS TEZİ

PATLİCAN (*Solanum melongena* L.) BİTKİSİNDE REGÜLATÖR RNA'LARIN ve TRANSPOZON LOKUSLARIN KARAKTERİZASYONU

Dilay MALYEMEZ

**Necmettin Erbakan Üniversitesi Fen Bilimleri Enstitüsü
Moleküler Biyoloji ve Genetik Anabilim Dalı**

Danışman: Doç. Dr. Ali Tefvik UNCU

2022, 48+xi Sayfa

Jüri

Doç. Dr. Ali Tefvik UNCU

Prof. Dr. Önder TÜRKMEN

Prof. Dr. Emrah TORLAK

Patlıcan *Solanaceae* ailesinin en önemli üyelerinden, besleyicilik özellikleri ve ekonomik değeri yüksek bir bitki türü olmasına rağmen genetik ıslah, moleküler genetik çalışmalar bakımından domates ve patates, biber gibi ailenin diğer önemli türlerine kıyasla daha az çalışılmıştır. Sunulan tez kapsamında patlıcan genomuna ait referans genom verisi transpozon lokusları, kodlamayan RNA ve transmembran protein sınıflandırması bakımından biyoinformatik yaklaşımlar ile karakterize edilmiştir. Çalışma sonucunda elde edilen kodlamayan RNA sekansları, transpozon sekansları ve patlıcanın var olan EST sekansı verileri regülatör markör geliştirme potansiyelleri bakımından araştırılmıştır. Çalışma kapsamında ilk kez patlıcan genomuna spesifik regülatör/fonksiyonel olma potansiyeline sahip tüm genom düzeyinde mikrosatellit markörler geliştirilmiş, yapısal ve fonksiyonları anotasyonları yapılmıştır. Çalışma sonucu üretilen veri patlıcan gen kaynaklarının karakterizasyonu başta olmak üzere patlıcanda markör destekli ıslah çalışmalarına katkı sağlayacaktır.

Anahtar Kelimeler: mikrosatellit, markör geliştirme, ncRNA, patlıcan, transpozonlar

ABSTRACT

MS THESIS

Characterization of Regulatory RNAs and Transposon Loci in Eggplant (*Solanum melongena* L.)

Dilay MALYEMEZ

**THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCE OF
NECMETTİN ERBAKAN UNIVERSITY
THE DEGREE OF MASTER OF SCIENCE
IN MOLECULAR BIOLOGY AND GENETICS**

Advisor: Doç. Dr. Ali Tevfik UNCU

2022, 48+xi Pages

Jury

**Doç. Dr. Ali Tevfik UNCU
Prof. Dr. Önder TÜRKMEN
Prof. Dr. Emrah TORLAK**

Although eggplant is one of the most important members of the *Solanaceae* family and is a plant species with high nutritional properties and economic value, genetic breeding has been less studied in terms of molecular genetic studies compared to other important species of the family such as tomato, potato and pepper. Within the scope of the presented thesis, the reference genome of the eggplant genome was featured by bioinformatics approaches in terms of transposon elements, non-coding RNA and transmembrane protein classification. The non-coding RNA sequences, transposon element sequences and the existing EST sequence data of eggplant obtained as a result of the study were investigated in terms of their regulatory marker development potential. Within the scope of my work, for the first time, microsatellite markers at the whole genome level, which have the potential to be regulatory/functional specific to the eggplant genome, were developed and their structural and functional annotations were made. The data produced as a result of the study will contribute to the characterization of eggplant gene sources and marker-assisted breeding studies in eggplant.

Keywords: eggplant, microsatellite, marker development, non-coding RNA, transposon elements

ÖNSÖZ

Tez çalışmamızın planlanmasında, araştırılmasında, yürütülmesinde ve oluşumunda ilgi ve desteğini esirgemeyen, engin bilgi ve tecrübelerinden yararlandığım, yönlendirme ve bilgilendirmeleriyle çalışmamı bilimsel temeller ışığında şekillendiren sayın Doç. Dr. Ali Tevfik UNCU ‘ya, çalışmalarımız boyunca bizden yardım ve desteklerini esirgemeyen çok kıymetli hocalarım Doç. Dr. Ayşe Özgür UNCU ‘ya ve Arş. Gör. Ayşenur Soytürk PATAT’a maddi ve manevi desteklerinden dolayı teşekkür ederim.

Tüm bu süreçte bana eşlik eden, güzellikleri ve zorlukları birlikte paylaştığımız takım arkadaşlarım Cansu YÜRÜKÇÜ, Şerife Nur TOSUN ve Fatıma ŞEN’e, desteklerinden, katkılarından ve işbirliğinden dolayı teşekkür ederim.

Son olarak hayatımın her alanında benim yanımda olan ve aldığım kararları destekleyen, beni cesaretlendiren ve moral veren sevgili aileme ve kardeşlerime her zaman yanımda oldukları için teşekkür ederim.

Dilay MALYEMEZ
KONYA-2022

İÇİNDEKİLER

ÖZET	iv
ABSTRACT.....	v
ÖNSÖZ	vi
İÇİNDEKİLER	vii
SİMGELER VE KISALTMALAR	ix
ŞEKİL DİZİNİ	x
TABLO DİZİNİ	xi
1. GİRİŞ	1
2. KAYNAK ARAŞTIRMASI	2
2.1. Patlıcan (<i>Solanum melongena</i> L.)	2
2.1.1. Patlıcanın Yetiştği Bölgeler	3
2.1.2. Patlıcanın Genetik Yapısı	3
2.2. Kodlamayan DNA	3
2.2.1. Transpozonlar	4
2.2.2. Kodlamayan RNA'lar	9
2.3. Transmembran Proteinleri	12
2.4. Sinyal Peptidleri.....	14
2.5. Moleküler Markörler.....	15
2.5.1. Mikrosatellit Markörler (SSR).....	16
2.5.2. Mikrosatellit Markörlerin Tanımlanması ve Genotiplenmesi	18
3.METERYAL METOD	21
3.1. Veri Analizi.....	21
3.1.1. Transpozonların Karakterize Edilmesi	21
3.1.2. Kodlamayan RNA'ların Karakterize Edilmesi	21
3.1.3. Transmembran Proteinlerinin Karakterize Edilmesi	22
3.1.4. Mikrosatellit Markörlerin Geliştirilmesi.....	23
3.1.5. Geliştirilen Markörlerin Fonksiyonlarının Belirlenmesi	24
4. ARAŞTIRMA SONUÇLARI VE TARTIŞMA.....	26
4.1. SONUÇLAR.....	26
4.1.2. Transpozonların Tespit Edilmesi	26
4.1.3. Kodlamayan RNA'ların Tespit Edilmesi.....	28

4.1.4. Transmembran Proteinlerin Tanımlanması	33
4.1.5. Mikrosatellit Markörlerin Geliştirilmesi	34
4.1.6. Geliştirilen Markörlerin Fonksiyonlarının Belirlenmesi	39
5. SONUÇLAR VE ÖNERİLER	47
KAYNAKLAR	49

SİMGELER VE KISALTMALAR

Kısaltmalar

DNA: Deoksiribo Nükleik Asit
RNA: Ribo Nükleik Asit
PCR: Polimeraz Zincir Reaksiyonu
WGS: Tüm Genom Dizileme
NGS: Yeni Nesil Dizileme
EST: İfade Edilmiş Dizi Etiketleri
TE: Transposable Element
NOR: Nükleolar Organizasyon Bölgeleri
LTR: Uzun Terminal Sekansları
GAG: Gruba Özgü Antijen
Pol: Polimeraz
PR: Proteaz
INT: İntegraz
RT: Ters Transkriptaz
ORF: Açık Okuma Çerçevesi
UTR: Çevrilmemiş Bölgeler
SINE: Kısa Serpiştirilmiş Nükleer Elementler
LINE: Uzun Serpiştirilmiş Nükleer Elementler
TIR: Terminal Ters Çevrilmiş Tekrarlar
MITE: Minyatür Ters Çevrilmiş Tekrarlar
ncRNA: Kodlamayan RNA
MiRNA: Mikro RNA
LncRNA: Uzun Kodlamayan RNA
TM: Transmembran
SP: Sinyal Peptid
RAPD: Rastgele Çoğaltılmış Polimorfik DNA
RFLP: Restriksiyon Fragmanı Uzunluk Polimorfizmi
AFLP: Amplifiye Edilmiş Fragman Uzunluk Polimorfizmi
SSR: Basit Tekrar Dizileri
SNP: Tek Nükleotit Polimorfizmi
QTL: Kantitatif Özelik Lokus
MAS: Makör Destekli Seçilim
GWAS: Tüm Genom İlişkilendirme Seçilim
AGE: Agaroza Jel Elektrophorezi
PAGE: Poliakrilamid Jel Elektrophorezi
BP: Biyolojik Süreç
CC: Hücresel Bileşen
MF: Moleküler Fonksiyon

ŞEKİL DİZİNİ

Şekil 2.1. Transpozonlar.....	7
Şekil 2.2. Kodlamayan RNA 'ların sınıflandırılması.....	10
Şekil 2.3. MikroRNA biyogenezi.....	11
Şekil 2.4. Transmembran protein çeşitleri.....	14
Şekil 2.5. SSR örneği.....	18
Şekil 2.6. SSR markörlerinin elektroforez görüntüsüne örnek.....	20
Şekil 2.7. Kapiler elektroforez SSR sonuçlarına örnek	20
Şekil 4.1. Karakterize edilen transpozonların oranları.....	28
Şekil 4.2. Patlıcan genomunda kodlanan ve kodlamayan RNA 'ların dağılımı.....	29
Şekil 4.3. Transpozonlarda geliştirilen SSR motiflerinin dağılımı.....	35
Şekil 4.4. Transpozonlarda geliştirilen SSR motiflerinden en çok bulunan markörlerin yüzdelik dağılım grafiği.....	35
Şekil 4.5. Kodlamayan RNA 'larda geliştirilen SSR motiflerinin dağılım grafiği.....	36
Şekil 4.6. Geliştirilen SSR motiflerinden bulunan markörlerin yüzdelik dağılım grafiği.....	37
Şekil 4.7. EST sekanslarında geliştirilen SSR motiflerinin grafik üzerinde dağılımı.....	38
Şekil 4.8. Geliştirilen SSR motiflerinden markörlerin yüzdelik dağılım grafiği.....	39
Şekil 4.9. Kodlamayan RNA gen anotasyon dağılımı biyolojik süreç (BP), hücresel bileşen (CC) ,moleküler fonksiyon (MF). Yeşil, sarı ve mavi sırasıyla BP, CC ve MF kategorilerini temsil eder	41
Şekil 4.10. Biyolojik süreç kısmında tanımlanan proteinlerin dağılımı.....	42
Şekil 4.11. Tanımlanan proteinleri moleküler fonksiyon dağılımı.....	42
Şekil 4.12. Tanımlanan proteinlerin hücresel dağılımı.....	43
Şekil 4.13. EST bölgelerinin gen anotasyon dağılımı biyolojik süreç (BP), hücresel bileşen (CC) moleküler fonksiyon (MF). Yeşil, sarı ve mavi sırasıyla BP, CC ve MF kategorilerini temsil eder.....	44
Şekil 4.14. Biyolojik süreç kısmında tanımlanan proteinlerin dağılımı.....	45
Şekil 4.15. Tanımlanan proteinleri moleküler fonksiyon dağılımı.....	45
Şekil 4.16. Tanımlanan proteinlerin hücresel bileşenlerin dağılımı.....	46

TABLO DİZİNİ

Tablo 3.1. PLEK programı için kullanılan parametreler.....	22
Tablo 3.2. Transmembran proteinleri parametreleri.....	23
Tablo 4.2. Karakterize edilen kodlamayan RNA'ların lokasyonları.....	29
Tablo 4.3. Patlıcanda analiz edilen abiyotik stres ile ilişkili miRNA'lar.....	31
Tablo 4.4. Transmembran proteinleri ve sinyal peptidleri.....	33
Tablo 4.5. Transpozonlarda geliştirilen SSR motiflerinin tekrar nükleotit sayısı.....	34
Tablo 4.6. Kodlamayan RNA bölgelerinde geliştirilen SSR motiflerinin tekrar nükleotit sayısı.....	36
Tablo 4.7. <i>Solanum melongena</i> EST sekanslarından geliştirilen SSR motiflerinin nükleotit sayısı.....	38
Tablo 4.8. Omicsbox programında analiz edilen kodlamayan RNA markörlerinin rastgele 10 tanesinin protein sonuçları.....	41
Tablo 4.9. Omicsbox programında analiz edilen EST markörlerinin rastgele 10 tanesinin protein sonuçları.....	44

1. GİRİŞ

Solanaceae ailesinin bir üyesi olan patlıcan (*Solanum melongena* L.), dünyanın tropikal ve ılıman bölgelerinde yetişen önemli bir sebze mahsulüdür (Sharma ve Katoch 2020). Genellikle brinjal veya patlıcan olarak adlandırılan patlıcan (*Solanum melongena*) Solunum cinsinden domates patates ve biberden sonra dördüncü en önemli sebze mahsulüdür (Hirakawa ve ark. 2014). Domates ve biber gibi patlıcan da 12 kromozomlu diploid bir bitkidir ($2n=24$). Genomu 1100 Mb boyutlarındadır. Patlıcan, aynı ailede bulunan domates başta olmak üzere biber ve patatese kıyasla hem tohum ıslahı da hem de moleküler biyoloji ve genetik alanında çok fazla çalışılmamıştır. Bu sebeple günümüzde patlıcana ait moleküler veriler sınırlıdır. Bunun nedeni olarak patlıcanın özellikle batı ülkelerinde daha az üretilmesi ve tüketilmesinden kaynaklı, ekonomik değerinin düşük olması gösterilebilir.

Özellikle son yirmi yılda yeni nesil sekanslama teknolojilerindeki (NGS) gelişmeler daha az çalışılmış türlere ait biyolojik verinin hızla üretilmesine yol açmıştır. NGS teknolojilerindeki bu gelişmeler ile birlikte tüm genom dizileme (WGS) ve transkriptom uygulamalarında ki ilerlemeler, moleküler bilgilerin üretimini hızlandırmış ve markör destekli sebze ıslahı alanında önemli bir ilerlemeye bir yol açmıştır (Arumuganathan 1991). Patlıcan genomunun da sekanslanması ile birlikte, genomik çalışmalar artmış olup üretilen verilerin ıslah çalışmalarına adaptasyonu hızlanmıştır.

Tez çalışması kapsamında; Patlıcan genomu kodlamayan RNA'lar, transpozon içeriği ve regülatör elementler bakımından biyoinformatik araçlar ile analiz edilmiş, daha sonra patlıcan genomuna özel regülatör sekansların içinden ve/veya çevre lokuslardan SSR markörleri geliştirilerek yapısal ve fonksiyonel anotasyonları yapılmıştır. Patlıcan tüm genom verisi yüksek oranda korunan ve gen ekspresyonlarını düzenleyen kodlamayan RNA'ların, genom çapında çeşitliliği ve dağılımı ile farklılık sağlayan transpozonların ve transmembran proteinlerin karakterizasyonu için hem *ab initio* hem de homoloji tabanlı yaklaşımlar ile incelenmiştir. Üretilen veriler patlıcan gibi genetik çeşitlilik bakımından dar bir havuza sahip gen kaynaklarının daha gerçekçi genetik uzaklık ölçümlerini mümkün kılacak, aynı zamanda üretilen veri patlıcanda önemli agronomik özelliklerin, biyotik ve abiyotik stres toleransında etkili gen/QTL lokusların yüksek çözünürlükte haritalanmasına imkan sağlayacaktır. Patlıcanda MAS çalışmaları literatüre önemli veri katkısı sağlayacaktır.

2. KAYNAK ARAŞTIRMASI

2.1. Patlıcan (*Solanum melongena* L.)

Solanaceae ailesinin bir üyesi olan patlıcan (*Solanum melongena* L.), dünyanın tropikal ve ılıman bölgelerinde yetişen önemli bir sebze ürünüdür. Patlıcan, patates, domates ve biberden sonra ailenin ekonomik açıdan en önemli dördüncü tarım mahsulüdür. Patlıcan, brinjal olarak da bilinmektedir (Ansari ve Singh 2014). Patlıcan, yumurta şeklinde (*S. melongena* var. *esculentum*); uzun ince şekilli (*S. melongena* var. *serpentium*) ve bodur tip (*S. melongena* var. *depressum*) olmak üzere üç ana çeşide ayrılmıştır (Kalloo 1993).

Demir (0,9 mg), kalsiyum (18 mg), kükürt (44 mg), potasyum, vitaminler ve proteinler için iyi bir kaynaktır (Noda ve ark. 1998). Beyaz brinjal ile karşılaştırıldığında, koyu mor olanların C vitamini açısından zengin olduğu bildirilmektedir (Tabing ve Tiwari, 2018). Domates (*S. lycopersicum*) ve biber gibi (*Capsicum annuum*), patlıcan 12 kromozomlu ($2n = 24$) otogam bir diploiddir. Genomu 1100 Mb boyutlarındadır. Genellikle "düşük kalorili sebze", olarak kabul edilse de patlıcan besin değeri diğer yaygın sebzelerin çoğuyla karşılaştırılabilir. Vitaminler, mineraller, proteinler, lif ve ayrıca çoğu antioksidan aktiviteye sahip olan fenolik bileşikler ve flavonoidler gibi önemli bitkisel besinler içerir (Raigona ve ark. 2008). İyi bir mineral ve vitamin kaynağıdır ve toplam besin değeri açısından domates ile karşılaştırılabilir (Kalloo 1993; Singh ve Kumar 2007). Ayrıca geleneksel tıpta patlıcan kullanılmıştır. Doku özleri astım, bronşit, kolera ve dizüri tedavisinde kullanılır; meyve ve yapraklar kan kolesterolünü düşürmede faydalıdır. Patlıcanın da antimutajenik özelliklere sahip olduğu gösterilmiştir (Khan 1979; Hinata 1986; Kalloo 1993; Collonnier ve ark. 2001; Kaşyap ve ark. 2003; Rajam ve Kumar 2006). Diğer solanli ürünler gibi patlıcan, insanlar üzerinde potansiyel toksik etki gösterebilen solasonine ve solamargin glikoalkaloidleri içerir. Aynı zamanda anti kanser özellikleri de göstermiştir; özellikle, cilt neoplazileri patlıcandan türetilen glikoalkaloidlere dayalı ticari kremlerle verimli bir şekilde tedavi edilir. Ayrıca ekonomik açıdan önemli birçok gen taşıyan birçok yabani patlıcan türü vardır. Yabani patlıcan türleri ticari patlıcan üretimini etkileyen önemli zararlılara karşı direnç göstermektedir.

Ekonomik olarak önemli olmasına rağmen, patlıcanın moleküler genetiği, bağlantı haritalarına sahip olan domates, patates ve biberin aksine çok az çalışılmıştır.

2.1.1. Patlıcanın Yetiştigi Bölgeler

Patlıcan domates, patates, biber ve tütün gibi diğer önemli *Solanaceae* bitkilerinden farklı olarak; Asya temelli bir bitkidir ve dünya çapında yetiştirilmektedir. Patlıcanın yabani tür *Solanum incanum*'dan evcilleştirildiği varsayılmıştır (Ranil ve ark. 2017; Taher ve ark. 2017). Birçok ülkede önemli bir sebze ürünüdür. Önde gelen patlıcan üreticileri Çin (dünya üretiminin%56'sı), Hindistan (dünya üretiminin%26'sı), Mısır ve Türkiye'dir (Van Der Weerden ve Barendse 2007).Popülaritesi nedeniyle 16. yüzyılda Avrupa'ya ve 1806'da Amerika'ya tanıtılmıştır; bu yerlerden ekimi dünya çapında hızla yayılmıştır. Muhtemelen Hindistan temelli bir bitki olduğu düşünülmektedir. İkinci merkezi ise Çin olarak bilinmektedir(Kalloo 1993).Bu ürün özellikle Asya'da (dünya üretiminin hem de hasat alanının %93'ü) yetiştirilmektedir. Bu mahsulün iyi verim vermesi için uzun süre ılık havaya ihtiyacı vardır; bu nedenle tropikal ve subtropikal iklimler ekimi için en uygun olarak kabul edilir. Ancak ılıman ve karasal bölgelerde hem açık alanda hem de sezon dışı üretim için koruma altında yetiştirilmektedir (Kumar ve ark. 2020) .

2.1.2. Patlıcanın Genetik Yapısı

Domates ve biber gibi patlıcan da 12 kromozomlu diploid bir bitkidir. Patlıcanın nükleer genomu domatesinkinden biraz daha büyüktür ve 1100 Mb DNA içerir (Arumugana than ve Earle 1991). Geleneksel olarak, solanli ürünlerdeki araştırmaların çoğu, genomları yayınlanmış olan patates ve domates üzerine odaklanmıştır. Patlıcan SME_r2.5.1'in referans genomu ise 2014 yılında Hirakawa ve arkadaşları tarafından yayınlanmıştır (Barchi ve ark. 2019).

2.2. Kodlamayan DNA

Milyonlarca yıllık evrim boyunca, bitki genomları karmaşık bir yapı kazanmıştır. Bitki genomları, en büyük ve en küçük taksonlar arasında 1000 kat farkla geniş bir boyut aralığına sahiptir (Schnable ve ark. 2009). Nispeten, bitki genomlarının en iyi karakterize edilmiş elemanı olan protein kodlayan genlerin sayısında taksonlar arasında sadece 10-100 katlık bir varyasyon görülür (Hou ve ark. 2019). Protein kodlayan genler, bitki

genomlarında eşit olmayan bir şekilde dağılmıştır ve gen yoğunlukları farklılık gösterir. Protein kodlayan genlerle serpiştirilmiş, hepsi çok daha az çalışılmış olan çok sayıda başka genomik elementler vardır. Bunlar, protein kodlayan genlerin intronlarını ve kodlamayan RNA, tekrar dizileri ve transpozonlar gibi intergenik dizileri içerir. Bu kodlamayan diziler, gen ekspresyonunu düzenlemeye yardımcı olur ve genom organizasyonu ve işlevinde önemli bir rol oynadığı giderek daha fazla kabul edilmektedir (Hou et al. 2019) .

2.2.1. Transpozonlar

Canlı organizmaların genomlarında birden fazla yerde bulunan dizilere "tekrarlayan diziler" denir. Tekrarlayan diziler (DNA tekrarları) terimi, DNA fragmanlarını ifade eder. Bu diziler, farklı yapı ve kökenlerden oluşan farklı DNA parçalarıdır. Tekrarlanan diziler "Tandem diziler" ve "Transpozonlar" olmak üzere iki sınıfa ayrılır. Tandem dizileri sıralı, benzer dizilerdir. Yer değiştiren (transposable) elementler (TE) veya transpozonlar, genom içinde yer değiştiren ve mutasyonlara neden olan ve genomdaki DNA miktarını değiştiren DNA dizileridir. Genomda birden fazla kopya halinde bulunurlar (Arvas ve ark.). Yer değiştirebilir elemanlar McClintock' un benekli mısır analizinden beri tanınmaktadır. Barbara McClintock'ın yapmış olduğu bu keşif 1983'te Nobel Ödülü'ne layık görülmesini sağladı. Başlangıçta önemsiz olduğu düşünülen transpozonlar üzerinde çok az çalışma yapılmıştır. Nadir olduğu düşünülse de genom dizileme projelerinin yaygınlaşması, memeli ve bitki genlerinin çoğunun intronlarında transpozon eklemeleri barındırdığını ortaya çıkardı ve bu durum ile düşünüldüğü kadar önemsiz olmadığı anlaşıldı (Naito 2006). Transpozonlar, genomdaki transpozisyon yoluyla konumlarını değiştirebilirler ve bu nedenle de 'sıçrayan genler' olarak da adlandırılırlar (Yepuri ve ark. 2020). Her transpozisyon olayı, mutasyonlar yaratarak ve bir hücrenin genom boyutunu değiştirerek yeni değişkenlik üretir (Yadav ve ark. 2015). Transpozonlar duplikasyon, delesyon, insersiyon gibi kalıcı genomik mutasyonlara neden olduğu için döllenme ile sonraki nesillere aktarılır (Arvas ve ark. 2021).

Transpozonlar hem prokaryotik hem de ökaryotik genomlar arasında yaygın olarak bulunmaktadır (Fattash ve ark. 2013). Transpozonlar, ökaryotik genom içeriğinin %85'ini oluşturmaktadır (Flynn ve ark. 2020). Bitki genomlarının da önemli bir

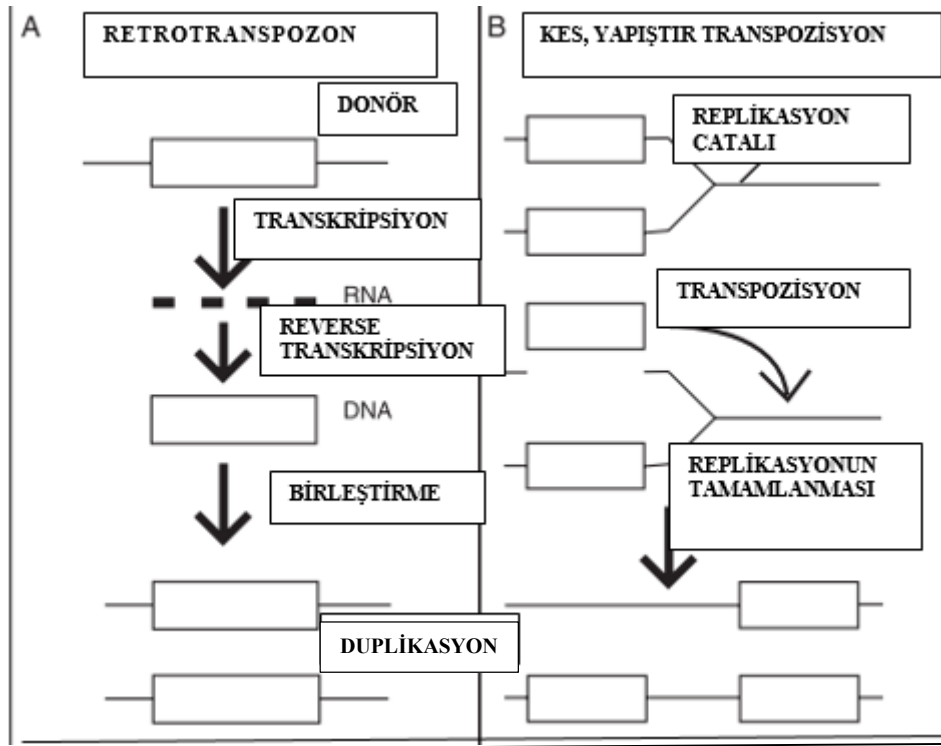
bölümünde transpozonlar yer almaktadır (Yadav ve ark. 2015). Transpozonlar mutasyon, replikasyon ve transpozisyon yöntemleri sayesinde bitki genomlarının yeni polimorfizmleri yaratırlar (Yepuri ve ark. 2020). Transpozonlar, bitki genomunun ana bileşenlerinden biridir ve fonksiyonel genom çeşitliliği ve fenotipik varyasyonlarda çok önemli bir rol oynadıkları bildirilmektedir. Bu nedenle, genomdaki transpozonların organizasyonunu analiz etmek, araştırmacıların transpozonların yakınındaki genlerin ekspresyonunu incelemesine olanak tanıyacaktır. Ayrıca transpozonların bitki fenotipik çeşitliliğini şekillendirmesindeki rolünü anlamayı sağlayacaktır (Yadav ve ark. 2015). Transpozisyon mekanizmalarına dayalı olarak, transpozonlar geleneksel olarak iki sınıfa ayrılır; sınıf I transpozonlar ve sınıf II transpozonlar (Finnegan 1989). Sınıf I transpozonlar, RNA ara maddeleri aracılığıyla "kopyala ve yapıştır" şekilde hareket eden retrotranspozonlardır. Sınıf II transpozonlar, DNA ara maddeleri yoluyla transpoze olan DNA transpozonlarıdır (Neha Samir Roy 2014).

2.2.1.1. Sınıf I Transpozonlar (Retrotranspozonlar)

Sınıf I transpozonlar veya retrotranspozonlar, DNA'nın ana ipliğinden RNA ara maddeleri yoluyla kopyalayarak ve yeni kopyaları yeni lokuslara yapıştırarak yer değiştirirler (Fattash ve ark. 2013) (Şekil 2.1). Genellikle, transpozonlar ve özellikle retrotranspozonlar sentromerik ve subsentromerik olarak oldukça zenginleştirilmiş ve nükleolar organizasyon bölgelerinde (NOR'ler) bulunur (Klein ve O'Neill 2018). LTR (Uzun Terminal Sekansları) retrotranspozonları ve LTR olmayan retrotranspozonlar olmak üzere iki ana alt sınıfa ayrılır (Yadav ve ark. 2015). Uzun terminal tekrar (LTR) retrotranspozonları; gruba özgü antijen (GAG) ve polimeraz (pol) kodlama bölgelerinden oluşur (Kidwell 2005). GAG, replikasyon için bir proteini kodlar ve pol, proteaz (PR), integras (INT), ters transkriptaz (RT) ve RNaz H (RH) içeren çoklu proteini kodlar (Roy ve ark. 2014). Bu elemanlar, her iki uçta promotör kabiliyetine sahip LTR'ler ile çevrilidir (Kidwell 2005). Transpozisyon sırasında bu elementler tarafından RNA şablonundan çift sarmallı bir DNA ara maddesi sentezlenir ve yeni kopya yeni lokusa entegre olur (Craig ve ark. 2002). Özelliklerine göre, LTR-retrotranspozonlar başlıca Ty1-copia ve Ty3-gypsy olarak iki süper aileye ayrılır (S. Li ve ark. 2020). Ty1/copia ve Ty3/gypsy hayvanlarda ve bitkilerde bol miktarda bulunur ve kapsamlı bir şekilde karakterize edilmiştir (Kidwell 2005). LTR retrotranspozonu yüksek kopya sayısı oluşumu

sağladığından, bu retrotranspozonlar bitki genomlarında bulunan en bol mobil elementlerdir (Yadav ve ark. 2015). Gerçekten de bazı bitkilerde LTR retrotranspozonları genomun %90'ına kadarını temsil eder. Örneğin; Mısır genomunun >%50'sini, *Arabidopsis thaliana* genomunun %14'ünü ve buğday genomunun %90'ına kadarını oluştururlar (Yadav ve ark. 2015). Diğer bitkilere benzer şekilde, çim genomları da kendilerini genomda RNA aracılı bir retrotranspozisyon süreci ile çoğaltan retrotranspozonlar açısından zengindir.

LTR olmayan retrotranspozonları transpozisyon sırasında ters transkriptaz (RT) kullanan TE'lerin ikinci ana alt sınıfıdır. Tam uzunlukta, otonom LTR olmayan retrotranspozonlar tipik olarak bir veya iki açık okuma çerçevesi (ORF'ler) içerir (Han 2010). Kısa olan ORF, nükleik asit bağlayıcı bir protein kodlar ve uzun olan ORF endonükleaz ve ters transkriptaz aktivitelerine sahip bir proteini kodlar (Neha Samir Roy 2014). Retrotranspozon hareketinin merkezi ters transkriptaz (RT) aktivitesidir ve dolayısıyla tüm otonom LTR olmayan retrotranspozonlar bir RT alanı içerir (Han 2010). LTR olmayan retrotranspozonların 5' ve 3' UTR'ler oldukça değişkendir. Fonksiyonel LTR olmayan retrotranspozonların çoğu karakterize edilmiş 5' UTR'leri promotör aktivitesi içermesine rağmen, promotör değişimi siktir ve bu nedenle promotör dizisi korunmaz. Yapılan çalışmalar, bu promotörlerin RNA polimeraz II tarafından kopyalandığını göstermektedir. LTR olmayan retrotranspozonların UTR'lerindeki geniş değişkenlik nedeniyle, LTR olmayan transkripsiyon başlatma ve sonlandırmanın belirli ayrıntıları hakkında kapsamlı genellemeler yapmak yanlış olabilir (Han 2010). Bu retrotranspozonlar LTR'lerden yoksundur ve RNA transkriptlerinin tamamlayıcı bir DNA (cDNA) kopyasını doğrudan kromozomal hedef bölgeye ters kopyalayarak transpoze ederler (Kidwell 2005). Kısa serpiştirilmiş nükleer elementler (SINE'ler), uzun serpiştirilmiş nükleer elementler (LINE'lar) LTR olmayan retrotranspozonlara örnek olarak verilebilir (Kajikawa ve Okada, 2002). LTR olmayan retrotranspozonlar çoğu ökaryotik genomda bulunur. İnsanlar gibi bazı türlerde, bu elementler en bol bulunan genom dizisidir ve bu güne kadar çoğalmaya devam ederek bir endojen mutasyon ve potansiyel stres kaynağı yaratırlar (Han 2010).



Şekil 2.1. Transpozonlar.

2.2.1.2. Sınıf II Transpozonlar (DNA Transpozonları)

DNA transpozonları, bitkilerde ilk tanımlanan transpozonlardır. Temel olarak hem germinal hem de somatik dokularda çok yüksek frekanslarda değiştirilmiş gen veya genom fenotiplerine yol açarlar (Bennetzen 2000). DNA transpozonları olarak bilinen sınıf II elementler, DNA ipliğinde bulunan transpozonun, tek iplikli veya çift iplikli DNA'sını keser ve yeni lokuslara yapıştırır (Fattash ve ark.2013) (Şekil 1). Bir RNA ara molekülü olmadan genomun başka bir lokasyonuna transpozon sekansını yapıştırırlar (Rebollo ve ark.2012). Transpozisyon sırasında kesilen DNA ipliklerinin sayısına bağlı olarak DNA transpozonları iki alt sınıfa ayrılırlar (Classification 2019).

DNA transpozonları alt sınıf I, terminal ters çevrilmiş tekrar (TIR) transpozonları ve alt sınıf II, TIR olmayan transpozonlar olarak sınıflanmışlardır. (Wicker ve ark. 2007; Schulman ve Wicker 2013). Alt sınıf I TIR DNA transpozonları, çeşitli uzunluklarda TIR'larla çevrilidirler (Lee ve Kim 2014). TIR transpozonları "kes ve yapıştır" mekanizmasıyla aktarılır. TIR'lar, transpozonun kendisi tarafından kodlanan bir transpozaz tarafından çıkarılarak elde edilen bir geriye katlamalı yapı oluşturur. Özellikle transpozaz bağlanma bölgeleri, transpozonun yer değiştirmesi için tek gerekli dizilerdir.

Kesilen çift sarmallı DNA, genomun başka bir yerine yeniden yerleştirilir. Bu şekilde genomun içinde yer değiştirmiş olurlar. TIR transpozonları LTR-retrotranspozonlar kadar yüksek kopya sayılarına ulaşmaz (Roy ve ark. 2014).

Minyatür ters çevrilmiş tekrarlı transpoze edilebilir elementler (MITE'ler) TIR transpozonlarıdır (Neha Samir Roy 2014). MITE'ler, yaklaşık 100 farklı ailede gruplandırılmıştır. 90.000'den fazla MITE içeren pirinçte olduğu gibi, çeşitli bitki genomlarında yüksek bir kopya sayısına ulaşmışlardır (Feschotte ve ark. 2003; Jiang ve ark. 2004; Juretic ve ark. 2004). MITE'ler genellikle gen ekspresyonunu etkileyebilecekleri genlerin yakınında veya içinde bulunur (Santiago ve ark. 2002; Yang ve ark. 2005; Ohmori ve ark. 2008; Piriyapongse ve Jordan 2008; Kuang ve ark. 2009).

Alt sınıf II, TIR içermeyen transpozonlar klasik DNA transpozonlarından farklı olarak, çift sarmallı bölünme olmaksızın replikasyon gerektiren bir transpozisyon sürecinden geçen DNA transpozonlarını içerir. Bu kopyala ve yapıştır transpozonları, yalnızca bir ipliğin yer değiştirmesini içeren replikasyon yoluyla yer değiştirir (Wicker ve ark. 2007). Alt sınıf II TIR içermeyen DNA transpozonları, Helitrons ve Maverick adı verilen iki süper aileden oluşur. Helitronlar yaygın olarak bazı bitkilerde bol miktarda bulunur, ancak terminal tekrar yapısının olmaması nedeniyle tanımlanmaları zordur. Helitronlar bir yuvarlanan daire mekanizması yoluyla çoğalırlar. Helitronlar çok yüksek kopya sayılarına ulaşabilir ve genomik çeşitlilik yaratabilir (Lee ve Kim 2014).

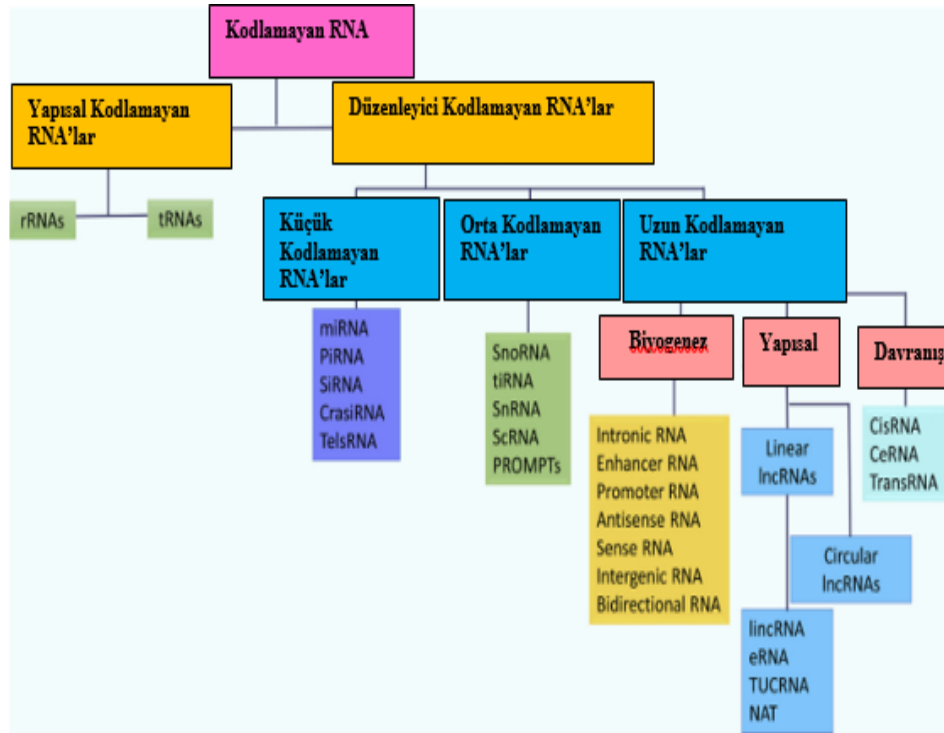
Transpozonların hayvan ve bitki genomlarındaki çok sayıda bulunması, yüksek oranda insersiyon polimorfizmi ve tespit kolaylığı gibi yararlı özellikleri göz önünde bulundurulduğunda moleküler markör sistemi için avantajlı olabilirler (Bhat ve ark. 2020). Transpozonlara dayalı moleküler markörler, moleküler genetik çeşitliliği yakalamada daha verimli olabilir. Bu elementler amplifikasyon patlamaları ile genom genişlemesi ve transpozisyonlar nedeniyle kromozomal bölgelerin yeniden oluşmasını sağlarlar. Bu şekilde canlıda yeni aleller oluşturabilirler. Bu sebeple de evrimsel çalışmalar için önem arz etmektedirler (Venkatesh ve Nandini 2020). Birçok evrimsel çalışmada ve filogenetik çalışmada kullanılmak üzere transpozonlardan moleküler markörler geliştirilmiştir.

2.2.2. Kodlamayan RNA'lar

Kodlamayan RNA'lar (ncRNA) veya proteinlere çevrilmeyen RNA'lar, ökaryotik genomlarda evrensel olarak gözlenir. Bitki genomlarında çok çeşitli RNA türleri tanımlanmış ve karakterize edilmiştir. Bitkilerde, kodlamayan RNA transkriptleri, tüm RNA transkriptlerinin %50'sinden fazlasını oluşturur (Mattic 2004; Shin ve ark. 2018). Kodlamayan RNA'lar farklılaşma, büyüme, metabolizma ve hastalığa direnç süreçleri gibi önemli süreçlerin düzenlenmesinde ve abiyotik, biyotik streslere yanıt olarak köklü rollere sahiptirler. Bununla birlikte, ncRNA araştırması, yüksek verimli deneysel çalışmaların eksikliği nedeniyle büyük ölçüde sınırlıydı. Küçük boyutları nedeniyle, genomik örnekler içindeki ncRNA'lar genellikle önemsiz DNA olarak yanlış tanımlanmaktaydılar (Grobhans ve Filipowicz 2008; Kawaji ve Hayashizaki 2008; Budak ve Zhang 2017). Bu nedenle, ncRNA'ların kapsamının ve çeşitliliğinin ayrıntıları, NGS teknolojilerinin geliştirilmesinin ardından ortaya çıkmaya başlamıştır (Eddy 2001; Shimon ve ark. 2007).

İlk keşiflerinden bu yana geçen süre içinde, ncRNA'lar, biyogenez yollarına dayalı olarak, birçoğu farklı, ancak örtüşen işlevlere sahip olan bir dizi farklı sınıfa ayrılır (Bonnet ve ark. 2006).

Kodlamayan RNA'nın çeşitliliği üzerine önceki çalışmalarla ilgili bu bulgular, ncRNA'ları iki ana kategoride sınıflandırılır: yapısal kodlamayan RNA'lar ve düzenleyici kodlamayan RNA'lar (Şekil 2.2). Yapısal kodlamayan RNA'lar, rRNA'lardan ve tRNA'lardan oluşur. Düzenleyici kodlamayan RNA'lar ayrıca küçük, orta ve uzun kodlamayan RNA'lar olmak üzere üç sınıfa ayrılır (Ponting ve ark. 2009; Alvarez-Dominguez ve Lodish, 2017). MiRNA, siRNA, piRNA, cisRNA, telsRNA, 20-50 nükleotit arasında bir boyuta sahip kısa kodlamayan RNA olarak kabul edilmiştir ve snoRNA, scRNA, tiRNA, snRNA orta sınıf olarak sınıflandırılmıştır. Uzun kodlamayan RNA'lar ise 50-200 nükleotit arasında bir boyuta sahip ve maksimum düzenleyici potansiyele sahip geniş bir RNA sınıfı olarak sınıflandırılmıştır (Dahariya ve ark. 2019). Kodlamayan RNA'lardan güncel olarak en fazla miRNA ve lncRNA çalışılmaktadır.

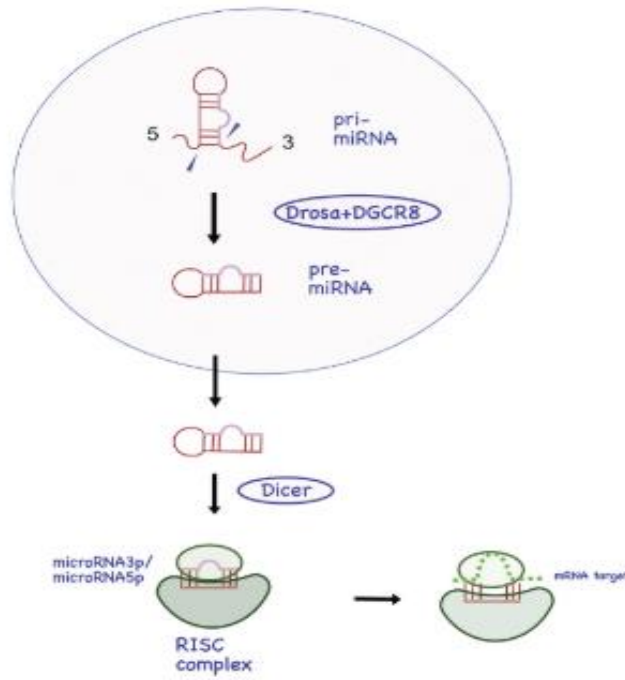


Şekil 2.2. Kodlamayan RNA'ların sınıflandırılması.

2.2.2.1. MikroRNA (miRNA)

MikroRNA'lar (miRNA'lar) 21-22 nükleotit (nt) uzunluğunda, diziye özgü düzenleyici elementlerdir. *Caenorhabditis elegans*'ta ilk miRNA'nın (lin-4) keşfedilmesinden bu yana (Lee ve ark. 1993), bitkilerde, hayvanlarda ve virüslerden yüzlerce miRNA tanımlanmıştır (Patil ve ark. 2020). Farklı düzenleyici elementler arasında miRNA'lar, mRNA'a ekspresyonunu düzenleyen küçük RNA molekülleridir. MikroRNA'ları kodlayan MIR genleri RNA polimeraz II tarafından pri- miRNA'lara çekirdekte bir saç tokası yapısı şeklinde kopyalanır. Pri-mikroRNA'lar daha sonra RNase III Drosha ve onun kofaktörü DGCR8 tarafından sitoplazmaya taşınan pre-mikroRNA'lara dönüştürülür. Pre-mikroRNA, Dicer'in RNase III tarafından, uçları sarkan yaklaşık 22 nükleotitten oluşan çift sarmallı bir mikroRNA'ya dönüştürülür. Olgun mikroRNA'nın iki ipliği 3p ve 5p olarak belirtilmiştir. Bazen her iki zincir de fonksiyonel mikroRNA'lara yol açar, ancak daha sıklıkla ikisinden biri bozulur. Olgun mikroRNA, RNA ile indüklenen susturma kompleksi RISC'ye dahil edilir ve transkripsiyon sonrası gen susturulmasını indüklemek için hedef mRNA'ya yönlendirilir (Şekil 2.3) (Panni ve ark. 2020). Gelişim, farklılaşma, büyüme, metabolizma ve hastalığa direnç süreçleri gibi

temel süreçlerin düzenlenmesinde ve biyotik, abiyotik streslere yanıt olarak köklü rollere sahiptirler (Sontheimer ve Erik 2009). Evrimsel olarak, bazı miRNA aileleri hem bitkilerde hem de hayvanlarda yüz milyonlarca yıl boyunca yüksek oranda korunur. Bireysel türler ayrıca çeşitli gelişimsel ve biyolojik süreçleri düzenlemek için türe özgü miRNA'ları kullanır (Voinnet 2009; Vitsios ve ark. 2017). Bu nedenle miRNA'lar, yakından ilişkili genotipler arasındaki genetik çeşitliliği incelemek için faydalıdır.



Şekil 2.3. MikroRNA biyogenezisi.

2.2.2.2. Uzun Kodlamayan RNA'lar (LncRNA)

Uzun Kodlamayan RNA'lar (LncRNA) >200 nt uzunluğunda kodlamayan RNA'lardır ve gen ekspresyonunun potansiyel olarak önemli düzenleyicileridir (Dykes ve Emanuelli 2017; Mach 2017; Qin ve ark. 2017). LncRNA'lar gen ekspresyonunun ve hücre kaderini kontrol eden çeşitli diğer süreçlerin ana düzenleyicileri olarak öne çıkar. LncRNA'lar, transkripsiyonel veya transkripsiyon sonrası gen ekspresyonunu nasıl etkilediklerine bağlı olarak çekirdeğin sınırları içinde lokalize olan ve hücre içinde hareket edenler olarak iki ayrı grupta mümkündür. DNA'ya veya DNA'ya bağlı proteinlere doğrudan bağlanma yoluyla transkripsiyonu etkileyebilir, böylece spesifik

genlerin transkripsiyonunu, kromatin durumunu veya kromozomal yapıyı modüle edebilirler. Gen transkripsiyonunu etkileyen lncRNA'ların, gen ekspresyonunun inhibisyonuna veya aktivasyonuna yol açan spesifik genomik lokuslar üzerindeki etkisi gösterilmiştir. Bazı lncRNA'ların ise, RNA polimeraz II ile etkileşim yoluyla transkripsiyon faktörlerinin hedef lokusları bağlamasını önleyerek transkripsiyonu düzenlediği de gösterilmiştir (Fort ve ark. 2021). Bitki türlerinde lncRNA'ların sayısı binlerce ile on binlerce arasında değişir. LncRNA dizileri farklı bitki türleri arasında bazı farklılıklar olduğu gözükse de, dizilerde göreceli olarak koruma gözlenmektedir. (Mohammadin ve ark. 2015). LncRNA'nın keşfedilmesinden bu yana tarımsal olarak önemli olan mahsüllerde birçok strese duyarlı lncRNA bildirilmiştir (Joshi ve ark., 2016; Liu ve ark. 2012; Xiao ve ark. 2015; Wang ve ark 2015; Zhu ve ark. 2015; Ding ve ark. 2011; Chen ve ark. 2018).

Kodlamayan RNA tabanlı markör sistemi, nispeten yüksek polimorfizm, tekrarlanabilirlik türler arasında aktarılabilirlik ve kullanım kolaylığı gibi avantajlar sağlamaktadır. Kodlamayan RNA'lardan geliştirilen moleküler markörler yakından ilişkili türlerde, genetik çeşitlilik analizi, popülasyonların ve tohumların genetik haritalaması dâhil olmak üzere bitki moleküler çalışmalar için büyük bir potansiyel sergilemektedir (Yadav ve Prasad 2014) .

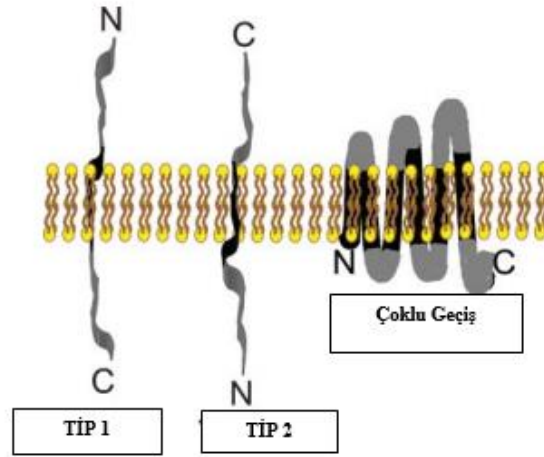
2.3. Transmembran Proteinleri

Plazma zarındaki proteinlerin çoğu, diğer maddelerin zarı geçmesine yardımcı olur. Plazma zarları ayrıca belirli tipte proteinler içerir. Bir zar proteini, bir hücrenin veya bir organelin zarına bağlı veya onunla ilişkili bir protein molekülüdür. Biyolojik zarların birçok işlevi vardır. Bir geçirgenlik bariyeri görevi görürler ve sonuç olarak, hücrenin dış ve iç bölgesi arasındaki sınır boyunca madde ve bilgiyi taşırlar (Guidotti 1972). Zar proteinleri, proteinin zarla nasıl ilişkili olduğuna bağlı olarak periferik ve integral protein olarak ikiye ayrılır. Periferik zar proteinleri, zara bağlı olamayan ancak zar yüzeyi ile değişen oranlarda ve çeşitli kısımlarla birleşen proteinlerdir. Bu proteinler genellikle zara, integral zar proteinleri ile etkileşim sağlarlar veya doğrudan lipit bölgeleri ile bağlanırlar. İntegral membran proteinleri, plazma zarının içine kalıcı olarak gömülür. Tek veya çoklu transmembran alanlarından oluşurlar. Bunlar tip 1, tip 2 veya çok geçişli proteinler olarak sınıflandırılabilir ve genomun sırasıyla %10 ve % 11-12'sini kapsadığı tahmin

edilmektedir (Şekil 2.4).Tip 1 proteinler, N-terminalleri hücre dışı boşluğa veya bir organeldeki lümen boşluğa bakacak şekilde yönlendirilir, Tip II proteinleri zıt yönelime sahiptir. Çok geçişli bir proteinin olası bir yöneliminin bir örneği, sitoplazmada bulunan hem N- hem de C-terminalleri Şekil 4 'te gösterilmiştir (Macher ve Yen 2007).

İntegral proteinlerinden biri de transmembran proteinleridir. Transmembran (TM) proteinlerinin ökaryotik genomların %15-30'unu oluşturduğu tahmin edilmektedir. Birkaç organizma için WGS analizi, tüm genlerin %20-25'inin transmembran proteinleri kodladığını göstermektedir (Jones, 1998; Wallin ve von Heijne, 1998). Transmembran proteinleri, hücre sinyali, zar geçirgen olmayan moleküllerin taşınması, hücre-hücre iletişimi, hücre tanıma gibi çok çeşitli temel biyolojik süreçlere aracılık eder. Bu nedenle, zar proteinlerinin yapısını ve işlevini anlamak, biyolojik ve farmakolojik araştırmalarda büyük önem taşımaktadır (Jones 2007). TM proteinleri sinyalizasyon için ayrıca önem taşımaktadır bu şekilde biyotik streslere yanıt oluşmasını sağlayabilirler. Sıralanan çeşitli hücresel süreçlerde membran proteinlerinin önemi, hastalıklarda oynadıkları roller ve ilaç hedefleri olarak potansiyelleri göz önüne alındığında, bu protein sınıfının daha iyi çalışılması zorunludur (Tan ve ark. 2008).

Bir proteinin işlevi belirlenmeye çalışılırken cevaplanması gereken önemli bir soru, proteinin hücrenin neresinde bulunduğuudur. Bir proteinin konumu, başka hangi protein türleri veya diğer moleküllerle etkileşime girebileceğini belirler. Bu süreçteki ilk adım, bir sinyal peptidine (SP) sahip olup olmadığını belirlemektir, çünkü bu bize bunun bir sitozolik protein olup olmadığını söyleyecektir (Käll 2010). Birçok integral membran proteini, stop-transfer dizisi olarak adlandırılan sinyal dizisine sahiptir (Rose ve ark. 1980; Yost ve ark. 1983).



Şekil 2.4. Transmembran protein çeşitleri.

2.4. Sinyal Peptidleri

Bir mRNA molekülü bir proteine çevrilirken birçok ribozom ona bağlanır ve bir poliribozom oluşturur. Sitozelda aynı ribozom havuzunu paylaşan iki ayrı poliribozom grubu vardır. Serbest ribozomlar herhangi bir zara bağlanmazlar. Zara bağlı ribozomlar endoplazmik retikulum (ER) yüzeyini kaplar ve ER yer değiştiren proteinlerin translokasyonlarını sağlar. Tüm bu proteinler yapılarında bulunan sinyal peptidleri sayesinde taşınabilirler. Hem prokaryotik hem de ökaryotik hücrelerdeki salgı proteinlerinin çoğu, sinyal peptidi (SP) olarak adlandırılan 20 ila 25 amino asit kalıntısının amino terminal uzantısına sahip bir öncüden sentezlenir.

Yeni sentezlenen polipeptit zincirinde, sinyal peptidi bir dizi protein-protein ve protein-lipid etkileşiminde yer alır. Translokasyon görevini yerine getiren sinyal peptidleri sinyalpeptidaz tarafından denatüre olarak yok olmaktadır. Sinyal peptidlerinin, kısa, pozitif yüklü bir amino terminal bölgesi (n-bölgesi); bir merkezi hidrofobik bölge (h-bölgesi); ve sinyal peptidaz tarafından bölünen bölgeyi içeren daha polar bir karboksi-terminal bölgesinden oluşan ortak bir yapısı vardır. Sinyal peptidi ribozomdan çıkarken, önce bir ribonükleoprotein kompleksi tarafından tanınır ve ER'ye aktarılır. Protein, zardan geçtikten sonra sinyal dizisi ER zarının lümen kısmında bulunan bir sinyal peptidazı tarafından hızla bozular. Ancak ER'ye giren tüm proteinler serbest bir şekilde kalmamaktadır. Bazı proteinler ER membranına gömülü kalır. Çoğu integral zar proteini, temizlenmemiş bir SP'ler ve art arda SP-benzeri "start -transfer" ve "stop -transfer"

dizileri içerir. ER 'ye gömülü kalan sinyal peptidleri bir proteinin nerden geldiğini ve hangi proteinlerle etkileşim halinde olacağını tanımlayabilmek açısından önemli olduğu düşünülmektedir.

2.5. Moleküler Markörler

1980'lerde moleküler markör teknolojilerinin gelişmesiyle bitki ıslahının kaderi değişti. Farklı tipte moleküler markörler geliştirildi ve sekanslama teknolojisindeki ilerlemeler mahsul gelişimini ilerletti. Bastein ve ark. (1980), DNA polimorfizmin saptanması "RFLP" kullanarak genetik haritalanmanın yapılandırılması hakkında rapor edilen ilk moleküler markör tekniğidir. Genetik markörler, belirli bir geni veya özelliği kontrol eden bilinen bir kromozom konumuna sahip bir gen veya DNA dizisidir. Genetik markörler, hedef gen ile yakından ilişkilidir. Genetik markörler genel olarak 2 kategoriye ayrılır. Klasik markörler ve DNA/moleküler markörler (Adhikari ve Ghosh, 2017)

Klasik Markörler:

Morfolojik Markörler; Morfolojik markörler tohum yapısı, çiçek rengi, büyüme alışkanlığı ve diğer önemli agronomik özellikler gibi görsel olarak ayırt etmemizi sağlar. Morfolojik markörlerin kullanımı kolaydır. Belirli aletlere gerek yoktur. Herhangi bir özel biyokimyasal ve molekül teknik gerektirmez. Sitolojik Markörler; Kromozomların sayıları, bantlama şekilleri, boyutu, şekli, sırası ve pozisyonunda var olan değişimlerde ilişkili markörlerdir. Bu varyasyonlar öökromatin ve heterokromatin dağılımdaki farklılıkları ortaya koymaktadır. Bu kromozom markörleri normal ve mutasyona uğramış kromozomları ayırt etmekte kullanılabilir. Bu tarz markörler bağlantı gruplarının tanımlanmasında ve fiziksel haritalamada kullanılabilir.

Biyokimyasal Markörler; Biyokimyasal markörler veya izozimler genetik çeşitliliğin, popülasyon yapısının, gen akışını ve popülasyon alt yapısının saptanmasında başarıyla kullanılır. Sayıları daha azdır ve daha az polimorfizm tespit eder. (Nadeem 2017).

Moleküler Markörler /DNA Markörleri:

Moleküler markörler nükleotit sekanslarıdır ve farklı bireylerin nükleotit sekansları arasında bulunan polimorfizm yoluyla araştırılabilir. İnsersiyon, delesyon, nokta mutasyon, duplikasyon ve translokasyon bu polimorfizmlerin temellerini oluşturur ancak genlerin aktivitesini etkilemez. (Nadeem 2017).

Moleküler markörler, hücrenin büyümesi, farklılaşması, gelişimi veya savunma durumundan bağımsız olarak tüm dokularda saptanabilir olduklarından diğer markör türlerine göre çok sayıda avantaj sağlar. Moleküler markörlerin ortaya çıkmasıyla DNA düzeyinde organizmalar arasında genetik çeşitlilik ve ilişkiler hakkında doğrudan çıkarımlar yapmak mümkündür. İdeal bir moleküler markör şu şekilde tanımlanabilir; polimorfik olmalı ve genom boyunca eşit olarak dağılmalıdır, genetik farklılıkların analizini sağlamalıdır, güvenilir olmalıdır, basit ve hızlı bulunmalıdır, az miktarda doku ve DNA örneğine ihtiyaç duymalıdır (Agarwal ve Padh 2008).

Bu moleküler markörler şu şekilde sınıflandırılır: (i) hibridizasyon bazlı markörler, yani restriksiyon fragmanı uzunluk polimorfizmleri (RFLP'ler), (ii) PCR bazlı markörler, yani polimorfik DNA'ların (RAPD'ler) rastgele amplifikasyonu, amplifiye edilmiş fragman uzunluğu polimorfizmleri (AFLP'ler), ara basit sekans tekrarlar (ISSR'lar) ve mikrosatellitler veya basit dizi tekrarları (SSR'lar) ve (iii) sekans bazlı işaretleyiciler, yani tek nükleotit polimorfizmleri (SNP'ler). Bu moleküler markörlerin çoğu ya genomik DNA kütüphanelerinden (örn. RFLP'ler veya SSR'ler) veya genomik DNA'nın rastgele PCR amplifikasyonundan (örn. RAPD'ler) veya her ikisinden (örn. AFLP'ler) geliştirilmiştir. Tüm bunlar arasında, SSR markörleri yüksek polimorfizm, çok alelli yapı, kodominant kalıtım, yeniden üretilebilirlik, kromozoma özgü konum, dahil olmak üzere birçok istenen özellik nedeniyle bitki genetiğinde ve ıslahında önemli bir önem kazanmıştır. (Agarwal ve Padh 2008).

2.5.1. Mikrosatellit Markörler (SSR)

Mikrosatellitler (SSR), kısa tandem tekrarlar veya basit sekans tekrarlar tüm ökaryotik genomlarda bulunur. SSR'lar mononükleotit motifi veya dinükleotit motifi gibi kısa (1-6 bp) bir motifin tekrarlarından oluşur (Şekil 2.5). Genomun her yerine, özellikle ökaryotların ökromatininde, nükleer ve organeller de DNA'da kodlanan ve kodlamayan bölgelerde çok sayıda SSR lokusu vardır. Bitkilerde her 50 kbp genomik sekans için yaklaşık bir SSR vardır (Mason 2014). SSR dizisi tamamen bir motifin tekrarlarından oluşuyorsa, "mükemmel" SSR olarak adlandırılırken, "kusurlu" veya "bileşik" SSR'lar birden fazla farklı tekrar birimi motifinden oluşur. SSR'lar, bitki genomlarının hem kodlanan hem de kodlamayan bölgelerinde yaygın olarak bulunur, ancak çoğunlukla

kodlamayan bölgelerde bulunurlar. SSR evrimi, yani SSR'da tekrar sayısında artış veya azalma ile ilişkili herhangi bir değişiklik, mutasyon oranı ile ilişkilidir (Senan S. 2014).

Mikrosatellit oluşumu evrimsel olarak dinamik bir süreçtir ve son derece karmaşık olduğu kanıtlanmıştır. Mikrosatellit oluşumu için olası mekanizmalar arasında; tek sarmallı DNA kayması, çift sarmallı DNA rekombinasyonu (eşit olmayan geçiş ve gen dönüşümü), DNA çift iplik tamir mekanizması yer alır. DNA replikasyonu sırasında, DNA polimeraz III'ün DNA şablon ipliği üzerinde tekrar bölgesinde kayması sonucunda oluşan mutasyonlar onarılmazsa, yeni oluşturulan DNA zincirinin tekrar bölgesinde genişlemesine veya büzülmesine neden olabilir ve bu şekilde genomda birçok SSR markörü oluşmaktadır. Dahası, UV ışınlaması, c-ışınlaması ve oksidatif stres vb. gibi harici stresin neden olduğu DNA hasarı, SSR'lar da kayma mutasyonlarına neden olabilir ve mutasyon oranlarını artırabilir. Kayma ve nokta mutasyonları arasındaki denge SSR'ların denge dağılımı için de önemlidir. Kaymada çoğalma büyümeyi desteklerken, nokta mutasyonu uzun tekrarlı bir diziyi iki veya daha fazla bölgeye böler. Bu nedenle, kayma ve nokta mutasyonlarının göreceli frekanslarındaki değişiklikler, bir genomdaki SSR 'ların dağılımı üzerinde doğrudan etkiye sahip olabilir (Adhikari ve Ghosh, 2017).

SSR alelleri, tekrarlanabilir sonuçların elde edilmesinin zor olduğu RAPD'ler gibi diğer bazı markör türlerinin aksine, yüksek oranda yeniden üretilebilir. SSR'lar ilk olarak bitkilerde 1984'te keşfedildi ve bunların moleküler bir markör olarak kullanımı sadece 5 yıl sonra 1989'da gerçekleştirildi (Mason 2014). O zamandan beri birçok SSR'lar geliştirildi ve çok çeşitli ürün türlerinde moleküler markörler olarak, başlangıçta buğday, pirinç ve soya fasulyesi gibi önemli mahsullerde ve daha yakın zamanda ayçiçeği ve fıstık gibi mahsullerde kullanıldı.

Mikrosatellitler, lokus başına yüksek mutasyon oranı, lokus özgüllüğü, yüksek polimorfizm, yüksek tekrarlanabilirlik, skorlama kolaylığı, çoklu alellik ile son derece bilgilendirici olduklarından önceki yıllarda bolca kullanılmıştır. SSR'ların kodominant markörler olduklarından dolayı, bireyde homozigot / heterozigot ayrımı yapabilirler. Ayrıca az bir DNA miktarı ile yeterli veriyi sağlarlar. Özellikle, genetik çeşitlilik, kantitatif özellik lokuslarının (QTL) keşfedilmesi, gen ve markör arasındaki bağlantı haritası yapımı, istenen özellikler için markör destekli seçim (MAS), adli tıp ve ebeveynlik analizi, DNA parmak izi, genom çapında ilişki çalışması (GWAS) ve genetik teşhis, ve genom organizasyonu çalışmalarında kullanılmaktadır (Nadeem 2017)



Şekil 2.5. SSR örneği.

2.5.1.1. Kloroplast Mikrosatellit Markörler

Kloroplast genomu karakterize eden daha az mutasyon oranı nedeniyle yeterli dizi varyasyonu tespit etmek zordur. Bunun aksine kloroplast mikrosatellitler (CpSSR) kolayca genotipleme ile daha yüksek polimorfizm seviyeleri sağlar. Bu da onları popülasyon genetiği çalışmaları için oldukça kullanışlı ve popüler bir markör haline getirir. Bu mikrosatellitler 8-15 kez tekrarlanan mononükleotit motifleri içerir. CpSSR'lar da polimorfizm seviyeleri türler ve lokus arasında oldukça değişmektedir.

Kloroplastlar birbirinden bağımsız olarak kalıtsaldır. Kloroplast kromozomu cpSSR lokuslarının bağlı olduğu rekombinant olmayan bir moleküldür (Agarwal ve Padh 2008).

2.5.1.2. Mitokondriyal Mikrosatellit Markörler

Bitki mitokondriyal DNA (MtDNA) ökaryotlar arasında en büyük ve en az gen yoğunluğuna sahiptir ve çok dinamiktir. MtDNA markörlerinin evrim hızı yavaştır. Bitkilerde mitokondriyal genomlar reorganizasyon hızı nedeniyle filogenetik analizlerde genellikle kullanılmazlar. Mitokondriyal tekrarlar popülasyonun özellik bazlı ayrımı için kullanılmıştır (Adhikari ve Ghosh 2017) .

2.5.2. Mikrosatellit Markörlerin Tanımlanması ve Genotiplenmesi

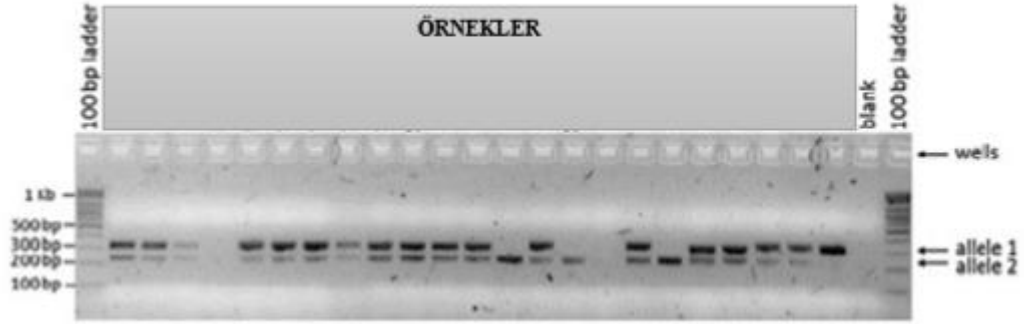
Bitki genetiğinde SSR markörlerin geniş uygulanabilirliğine rağmen, SSR markörlerinin gelişimi türlerin çoğunda, özellikle küçük mahsullerde özgü SSR geliştirmek sorun haline gelmektedir. SSR markörleri, hayvan genomuna kıyasla bitki genomlarında nispeten düşüktür (Mason 2014).

Biyoinformatik araçların gelişmesiyle birlikte NGS teknolojileri, SSR markörlerini geliştirmek için kullanılan geleneksel yöntemlere üstün bir alternatif sağlar. Bu, SSR ile zenginleştirilmiş DNA kütüphanelerinin oluşturulmasını önler ve büyük ölçekli SSR lokuslarının oluşturulması için hızlı bir yaklaşım sağlar. Son zamanlarda büyük ölçüde NGS teknolojisindeki gelişmeler ve dizileme maliyetlerindeki azalma, SSR markörlerinin daha kolay, daha ucuz ve hızlı bir şekilde tanımlanmasını sağlayacaktır. SSR markör genotiplemesindeki ilk adım, SSR bölgesinin tekrar sekansına özel primerler tasarlamaktır (Taheri ve Shamshiri 2018). Başlıca mahsul türlerinde ve model bitkilerde, uygun SSR primerleri hâlihazırda kamuya açık veri tabanlarında mevcuttur, ancak diğer durumlarda SSR markörleri sıfırdan tasarlanmalıdır. Genomik dizi mevcutsa, SSR'leri tanımlamak için yazılım programları kullanılabilir ve SSR bölgelerini büyütmek için genomik diziden primerler tasarlanabilir. Bununla birlikte, birçok tür için genomik sekans mevcut değildir ve bu durumda genomik sekans kütüphaneleri oluşturulmalıdır. SSR'leri içerdikleri düşünülen bölgeler sekanslanmalı, primer sekanslar tasarlanmalı ve primerler polimorfik ürünleri amplifiye etmek için denenmelidir (Fu ve An 2013).

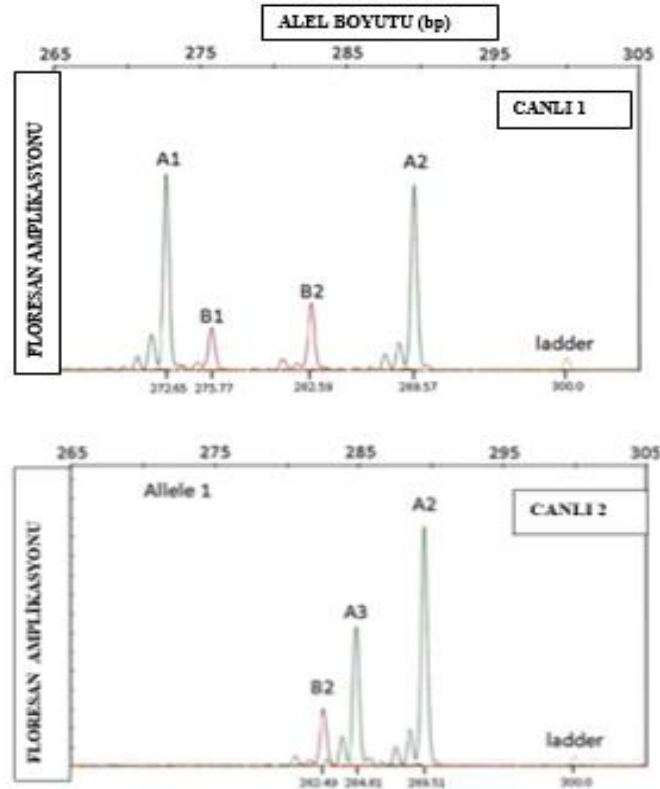
SSR primerleri mevcut olduğunda, SSR bölgesinin PCR esaslı amplifikasyonu genotipleme için kullanılabilir ayrıca hibridizasyon bazlı diğer yöntemler SSR'leri tespit etmek için tercih edilebilir. PCR bazlı amplifikasyon açık arayla en basit ve en yaygın olanıdır. Hibridizasyon yöntemi, özellikle SSR yönünden zengin genomlar için nispeten basit olmasına rağmen, düşük SSR frekanslarına sahip türler için son derece verimsiz olabilir. (Mason 2014)

PCR primerleri, hedef SSR sekansının her iki tarafına da spesifik olan oligonükleotitlerdir (genellikle 18–25 bp uzunluğunda) 5' – 3' yönünde sekans için spesifik bir ileri primer ve spesifik bir ters primer 3' – 5' yönü ile DNA'nın amplifikasyonu sağlanır. Daha sonrasında ise oluşan PCR ürününün boyut polimorfizmleri görselleştirmektir. Bunu yapmanın en yaygın ve kolay erişilebilir yolu, PCR ürününün katı bir jel üzerindeki bir kuyuya yüklendiği ve içinden bir elektrik akımı geçtiği agaroz jel elektroforezidir (Şekil 2.6). Daha küçük parçalar daha büyük parçalardan daha hızlı hareket ettiğinden, bu işlem zamanla negatif yüklü DNA parçalarını ayırır. Agaroz jel elektroforezine (AGE) popüler alternatifler arasında poliakrilamid jel elektroforezi (PAGE) ve kapiler jel elektroforezi bulunur. PAGE, teknik olarak AGE'den daha zor olsa da fragman boyutundaki daha küçük farklılıkları%5 veya daha az bir farka kadar çözebilme avantajına sahiptir. Kapiler jel elektroforezi genellikle

DNA fragmanlarının floresan etiketlenmesini ve ardından bir Sanger dizileme makinesi kullanılarak saptamayı içerir. Bu yöntemde DNA,kılcal tüp jellerine yüklenir ve elektroforezden sonra, floresan olarak etiketlenmiş DNA fragmanları kaydedilir (Şekil 2.7) (Taheri ve Shamshiri 2018).



Şekil 2.6. SSR markörlerinin elektroforez görüntüsüne örnek.



Şekil 2.7. Kapiler elektroforez SSR sonuçlarına örnek.

3.METERYAL METOD

3.1. Veri Analizi

3.1.1. Transpozonların Karakterize Edilmesi

Solanum melongena kromozom sekanslarına NCBI veri tabanından erişim sağlanmıştır (GenBank kayıt no: PRJDB1505). Genomda var olan transpozonları tespit etmek için Galaxy Europe (v 22.01) (Afgan ve ark. 2018) sunucusundan erişilebilen RepeatModeler (Version 2.0.3) (Flynn ve ark. 2020) programı kullanılmıştır. RepeatModeler programı, bir genom sekansında transpozonların tanımlanması ve sınıflandırılmasını gerçekleştirmektedir. Kullanılan program transpozonları bir veri tabanına bağlı olmadan tanımlamaktadır. RepeatModeler programında transpozon belirteci modeller önceden tanımlanmıştır ve bu sayede var olan transpozonları patlıcan genomunda *de novo* olarak karakterize edilebilmiştir.

3.1.2. Kodlamayan RNA'ların Karakterize Edilmesi

Solanum melongena genomunda bulunan kodlamayan RNA'ları karakterize etmek için PLEK (A. Li ve ark. 2014) programı kullanılmıştır. Program Phython aracılığı ile çalışmaktadır. Öncelikle PLEK programının doğru bölgeleri bulabilmesi için dikotil bitkilerin model organizması olan *Arabidopsis thaliana* ve patlıcanında içinde bulunduğu *Solanaceae* ailesinin model organizması olan *Solanum lycopersicum* kodlamayan RNA'larına miRBase (Kozomara ve Griffiths-Jones 2014) veri tabanından erişim sağlanmıştır. Kodlamayan RNA dizileri kullanılarak oluşturulan tanıma dosyası PLEK programına tanıtılmıştır ve bu şekilde kodlamayan RNA'lar programa öğretilmiştir. Program SVM (Support Vector Machine) algoritması üzerinden ikili sınıflandırma yaparak kodlamayan RNA'ları ve kodlanan RNA'ları ayırmaktadır. SVM algoritması bu gruplamayı k-mer değeri üzerinden yapmaktadır. K-mer değeri ne kadar yüksek ise programın o kadar iyi bir ayırım yaptığı sonucuna varılmıştır. Fakat yapılan çalışmalar sonucunda bu değer için 5 değerinin üstünün herhangi bir ayırım yapmadığı belirtilmiştir. Bu sebeple tez çalışmasının kapsamında k-mer değeri 5 olarak alınmıştır. Kodlanan ve kodlamayan RNA ayırımı için program üzerinden 1364 model hesaplanmış sonuç olarak

iki grup birbirinden ayrılmıştır. Diğer kullanılan parametreler için SVM algoritmasının çalışması için önerilen parametreler kullanılmıştır. Kullanılan parametreler Tablo 3.1 'de verilmiştir. Kodlamayan RNA'ların çoğu düzenledikleri proteinlerin gen bölgelerine yakın bölgelerde bulunurlar. Düzenledikleri fonksiyonel olma ihtimali olan bölgeleri de analiz edebilmek için kodlamayan RNA sekanslarının 2000 nt çevresi alınarak analiz gerçekleştirilmiştir. Patlıcan genom sekans verisi FASTA formatında input olarak kullanılmış ve bulunan kodlanan ve kodlamayan RNA'lar ayrılmıştır.

Tablo 3.1. PLEK programı için kullanılan parametreler.

Parametreler	Değerler
mRNA	own_model/plek_model_mirna.fa
lncRNA	own_model/plek_model_lncrna.fa
Minimum sekans uzunluğu (nt)	200
Thread sayısı	12
n-kat çapraz doğrulama	10
log2c	0,5,1
log2g	0,-5,-1
k-mer	5

3.1.3. Transmembran Proteinlerinin Karakterize Edilmesi

Transmembran proteini tahmin etmek için Galaxy Europe (v 22.01) (Afgan ve ark. 2018) sunucusu üzerinden TMHMM (Krogh ve ark. 2001) programı kullanılmıştır. TMHMM, Markov modeli (HMM) yaklaşımına dayalıdır. Bir HMM'nin ana avantajlarından biri, diğer yöntemlerin çoğunda oldukça kaba bir şekilde yapılmış olan membran proteinleri sarmal uzunluğu için üst ve alt sınırlar belirleyerek modellemenin mümkün olmasıdır (Krogh ve ark. 2001). HMM hidrofobiklik, sarmal uzunlukları gibi özellikler üzerinden parametre tahmini yapmaktadır ve tahmin için algoritmalarını zaten mevcut olduğu bir modelde birleştirebildiğinden, transmembran proteinlerin tahmini için çok uygundur. TMHMM transmembran proteinlerini % 97-98 olasılık ile doğru tahmin etmektedir (Krogh ve ark. 2001). Sinyal peptidleri mevcut olduğunda doğruluk düşse de, %99'dan daha iyi özgüllük ve hassasiyetle sinyal peptidi ve zar proteinleri arasında ayırım yapabilir (Krogh ve ark. 2001). Araştırılan dizinin amino asit dağılımı önceden bir

transmembran proteini için hesaplanmış ve bilinmektedir. Amino asit dizilimi beklenen amino asit dağılımıyla karşılaştırılır, incelenen protein amino asit dağılımları ile topolojik bölgelerde beklenen amino asit dağılımları arasındaki korelasyon göz önüne alındığında en olası dizilim seçilir. Transmembran sarmallarında beklenen amino asit sayısı yüksek ise, bunun bir membran proteini olma olasılığında yüksek olduğunu göstermektedir (Krogh ve ark. 2001). *Solanum melongena* genom sekansı üzerinden TMHMM programı çalıştırılmıştır. En kısa transmembran heliks yapıları yaklaşık 18 amino asitten oluşmaktadır, dolayısı ile aminoasit sayısı buna yakın olmalıdır. ExpAA parametresi transmembranda bulunan aminoasit kalıntı sayısını temsil etmektedir. Bu sebeple patlıcan genom sekansında bulunan transmembran protein sekanslarının ExpAA değerleri en az 18 olmalıdır. Elde edilen sonuçlar bu sebeple minimum ExpAA değeri 18 olacak şekilde filtrelenmiştir. TMHMM programı aynı zamanda transmembran proteinlerinden sinyal peptidlerini mevcut olduğunda iyi bir çözünürlük ile ayırmaktadır. Bulunan dizinin sinyal peptid olarak tanımlanabilmesi için EXP60 parametresi en az 2 olmalıdır (Krogh ve ark. 2001). Çalışmada kullanılan parametreler Tablo 3.2’de verilmiştir. Bulunan sinyal peptidlerini ayırt etmek için ise bu parametrede filtreleme yöntemi kullanılmıştır. Bu bağlamda sinyal peptid proteinleri ve transmembran proteinleri karakterize edilmiştir.

Tablo 3.2. Transmembran proteinleri parametreleri.

Parametreler	Tanımları
ExpAA	TM sarmalında amino asit sayısı
Exp60	Bulunan ilk 60 amino asit sayısı
N-best	Tahmin edilen heliks sayısı
Topoloji	TM konumu

3.1.4. Mikrosatellit Markörlerin Geliştirilmesi

Tez çalışması kapsamında geliştirilen transpozonlar, kodlamayan RNA’lar ve *Solanum melongena* genomuna ait EST (İfade Edilmiş Dizi Etiketleri) sekansları, SSR markörlerinin geliştirilmesinde kullanılmıştır.

SSR markörlerinin tanımlanması GMATA (Genom çapında Mikrosatellit Analiz Aracı Paketi) (Wang X. ve Wang L. 2016) programı kullanılarak biyoinformatik olarak gerçekleştirilmiştir. Program FASTA uzantısı ile çalışmaktadır. Bu sebeple *Solanum*

melongena genomuna ait EST sekans koleksiyonu NCBI veri tabanından FASTA formatında indirilmiştir. Kodlamayan RNA sekans contigleri *Solanum melongena* genomunda bulunup FASTA formatında düzenlenmiş ve GMATA programı ile mikrosatellit lokus tanımlama analizlerinde kullanılmıştır. Karakterize edilen transpozon sekansları da FASTA formatında markör tanımlama analizleri için input olarak kullanılmıştır. GMATA programı Perl, R ve Java programları ile birlikte çalışmaktadır. Program mikrosatellit lokusların tanımlanması, lokusların amplifikasyonu için primer sekans dizaynı ve e-map analizi ile markörlerin in silico haritalamasını gerçekleştirmektedir.

Patlıcan genomundan karakterize edilen TE, ncRNA ve patlıcan genomunun var olan EST sekansı için de aynı parametreler kullanılarak GMATA programında analizler gerçekleştirilmiştir. GMATA yazılımı kapsamında bulunan “SSR Geliştirme” bölümü için sekanslar FASTA formatında bu bölüme yüklenmiştir. Genomda SSR geliştirmek için minimum ve maksimum nükleotit uzunluğu değerleri sırası ile 2 ve 6, minimum motif tekrar sayısı ise 5 olacak şekilde uygulanmıştır. Yapılan analiz sonucunda (.ssr) çıktı dosyaları alınmıştır.

Bulunan SSR sekansları GMATA programının “Markör Dizaynı” bölümü kullanılarak PCR markörlerine dönüştürülmüştür. SSR geliştirme bölümünde alınan SSR dosyası (.ssr) sonuçları kullanarak “Markör Dizaynı” bölümü çalıştırılmıştır. Primer dizaynı işlemlerinde amplikon boyutu parametreleri; minimum amplikon boyutu 120 bp, maksimum 300 bp olacak şekilde seçilmiştir. Böylece bulunan SSR lokusları için PCR markörleri geliştirilmiştir. Bu işlemin sonucunda primer sekansı, primer uzunluğu, primerin bağlanma sıcaklığı verileri çıktı olarak elde edilmiştir. Bu analiz sonucunda (.mk) ve (.sts) uzantılı iki tür çıktı dosyası oluşturulmuştur. GMATA programının e-map modülü ise, markör haritalamasında kullanılmıştır. E-map analizi ile patlıcan genomunda geliştirilmiş olan (.sts) uzantılı dosya input olarak kullanılarak e-PCR yöntemi ile genom sekansı üzerinde fiziksel olarak haritalanmıştır.

3.1.5. Geliştirilen Markörlerin Fonksiyonlarının Belirlenmesi

Kodlamayan RNA'lara komşu genomik sekanslar ve EST sekanslarından geliştirilen aday moleküler markörlerin *Solanum melongena* bitkisi için biyolojik süreçlerdeki, moleküler işlevdeki ve hücresel bileşiminde önemli özellikleri ifade ettiği

düşünülmektedir. Bu bağlamda Omicsbox (<https://www.biobam.com/omicsbox/>) aracılığıyla geliştirilen markör sekanslarının protein anotasyonları (tanımlama) yapılmıştır. Omicsbox programı, kodon biansı göz önüne alarak homoloji tabanlı sekans üzerinden anotasyon yapmaktadır. NCBI veri tabanında *Solanum melongena* türüne ait tanımlanmış 617 adet peptit sekansı bulunmaktadır. Veri tabanlarında patlıcan türüne spesifik peptit sekansı kayıtlarının geliştirilen markörlerin fonksiyonlarını belirlemek için yeterli olmaması sebebi ile türün en yakın akrabası olan domatese (*Solanum lycopersicum* L.) ait peptit sekans koleksiyonu (34,489 adet peptit sekansı) NCBI veri tabanından indirilmiştir. Markör lokusların fonksiyonel anotasyonu için Omicsbox programı kullanılmış, *S. lycopersicum* protein sekans koleksiyonu kullanılarak Omicsbox programında blast analiz veritabanı oluşturulmuştur. EST sekansları, transpozonlar ve kodlamayan RNA'lara komşu sekanslar, blastx analizi ile *S. lycopersicum* peptit sekans veri tabanı ile eşleştirilmiştir. Bunun sonucunda EST, transpozon ve kodlamayan RNA'lara komşu bölgelerde protein kodlama potansiyeli belirlenmiştir.

4. ARAŞTIRMA SONUÇLARI VE TARTIŞMA

4.1. SONUÇLAR

Son on yılda tüm genom dizileme (WGS) ve transkriptom uygulamalarındaki hızlı ilerlemeler, moleküler bilgilerin üretimini hızlandırmıştır ve sebze ıslahı alanında bir devrime yol açmıştır (Tiwari 2019). Domates gibi sebzeler, genomik araçların uygulanmasında model mahsuller olarak hizmet etmişlerdir ancak en büyük sebze bitki grubu ve dördüncü en büyük ekonomik bitki grubu (FAO 2000) olmasına rağmen patlıcan gibi mahsuller geride kalmaktadır ve karakterizasyonları için çok az moleküler veri mevcuttur (Doganlar ve ark. 2002). Tezin amacı patlıcan genomunda yüksek oranda korunan ve gen ekspresyonlarını düzenleyen kodlamayan RNA’larda ve genom çapında çeşitliliği ve dağılımı ile farklılık sağlayan transpozon lokasyonlarının karakterizasyonu amaçlar. Ayrıca karakterize edilen bu bölgelerden ve patlıcanın bulunan EST sekansından aday moleküler markörler dizayn edilmiş ve fonksiyonel olarak tanımlanmıştır.

4.1.2. Transpozonların Tespit Edilmesi

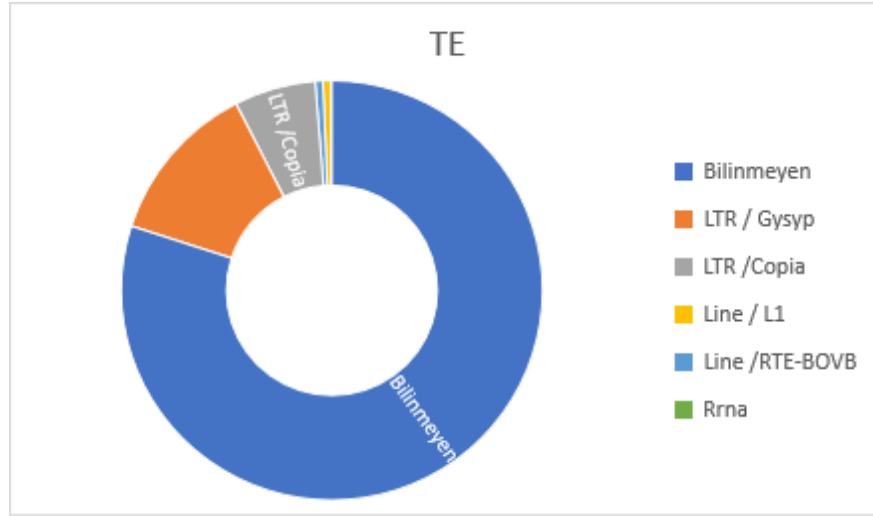
Transpozonlar patlıcan genomunda in silico olarak ilk defa analiz edilmiştir. *Solanum melongena* genomunda RepeatModeler programı ile *de novo* olarak 11.923 tane transpozon karakterize edilmiştir. Keşfedilmiş transpozonların bir kısmı Tablo 4.1 ‘de verilmiştir (<https://doi.org/10.6084/m9.figshare.20000054>). Yapılan analiz sonucunda karakterize edilen transpozonların bir kısmının hangi sınıfa ait olduğu belirlenirken, bir kısmının belirlenememiştir (Şekil 4.1). Karakterize edilen transpozonlardan; 128 tane LTR / Gypsy, 44 tane LTR / Copia, 6 adet LINE / L1, 6 adet LINE / Rte-bovb transpozon karakterize edilmiştir. Bu sınıflandırılan transpozonların hepsinin birer retrotranspozon olduğu söylenebilir. Retrotranspozonlar bitki genomlarında DNA transpozonlarına oranla daha fazla bulunurlar. Geriye kalan sekanslar ise bir sınıfa atanamamıştır.

Transpozonlar evrimsel açıdan önemlidirler. Hem genom büyüklüğünü değiştirirler hem de canlı da fenotipik olarak etki gösterirler. Eğer genomda gerçekleşen transpozisyon canlı için yarar sağlıyorsa sonraki nesillere aktarılır. Dolayısıyla karakterize ettiğimiz transpozonlar evrimsel çalışmalar için önem arz etmektedir. Türler

arasında veya aileler arasında transpozonların varlığını analiz ederek evrimsel yakınlıklarını, taksonomik olarak ait oldukları yerler bulunabilir.

Tablo 4.1. *Solanum melongena* genomunda karakterize edilen transpozonlar.

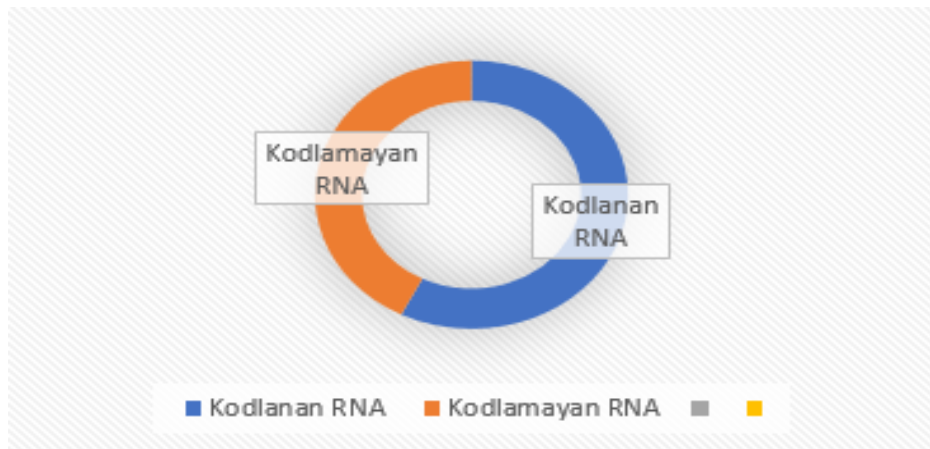
Transpozonlar	Sınıf	Sekans
Transpozon 1	-	TGTTAGAACCCGTACTTTTGGTACGTTGGGCCAATTACGAAATT TGGAACC AATGTTACAATTTTTGGAAAATTCGGAANTTANGTGTCAATTT TAGAAAT TT
Transpozon 2	-	TTTTGGTCCTGTTTTTTTACTTCGGTCCAAGACTTTTAAGTCT TGGGCC TAAGGGTTCACACACTTGAAAATACCTATACTTAGCCTTATTT CACCTC ATATCATCCTAAACATTCCCCAATACTCACTCAAACACTTAGN CNAATAT TTCCTCCTAAATTCAAGTCCCAA
Transpozon 3	LTR/ Copia	AATAAAGATGTGGCTTCNGTTAAAGTGCTGTGGCGAAACAAT AACNGAGA GGAGATGACCTGGGAAGCNGAGGAGCAGATGAAGAAGAAGT ATCCTCACT TGTTTTCACTGCCTNCAGGTAATTCAATTCTTCGCTTGACTTTG TTGCNT GATAAATAAATT
Transpozon 4	LTR / Gypsy	ACATCGGGNACCTCATCAACCCCTTGCCTCCCATGCAAAGCAT AAAACCG ATTGTGCCTTGGGGCACCACCTTGTCGGCCACCTCTCCCCAAA GGAACCG CATTAGCACCTTGTGGACGAACCTCCTTACCCTTGTGCGATT TCGAGGG CAATCCATCNCCTTATGNCCCACTTCCCCACAACCATAACANG CATTAGT ATCCCTCAAGCATGGGCCATTATCCTTATGCCCAACTTCCCCA CAACCAT AACANGCATTAGTATCCCTCAAGCATGGGCCACTATGACCCCT ACCACAT TTTGGACACACATCATTAGGAACCACACCAACACCTTGG GCCCTTGG GAAGTCATTGGCAAACCTCCTATTAGAAGCATGAGAAGACCC TTGGCCTT GATGGAACCGGCCACCCTTTTGGAACTT



Şekil 4.1. Karakterize edilen transpozonların oranları

4.1.3. Kodlamayan RNA'ların Tespit Edilmesi

Patlıcanın tüm genomunda kodlamayan RNA'lar in slico olarak ilk defa karakterize edilmiştir. Kodlamayan RNA analizi için kullanılan PLEK programında *Arabidopsis thaliana* ve *Solanum lycopersicum* model organizmalarının kullanılması ile birlikte *Solanum melongena* genomundan toplam 33.873 kodlanan RNA, 25.133 kodlamayan RNA bölgesi karakterize edilmiştir (Şekil 4.2). *Solanum melongena* genomunda karakterize edilen bir kısım kodlamayan RNA lokasyonları Tablo 4.2'de verilmiştir (<https://doi.org/10.6084/m9.figshare.19991504>).



Şekil 4.2. Patlıcan genomunda kodlanan ve kodlamayan RNA'ların dağılımı

Tablo 4.2. Karakterize edilen kodlamayan RNA'ların sekans üzerindeki konumları.

RNA'lar	Sekans Üzerindeki Lokasyonları
Kodlamayan RNA 1	-0.878668
Kodlamayan RNA 2	-0.151231
Kodlamayan RNA 3	-1.631.530
Kodlamayan RNA 4	-1.725.270
Kodlamayan RNA 5	-0.392965
Kodlamayan RNA 6	-0.431396
Kodlamayan RNA 7	-2.142.950
Kodlamayan RNA 8	-0.548811
Kodlamayan RNA 9	-1.503.820
Kodlamayan RNA 10	-1.109.720
Kodlamayan RNA 11	-0.695081
Kodlamayan RNA 12	-0.514957
Kodlamayan RNA 13	-1.561.470
Kodlamayan RNA 14	-0.787901
Kodlamayan RNA 15	-0.880656
Kodlamayan RNA 16	-1.237.200
Kodlamayan RNA 17	-1.914.710
Kodlamayan RNA 18	-0.700525
Kodlamayan RNA 19	-0.884963
Kodlamayan RNA 20	-1.044.600

Kodlamayan RNA analizlerinin sonucun da bulunan ncRNA 'ların çoğunun miRNA olduğunu söyleyebiliriz. Yapılan fonksiyonel homoloji tabanlı olan blast karakterizasyonu ile domates ve Arabidopsis thaliana 'da daha öncesinde tanımlanmış olan miRNA'lar patlıcan genomunda bulunmuştur (<https://doi.org/10.6084/m9.figshare.20158226>). Arabidopsis thaliana' nın tanımlanmış olan 664 adet miRNA'sı bulunmaktadır (Kozomara ve Griffiths-Jones 2014). Tanımlanmış olan miRNA 'ların 238 tanesi patlıcan da biyoinformatik olarak bulunmuştur. Arabidopsis thaliana patlıcanın uzaktan bir akrabası olmasına rağmen patlıcanda ncRNA'lara rastlanmış olması aslında ncRNA'ların çok fazla korunduğunun

bir göstergesi olabilir. Yapılan önceki çalışmalar sonucunda domateste ise 112 adet tanımlanmış miRNA bulunmaktadır (Kozomara ve Griffiths-Jones 2014). Analizi yapılan patlıcan da ise domateste bulunan 112 adet miRNA ‘dan 90 adet miRNA bulunmuştur. miRNA’lar hemen hemen tüm biyolojik süreçte etkisi olan RNA’lardır. Transkripsiyon sonrası gen susturmasına aracılık ederler. Farklılaşma, büyüme, metabolizma ve hastalığa direnç süreçleri gibi önemli süreçlerin düzenlenmesinde abiyotik ve biyotik streslere yanıt olarak köklü rollere sahiptirler. Özellikle abiyotik stres koşullarında canlıyı toleranslı duruma getirebilecek düzenlemeler yaparlar. Patlıcan da abiyotik stres ile ilgili olan 16 adet miRNA analiz edilmiştir (<https://doi.org/10.6084/m9.figshare.20158226>). Analiz edilen abiyotik stres ile ilgili miRNA’lar Tablo 4.3’de verilmiştir. Patlıcanda in slico olarak tanımlanan abiyotik stres ile ilgili miRNA’lar, patlıcanın stres koşullarına duyarlı ya da toleranslı olması açısından önemli olduğu düşünülmektedir. Karakterize edilen 16 miRNA, patlıcanda yüksek verim ve kaliteye sahip yeni mahsul çeşitlerinin ıslahı için umut verici olabilir.

Tablo 4.3. Patlıcanda analiz edilen abiyotik stres ile ilişkili miRNA'lar.

miRNA'lar	Sekanslar
>sly-MIR156a MI0009971	UUGACAGAAGAUAGAGAGCACUAAUG AUGAUUAGCUAAUUUCAUUCAAUUUG GGCAGCAAAAGCAUCUCAAUUCAUUU GUGCUCUCUAUGCUUCUGUCAU
>sly-MIR156b MI0009972	UUGACAGAAGAUAGAGAGCACGAAUA AUGAGGUGCUAAUUGGAAGCUGCACC UUAAUUCUUGUGAUCUCUAUACUUC UGUCAU
>sly-MIR156c MI0009973	UUGACAGAAGAUAGAGAGCACGAAUA ACGAGGUGCUAAUUGGAACAUGCACC UUAAUUCUUGUGCUCUCAUUCUUUU GUCAU
>sly-MIR159b MI0033710	AAGGUAGAGCUCUUCUGAAUUCGAAA GAGGAUCUAAGAGGGGAAAUCGAGCU GCUAAUCUAUGGAUCCCUCAAUCCUA UCUAUUCUCAAUAAUUAUUUAGAUAG GUUUGUGGUUUGCAUUAUCAGGAGC UGUAUUUACUCUUCUAGGAUCCUUAU UGGAAAGAAGGGAGCUCUACACUUG
>ath-MIR164c MI0001087	UAACACUUGAUGGAGAAGCAGGGCAC GUGCGAACACAAAUGAAAUCGAUCGG UACUUGUUGAUCUAUUUUUCGCACGU GUUCUACUACUCCAACACGUGUCU
>ath-MIR8166 MI0026794	AAUAAAAAUAAUGAAAAUGAAGGAGA GAGUGUAGAAAGUUUCUCACUUGAGA UACUUUUUACAACCCAUCAUGACAAU UUUCAUUUUAAU
>sly-MIR396a MI0029111	UGAUUCUUUGUGUACUUUCCACAGC UUUCUUGAACUGCAUACAUUAACAA AAAAAUCUGAUUUUUUUUGAUUUUU UUUUACUUUUUAUAAAGAAAAAUUUA GAUGGAAAGAAUUAUUUGUUGCGGU UCAAUAAAGCUGUGGGAAGAUACAUA UAGGAUCA
>sly-MIR167b MI0029112	UCUCACUUCUGUAUGUUGGUUUUUUG AGAGGUUAAAGCUGCCAGCAUGAUCU GGUAAGAACAAAAUCCUAUUUAUUU CAUCUAUUAUUUCUUAUAAAAAGAAG AAUUGAAUUUAAAUUUGAAUUUUUG UAAAAAAAAAUGUUGAGGUUUUUAA UUUCCUUCUUUUGGAUUUGAAAAAA GAAAAUUAUUUUUGUAAGGAUUUU AUUUCUUUGAAGAAAAUAUAGAUAGA UAGCUUUGGUUCCAUCAGGUCAUCUA GCAGCUUCAUCACUCACCAACAAUA CACACAAAGGA
>sly-MIR168a MI0024352	CUCUUAUUCGCUUGGUGCAGGUCGGG ACCUCAUUCGCCGGCGCGGGAUAA UGCCGGACGAACGACGGCGGUGUAA UUCUACUAAAGCUGUCACCGACGGAU AGAUGUUUGAUUAGCGGCGAAAUUG GGUCCUGCCUUGCAUCAACUGAAUUG GAG

Tablo 4.3. Patlıcanda analiz edilen abiyotik stres ile ilişkili miRNA'lar (Devam).

>sly-MIR169a MI0008361	GGUCUAUAAGGUUUGCAUGAAGUUGA AGAGAGUCAAGUUGCAGCCAAGGAUG ACUUGCCGGACCAAAAGUUGUGACCA UUUUCUAGUUCGGCAAGUUGUUUU GGCUAUAAGUUUGCUCUCUUCUUCUC AUGUUAUCCUCGUUAGACC
>sly-MIR169b MI0008362	UUUUAAUUUGGUUUCAGGCAGUCUCC UUGGCUACCUUGACAUGCUCUUUUUCU UUUAUGCAAGAUUUCUUUCAAUAUA GAAGUUUUAGUCAUCGUUAAAAUAUC GAGACGGGGUACAAGUCUUUUUAA AAGAAGAGAAGUCUAGCAUGAUGAGA AGAGUCUUAUUUGGUAGCCAAGGAUG ACUUGCCUGCACCUGACUUGAAA
>sly-MIR169c MI0008363	GUCUAGAGAGUUUAUUGUAUGAAGAGG UAGGGAGUGGAGUGCAGCCAAGGAUG ACUUGCCGAGAUGUUGUAGCCUAAUG CAAUGACAUUUCGGCAGGUCUUCUU GGCUACAUUUUAUUCUCUCCUAUUCA UGUAAAACUUUAUAGAC
>sly-MIR169d MI0008364	GAAGAGGGGUUUUAUUGAAGUUGAG AGUAUUUUUAGUAGCCAAGGAUGAC UUGCCUACUUCAUUUUGAUUAUAGUU UAAAAUAAUAAUUAUUGACAAUUAUUG UCAUUUGGUCGUAGGCAAGUUAUCCU GGCUAUCUUGACAUGCUCGUAAUUUU UCAUGUAAGACCUUAUCUUC
>sly-MIR395a MI0008369	AGGGCUUACUUUGAUGUUGGAAGCCU UGCUGAAGUGUUUGGGGAACUCCCG GGUUGACUUUGAAAAUUACGCAAACA AGUAUAUAUAAAUGCAUGAGUACGUA UUAGCUAAAGAUGAUGAGUACAAUAU UGAUUCAUGUACCUUUGGAGUUCCCU UGACCACUUCUAGAGAGCUUAAUAU UGAUUGUUUGAAGUCCU
>ath-MIR396a MI0001013	CUCUGUAUUCUCCACAGCUUUCUUG AACUGCAAAACUUCUUCAGAUUUUUU UUUUUUUCUUUUGAUUUCUUCACGC AUAAAAUAGUGAUUUUCUUCAUUUCU CUGCUCGAUUGAUUUUGCGGUUCAUA AAGCUGUGGGAAGAUACAGAC
>ath-MIR396b MI0001014	GGUCAUACUUUCCACAGCUUUCUUG AACUUUCUUUUUCAUUCCAUGUUUU UUUUCUUAACAAAAGUAAGAAGAAA AAAAACUUUAAGAUUAAGCAUUUUGG AAGCUCAAGAAAGCUGUGGGAAAACA UGACA

4.1.4. Transmembran Proteinlerin Tanımlanması

Markov Hidden modeline dayalı Galaxy Europe sunucusu üzerinde bulunan TMHMM programı kullanılarak *Solanum melongena* genomunda var olan transmembran proteinleri ve sinyal peptidleri karakterize edilmiştir. Yapılan çalışma sonucunda 33.873 veri elde edilmiş ve alınan veriler ExpAA ve Exp60 parametreleri ile filtrelenmiştir. Sonuçlar ExpAA değeri için ortalama 236.3573 değerinin üstü, Exp60 için 9.03 ortalama değerinin üstünde olanlar transmembran proteini ve sinyal peptidi için sonuç olarak eklenmiştir (<https://doi.org/10.6084/m9.figshare.19999985>). Transmembran protein sonuçlarının bir kısmı Tablo 4.4 'de verilmiştir.

Tablo 4.4. Transmembran proteinleri ve sinyal peptidleri.

ID	Uzunluğu	ExpAA	Exp60	N-Best	TOPOLOJİ
DF359985.1	67816	278,69	9,03	0	o
DF359370.1	101949	1792,32	9,06	0	o
DF357788.1	173527	276,42	9,11	0	o
DF365395.1	26273	554,05	9,11	0	o
DF361987.1	46769	715,88	9,11	0	o
DF358469.1	108116	295,47	9,13	0	o
DF359058.1	82716	309,57	9,15	0	o
DF357414.1	173458	901,11	9,15	0	o
DF359564.1	73823	600,77	9,17	0	o
DF360308.1	72660	543,77	9,2	0	o
DF376599.1	4999	240,61	9,23	0	o
DF359688.1	71751	597,04	9,23	0	o
BAUE01069822.1	24929	339,64	9,24	0	o
DF357497.1	159534	853,03	9,24	0	o
DF361108.1	54976	334,08	9,25	0	o
DF360639.1	59827	392,46	9,26	0	o
DF359592.1	73253	544,18	9,26	0	o
DF359389.1	76701	861,86	9,27	0	o
DF365099.1	38709	514,75	9,28	0	o
DF357301.1	214876	1088,84	9,36	0	o

Alınan sonuçlarda transmembran proteinlerinin hücre içindeki konumları belirlenmiştir. Hücre içine doğru i, dışına doğru olan heliks yapıları için o olarak

adlandırılmıştır. Bu şekilde tahmin edilen transmembran proteinlerinin ve sinyal proteinlerinin hücre içindeki konumları belirlenmiştir.

Hücre-hücre iletişimi, taşıma, sinyalizasyon ve biyotik stres ile ilgili bu genlerin karakterizasyonu önemlidir. Bitki oluşan sinyalizasyon sayesinde biyotik stres karşısında bir savunma mekanizması oluşturur. Biyotik stres anında hücrede oluşan sinyalizasyon takibi açısından bu bölgelerin tahmininin önemli olduğu düşünülmektedir

4.1.5. Mikrosatellit Markörlerin Geliştirilmesi

Karakterize edilen transpozonlardan, kodlamayan RNA’lardan ve patlıcana ait EST sekansından GMATA programı kullanılarak mikrosatellit markörler geliştirilmiştir.

Karakterize edilen transpozonların FASTA uzantılı dosyası GMATA programının “SSR Geliştirme” bölümünde analiz edilmesi sonucunda en fazla 5 tekrara sahip olan toplam 280 SSR lokusu geliştirilmiştir (<https://doi.org/10.6084/m9.figshare.20000291>). Transpozonlar sekans olarak uzun tekrarları içerdikleri için kısa tekrarlara sahip olan SSR markörleri transpozonlarda az bulunmuştur. SSR lokuslarından 165 primer çifti dizayn edilmiştir. Bu markörler en çok 2 ve 3 tekrardan oluşmaktadır (Tablo 4.5). Yapılan analiz sonucunda 5 ve 6 tekrarlı motif yoğunluğu TE sekanslarında az bulunmuştur.

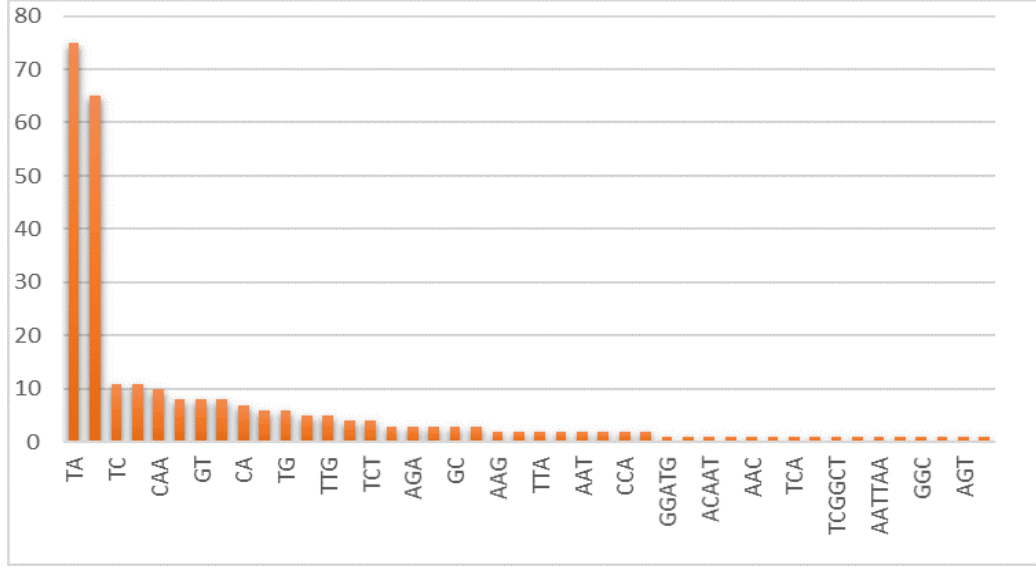
Tablo 4.5. Transpozonlarda geliştirilen SSR motiflerinin tekrar nükleotit sayısı.

Motif(-mer)	Total
2	207
3	67
6	2
4	2
5	2

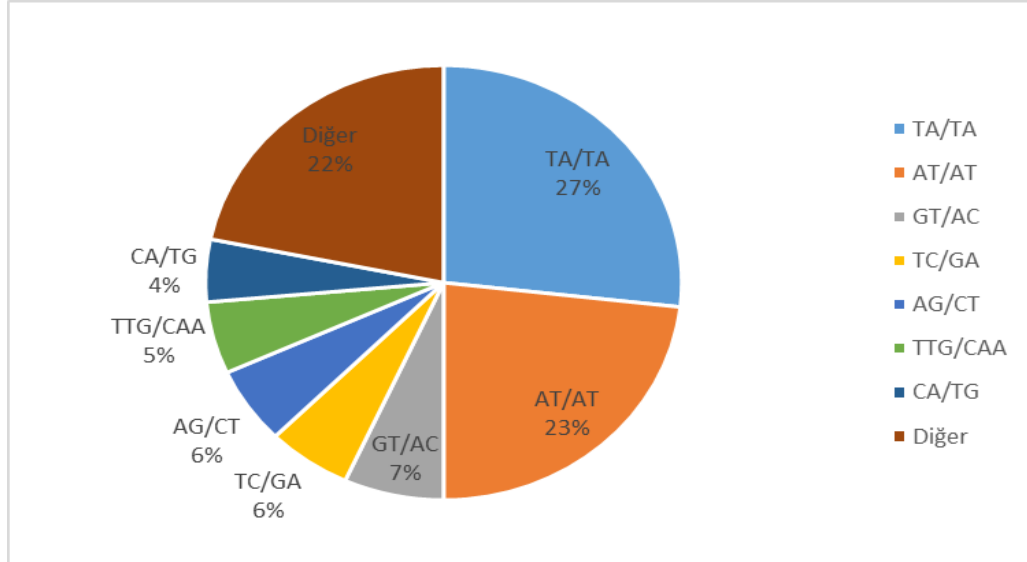
Transpozonlar genom boyunca en fazla tekrar eden SSR motifi 75 lokasyonda bulunan “TA” motifi olmuştur 65 lokasyonda ise “AT” motifi bulunmuştur. En çok bulunan trinükleotit ise “CAA” motifi olmuştur (Şekil 4.3-4.4) .

Transpozonlardan geliştirilen aday moleküler markörler evrimsel çalışmalarda kullanılabilirler. Bize türlerin birbirlerine olan uzaklıkları hakkında ve

taksonomik konumları hakkında bilgi verir. Bu sebeple de geliştirilen aday SSR markörlerin evrimsel çalışmalarda kullanımının etkili olacağını söyleyebiliriz.



Şekil 4.3. Transpozonlarda geliştirilen SSR motiflerinin dağılımı.



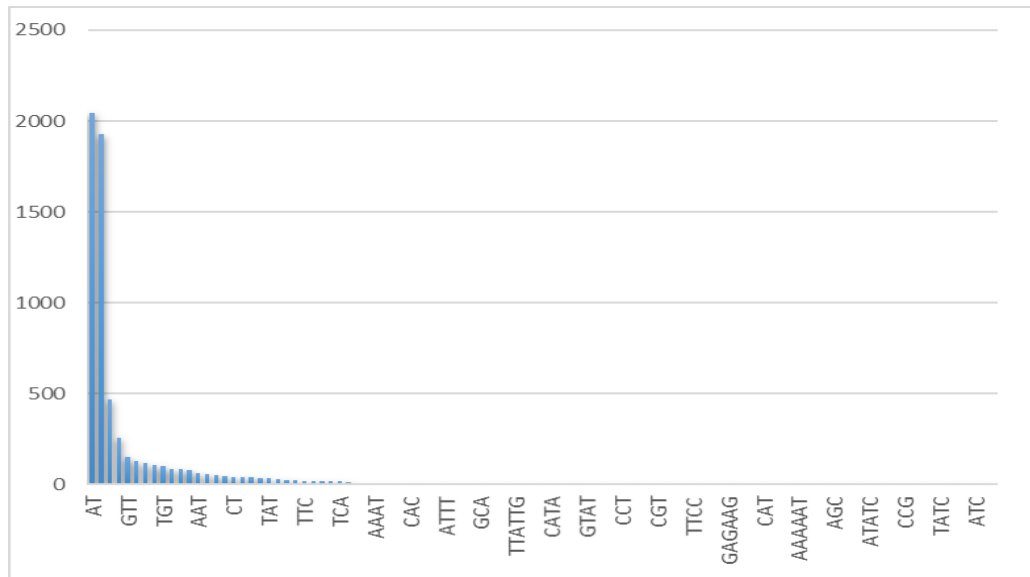
Şekil 4.4. Transpozonlarda geliştirilen SSR motiflerinden en çok bulunan markörlerin yüzdelik dağılım grafiği.

Geliştirilen kodlamayan RNA sekansları için GMATA analizinden sonra ise 6318 tekrar motifine sahip SSR lokusları bulunmaktadır. Bulunan SSR sekanslarından 2 ve 3 tekrarlı SSR motifleri daha fazla gelişmiştir (Tablo 4.6).Yapılan analiz sonucunda 5 ve 6 tekrarlı motif yoğunluğu kodlamayan RNA sekanslarında az bulunmuştur. Geliştirilen

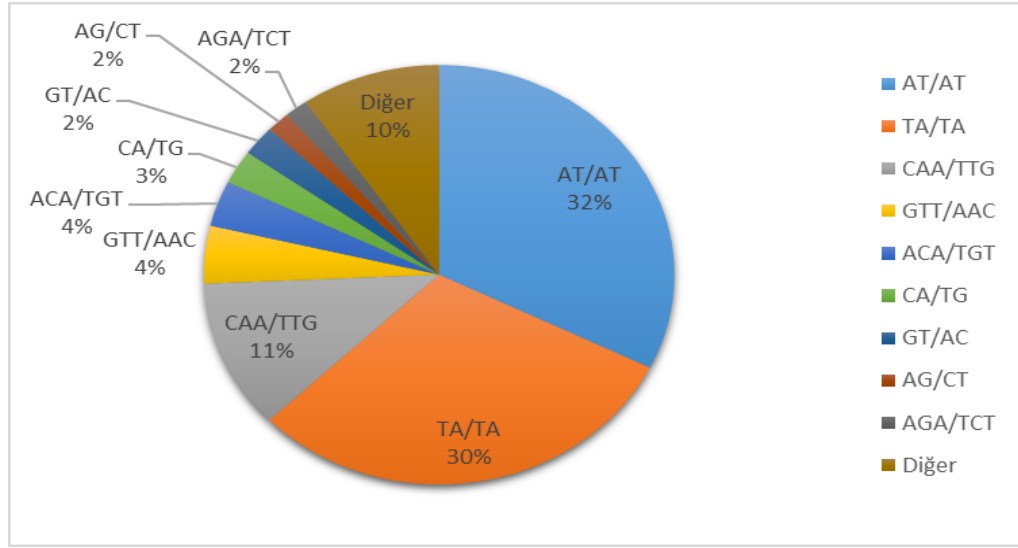
kodlamayan RNA SSR'larından en fazla 2042 lokasyonda bulunarak "AT" motifi geliştirilmiştir. 1927 lokasyonda ise "TA" motifi bulunmuştur. Bulunan trinükleotit sekanslarından en fazla "CAA" motifi, tetranükleotit sekanslarından ise "AAAT" geliştirilmiştir (Şekil 4.5-4.6). Karakterize edilen SSR lokuslarından 5416 lokus primer çifti olarak tasarlanmıştır (<https://doi.org/10.6084/m9.figshare.20000150>).

Tablo 4.6. Kodlamayan RNA bölgelerinde geliştirilen SSR motiflerinin tekrar nükleotit sayısı.

Motif(-mer)	Total
2	4477
3	1762
4	49
6	21
5	9



Şekil 4.5. Kodlamayan RNA'larda geliştirilen SSR motiflerinin dağılım grafiği.



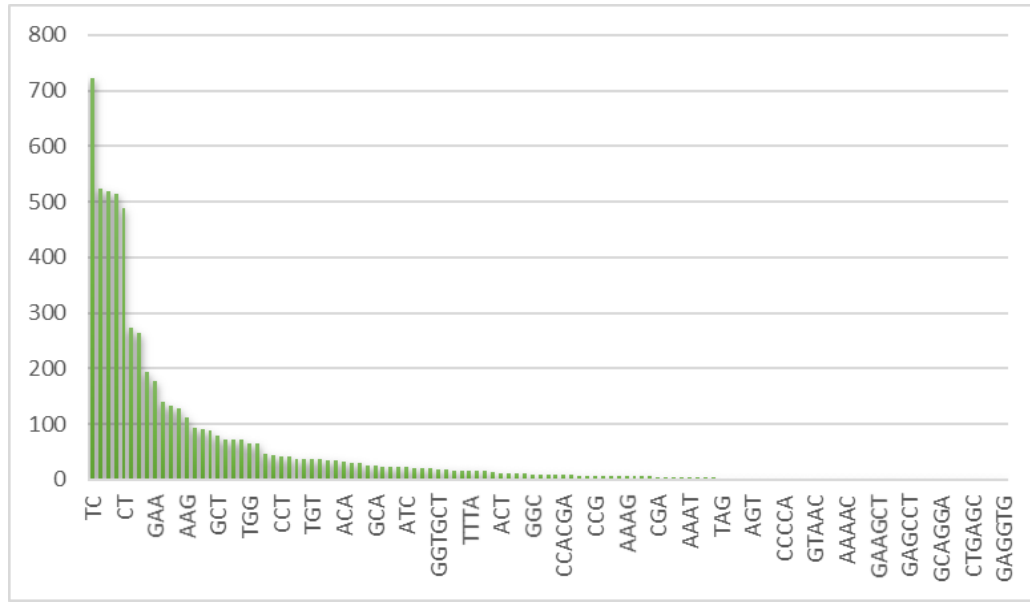
Şekil 4.6. Geliştirilen SSR motiflerinden bulunan markörlerin yüzdelik dağılım grafiği.

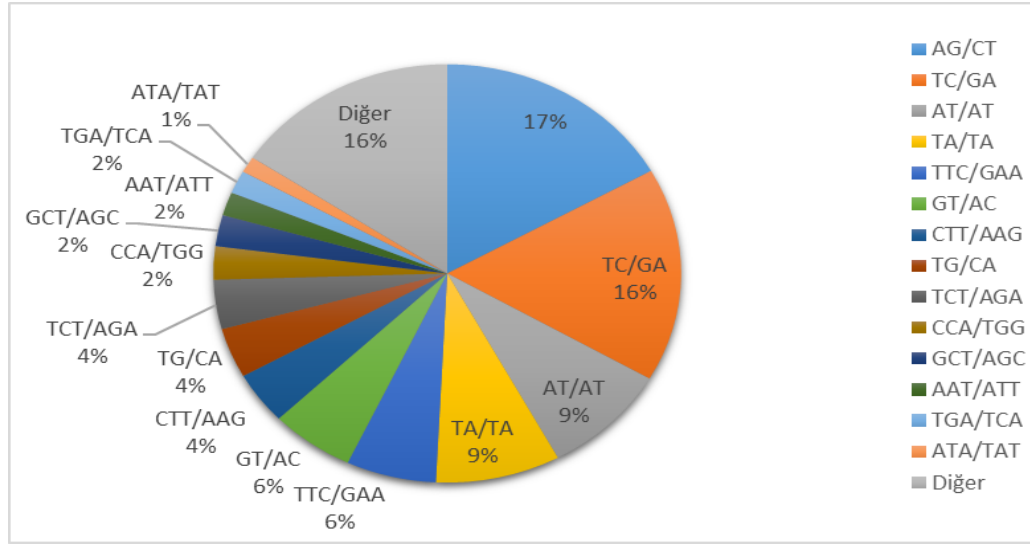
Kodlamayan RNA'lar aşırı derecede korunmuş bölgelerdir. Bu sebeple de bu bölgelerden polimorfizm bulmak zordur. Bu bölgelerden bulduğumuz aday SSR markörlerinin fonksiyonel olma ihtimali daha yüksektir.

Solanum melongena genomuna ait EST sekanslarından toplam 5973 SSR markörü geliştirilmiştir (<https://doi.org/10.6084/m9.figshare.20000294>). Geliştirilen SSR lokuslarından en fazla 2 ve 3 tekrarlı motifler bulunmaktadır (Tablo 4.7). Bulunan SSR sekanslarından en fazla "TC" motifi 723 lokasyonda, "AT" ise 524 lokasyonda belirlenmiştir. "TTC" motifi en fazla bulunan trinükleotit ve "ATTA" motifi ise tetranükleotit SSR'larda en fazla karakterize edilen motiftir (Şekil 4.7-4.8). Bulunan SSR motifleri arasında pentanükleotit ve heksanükleotit motifleri için EST genomunda kayda değer bir sonuç alınamamıştır. Geliştirilen 5973 SSR marköründen 3927 primer çifti geliştirilmiştir.

Tablo 4.7. *Solanum melongena* EST sekanslarından geliştirilen SSR motiflerinin nükleotit sayısı.

Motif(-mer)	Total
2	3616
3	2157
4	98
5	14
6	88

**Şekil 4.7.** EST sekanslarında geliştirilen SSR motiflerinin grafik üzerinde dağılımı.



Şekil 4.8. Geliştirilen SSR motiflerinden markörlerin yüzdelik dağılım grafiği.

Yapılan analizler sonucunda en fazla SSR markörü kodlamayan RNA sekanslarında geliştirilmiştir. Gelişen SSR sekansları EST, transpozon sekansları ve kodlamayan RNA sekansları da dâhil olmak üzere genellikle 2 ve 3 tekrardan oluşan motiflerden oluşmaktadır. Geliştirilen markörler daha sonra gen anotasyonları yapılmıştır.

Geliştirilen markörlerin 3 tekrardan oluşan motiflerin SSR bölgeleri açısından değerli olduğunu söylemek mümkündür çünkü bu bölgelerin fonksiyonel olma ihtimali daha yüksektir.

4.1.6. Geliştirilen Markörlerin Fonksiyonlarının Belirlenmesi

EST, transpozon ve kodlamayan RNA 'lardan geliştirilen markörlerin fonksiyonlarını belirlemek amacıyla Omiksbox programı kullanılarak *Solanum melongena* genomu için önemli fonksiyonlar analiz edilmiştir.

4.1.6.1. Transpozon Markör Sekanslarının Fonksiyonlarının Belirlenmesi

GMATA programından geliştirilen 280 SSR lokusu NCBI domates protein sekansı ile analiz yapılmıştır. Yapılan anotasyon sonucunda transpozonlar çoğunlukla protein kodlamadığından dolayı beklenildiği gibi hiçbir proteinle fonksiyonel olarak bir

anotasyonu gerçekleştirilmemiştir. Dolayısıyla yapılan Omicsbox analizi ile bir veri elde edilememiştir.

4.1.6.2. Kodlamayan RNA Markör Sekanslarının Fonksiyonlarının Belirlenmesi

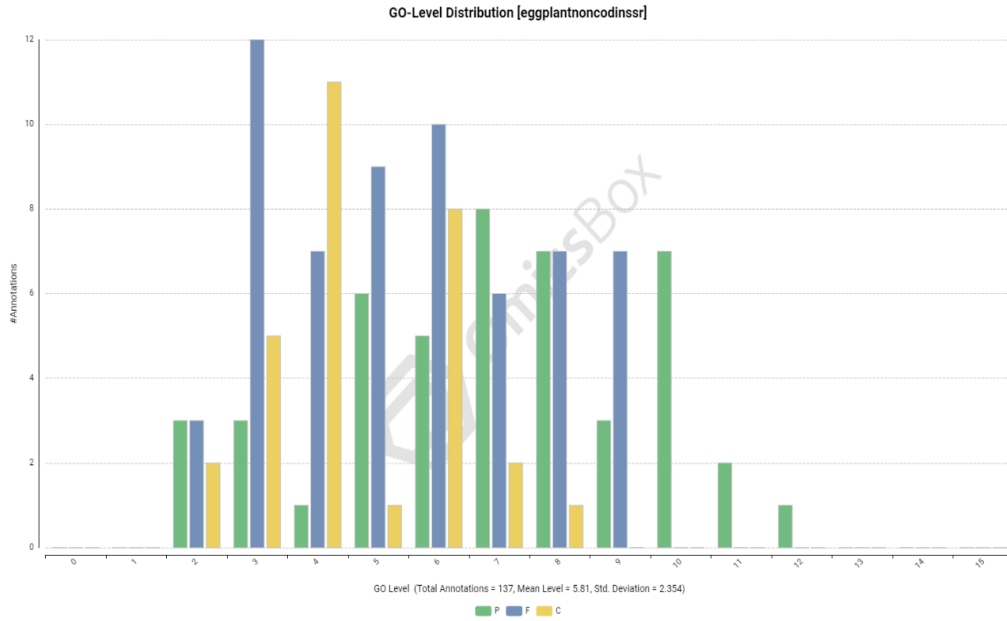
Kodlamayan RNA'larda karakterize edilen 6318 SSR lokusu, *Solanum lycopersicum* protein sekansı ile analiz edildi. Omicsbox programı biyolojik süreç (BP), hücresel bileşen (CC) ve moleküler fonksiyon (MF) olmak üzere üç gen ontoloji prensibiyle sonuçları vermektedir. MF; gen ürünün yani tanımlanan proteinin biyokimyasal olarak hangi görevi yaptığını, BP; proteinin hücrede hizmet ettiği biyolojik amacını ve CC; proteinin hücrenin hangi alanlarında görev aldığını belirtmektedir.

Patlıcanda karakterize edilen kodlamayan RNA'lar için yapılan analiz sonucunda anotasyonu sağlanan proteinler en fazla sitoplazma bölgesinde bulunmuştur. Sonrasında ise en çok çekirdekte görevli olan proteinlerle anotasyon sağlanmıştır. Karakterize edilen SSR anotasyonları biyolojik süreçte en çok bir immün savunma yolu olan MAPK yolu ve birçok hücresel olay için önemli olan protein fosforilasyonunda eşleşme gerçekleştirilmiştir. Gen anotasyonu en çok bir mitojenik bir protein kinaz olan MAP kinaz ile gerçekleştirilmiştir. MAP kinazlar bitkilerde patojen oluşturma durumunda bitkinin savunma mekanizmasını etkinleştirmektedir. İkinci olarak kodlamayan RNA SSR'larında en fazla tanımlanan protein bir transmembran proteini olan ATP bağlayan kaset taşıyıcıları olmuştur. Bu membran proteini sitoplazma ile hücre çevresi arasında oluşturduğu kanal ile ATP 'nin parçalanması sonucunda oluşan enerjiyi kullanarak madde taşınması sağlamaktadır. Yapılan analiz sonucunda metal iyon aktivitesi, protein kinaz aktivitesi, protein bağlama gibi birçok önemli proteinle anotasyonu sağlanmıştır. Sağlanan anotasyon sonucunda moleküler fonksiyonlarda, hücresel bileşenlerde ve biyolojik süreçlerde bulunan proteinler Şekil 4.9-4.10-4.11-4.12'de verilmiştir. Anotasyonu sağlanan markörlerin rastgele 10 tanesi Tablo 4.8'de verilmiştir.

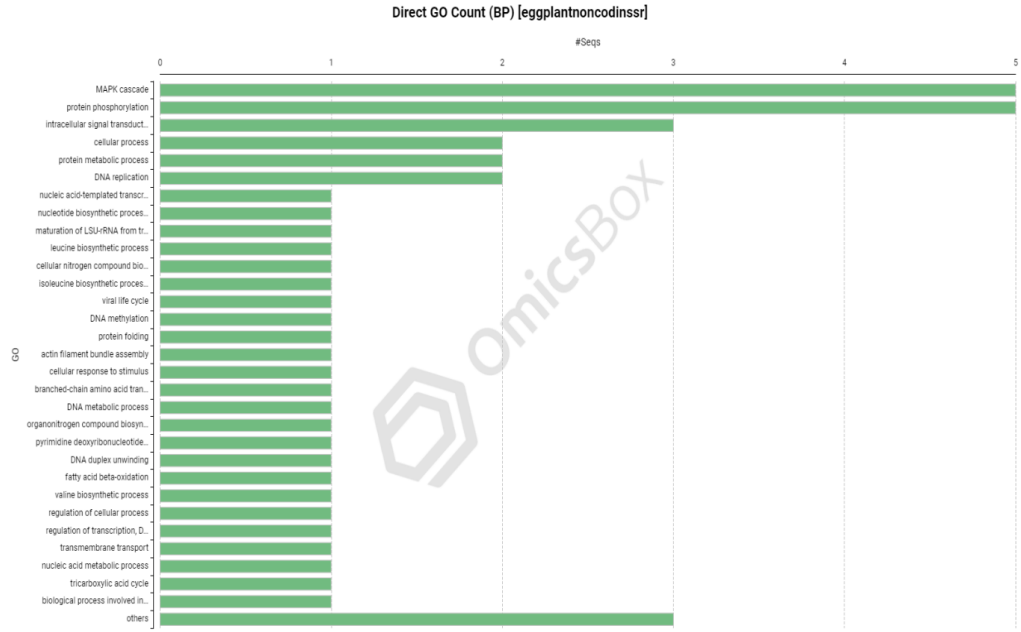
Kodlamayan RNA bölgelerinden elde edilen anotasyon sonuçlarının muhtemelen düzenledikleri proteinler olduğu söylenebilir.

Tablo 4.8. Omicsbox programında analiz edilen kodlamayan RNA markörlerinin rastgele 10 tanesinin protein sonuçları.

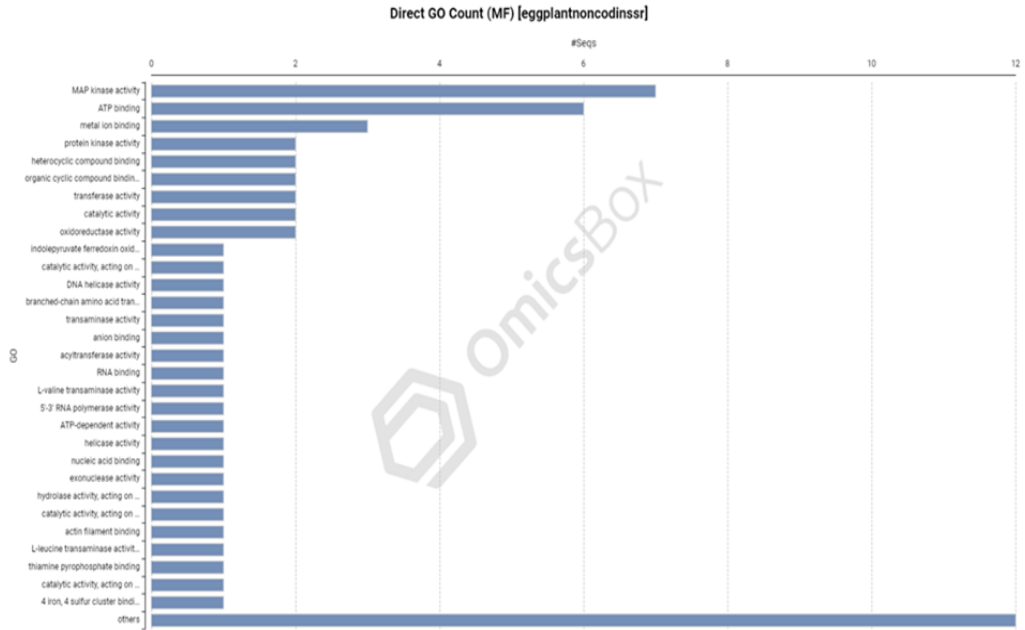
MARKÖR	PROTEİN
DF373458	Serin/treonin-protein fosfataz
DF381218	Ters transkriptaz içeren protein
DF381130	Gliserofosfodiester fosfodiesteraz
DF383756	Fenilalanin amonyum liyaz
BAUE01138976	PWWP proteini
DF372111	CCHC-Tip protein
DF382926	Gag-pol polyprotein
DF373807	Protein kinaz
BAUE01136329	Alkol açıl transferaz
DF378072	Sülfotransferaz



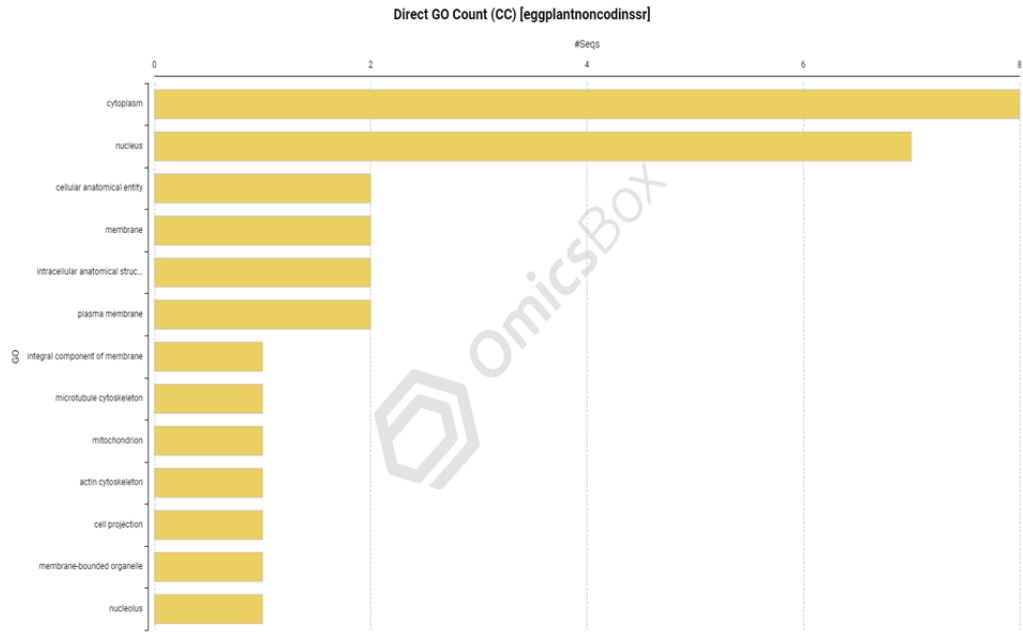
Şekil 4.9. Kodlamayan RNA gen anotasyon dağılımı biyolojik süreç (BP), hücresel bileşen (CC) moleküler fonksiyon (MF). Yeşil, sarı ve mavi sırasıyla BP, CC ve MF kategorilerini temsil eder.



Şekil 4.10. Biyolojik süreç kısmında tanımlanan proteinlerin dağılımı.



Şekil 4.11. Tanımlanan proteinlerin moleküler fonksiyon dağılımı.



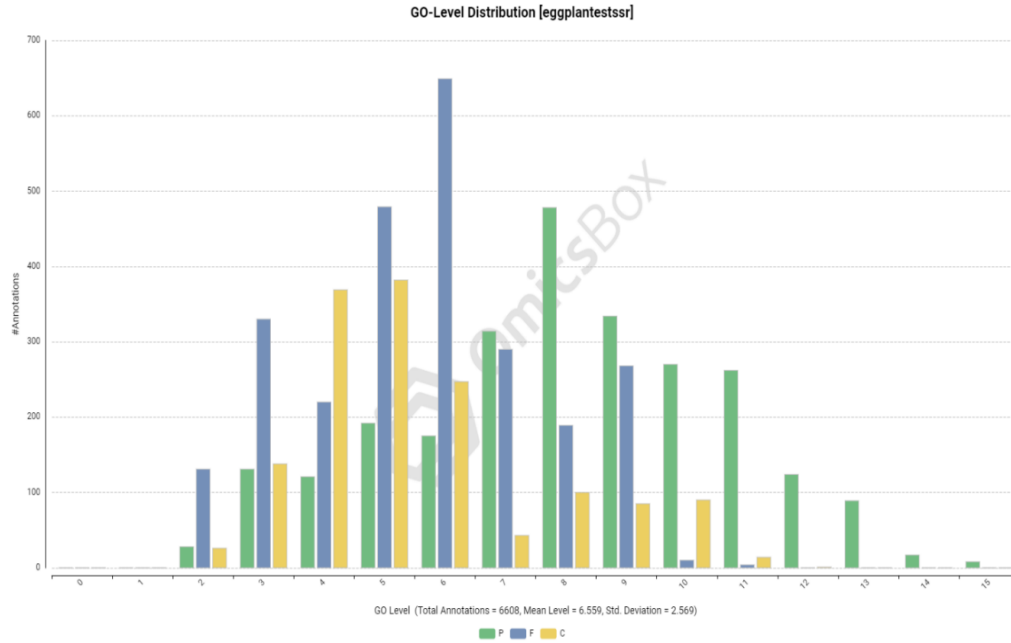
Şekil 4.12. Tanımlanan proteinlerin hücresel bileşenlerin dağılımı.

4.1.6.3. EST Markör Sekanslarının Fonksiyonlarının Belirlenmesi

Geliştirilen 5973 EST markörü Omikssbox programı ile analizi sonucun da en fazla zar bileşeninde ve sitoplazma da protein anotasyonu gerçekleştirilmiştir. Geliştirilen markörler hücre gelişme, biyolojik gelişim, moleküler fonksiyon gibi birçok önemli olayda yer almaktadır (Şekil 4.13.). Çoğu bölgede tanımlanan proteinler transkripsiyon düzenleyici, , translasyon, çift iplik tamiri mekanizmasında görev yapmaktadır. ATP bağlama, DNA bağlama, metal iyon bağlama gibi önemli moleküler fonksiyonları olduğu karakterize edilmiştir. Hücresel gelişme, biyolojik gelişme, uyarılara cevap verme gibi önemli kısımlarda görev yapmaktadır. Anotasyonu sağlanan markörlerden rastgele 10 tanesi Tablo 4.9’de verilmiştir. Tanımlamadan sonra markörlerin genel olarak dağılımı Şekil 4.13-4.14-4.15-4.16 ‘ da verilmiştir.

Tablo 4.9. Omicsbox programında analiz edilen EST markörlerinin rastgele 10 tanesinin protein sonuçları.

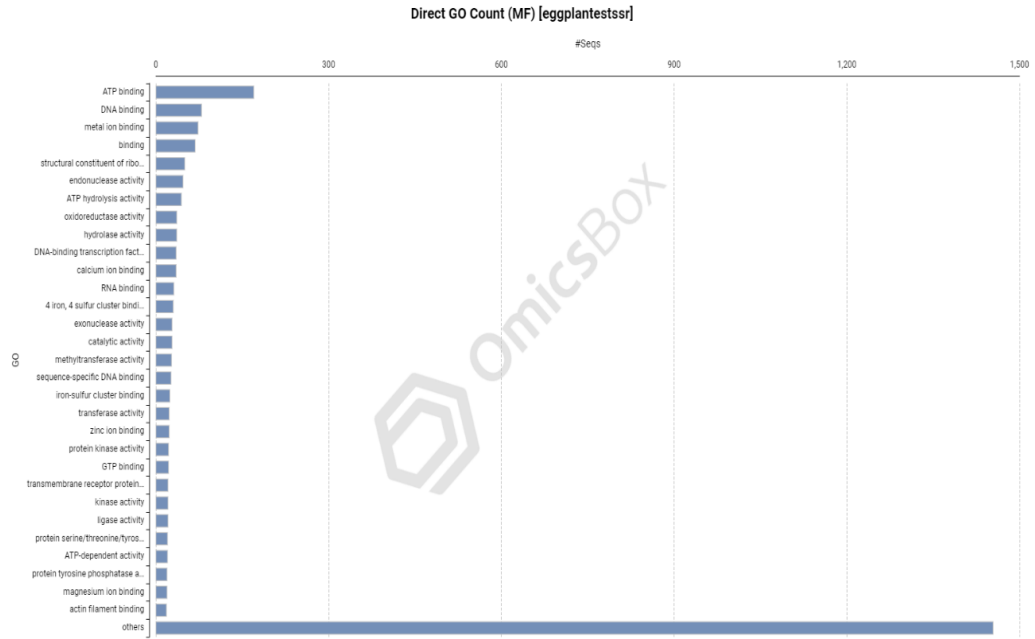
Markör	PROTEİN
FP030686.1	60S ribosomal protein L10
FS077884.1	NADH-sitokrom b5 redüktaz
FS071016.1	DDE Tnp4 içeren protein
FS089513.1	RRM içeren protein
FS071280.1	TPX2 içeren protein
FS075057.1	TCP içeren protein
FS091461.1	DNA-bağımlı protein
FS081300.1	SBP-tip içeren protein
FS070808.1	Serin/treonin protein kinaz
FS084142.1	Ring tipi E3 ubiquitin transferaz



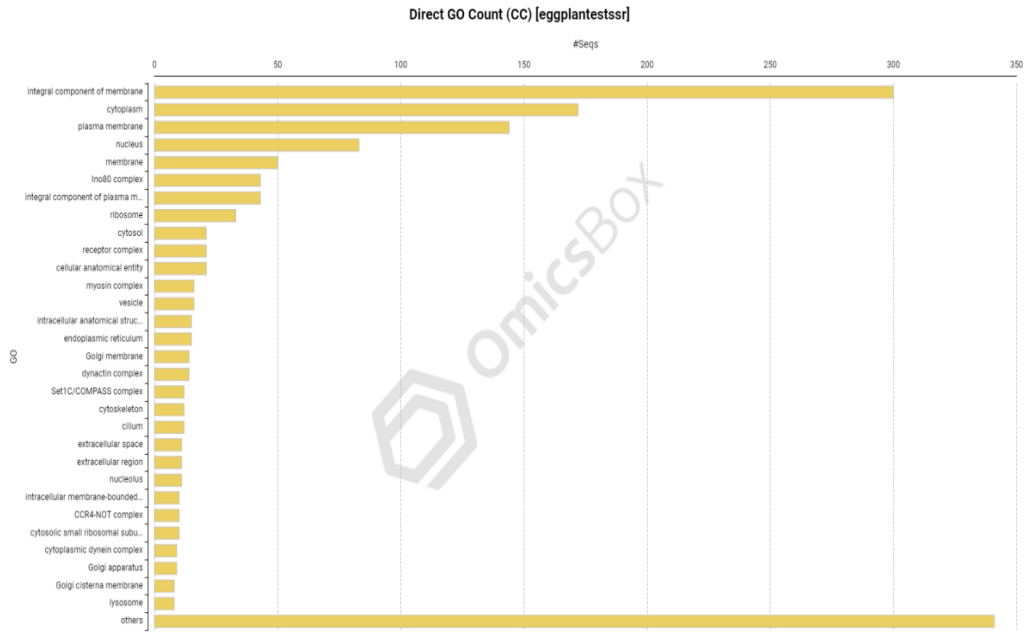
Şekil 4.13. EST bölgelerinin gen anotasyon dağılımı biyolojik süreç (BP), hücresel bileşen (CC) moleküler fonksiyon (MF). Yeşil, sarı ve mavi sırasıyla BP, CC ve MF kategorilerini temsil eder.



Şekil 4.14. Biyolojik süreç kısmında tanımlanan proteinlerin dağılımı



Şekil 4.15. Tanımlanan proteinlerin moleküler fonksiyon dağılımı.



Şekil 4.16. Tanımlanan proteinlerin hücresel bileşenlerin dağılımı.

Tanımlanan proteinler en fazla transkripsiyon düzenleme kısmında bulunmaktadır. Biyolojik süreçte ve gelişim süreçte birçok genin arttırılıp, azaltılması veya çalışıp çalışmamasını etkilediği için bu süreçte etkileşen proteinler bizim için değerlidir. Transkripsiyon düzenleme de ilişkili markörlerin bulunması patlıcan bitkisinin gen mekanizması hakkında birçok bilgi edinmemizi ve genetik mühendisliği adı altında birçok çalışmaya adım atabileceğimizi göstermektedir.

Aynı şekilde antioksidan aktivitesi birçok insan beslenmesinde savunma mekanizmasına katkıda bulunduğu için bitkisel bir gıda olan patlıcanda bulunan markörler besin değerini arttırmaya yönelik ıslah çalışmaları için büyük bir öneme sahiptir.

5. SONUÇLAR VE ÖNERİLER

Tez çalışması kapsamında patlıcan genomu, ncRNA'lar bakımından makine öğrenmesi algoritmaları ile analiz edilmiş olup literatürde ilk kez tüm genom düzeyinde ncRNA tahmin edilmiştir. Daha sonra tanımlanmış ncRNA lokuslarına yakın genomik bölgelerden SSR markörler adayları geliştirilmiş ve fonksiyonel olarak anlamlandırılmıştır. Ayrıca tüm genom düzeyinde *de novo* yaklaşımlar ile transpozon lokuslar tahmin edilerek markör tanımlama çalışmalarına tabi tutulmuştur. Patlıcan genomu son olarak hem transmembran protein sınıflandırılması için motif örüntüsü taranmış hem de tanımlı EST lokusları için EST-SSR markörleri geliştirilmiştir. Çalışma kapsamında 10000 üzerinde SSR markörü geliştirilmiş olup bu markörlerin yarısına yakını regülatör genomik lokuslardan üretilmiştir.

Çalışma kapsamında patlıcanın 12 kromozomunu yüksek yoğunlukta kapsayacak şekilde regülatör ve/veya fonksiyonel olma potansiyeli yüksek, polimorfik olma oranı yüksek olabilecek yüksek verimlilikte yeni moleküler markörler geliştirilmiştir. Bu markörlerin başta patlıcan gen kaynaklarının karakterizasyonunda kullanılabilecektir. Ayrıca genetik/agronomik özelliklerin haritalanmasında ve ilişkili markörlerin geliştirilerek markör destekli ıslah çalışmaları için kullanılmasına olanak sağlayacak yüksek miktarda tanımlı moleküler veri üretilmiştir.

Karakterize edilen TE'ler evrimsel çalışmalarda birçok bilgi sağlayacağı düşünülmektedir. Çünkü TE hem gen evrimi hem de fenotipik evrim açısından önemli olduğu söylenebilir. Bu sebeple de patlıcan genomunda karakterize edilen TE'lerinin yapılan filogenetik çalışmalarda kullanımının etkili olacağı tahmin edilmektedir.

Özellikle abiyotik stres toleransında bitki türleri arasında korunmuş miRNA'ların çalışma kapsamında patlıcan genomunda hem kromozomal lokalizasyonun hem de pre-miRNA genom sekanslarının tanımlanması bu önemli miRNA ailelerinin patlıcanda hedefli biçimde ekspresyon analizlerine tabi tutulabileceği öngörülmektedir. *In silico* olarak karakterize edilen miRNA'ların belirli bir abiyotik stres koşuluna toleranslı patlıcanlar için seçim yapmak ve bu stres koşuluna toleranslı patlıcan hatları geliştirmek için umut vadettiği düşünülmektedir.

Geliştirilen aday markörler çeşitlilik analizlerinde, filogenetik analizlerde, genetik haritalama gibi birçok patlıcan genetiği ve ıslahında önemli olan birçok çalışmada etkili

bir biçimde kullanılabileceği öngörülmektedir. Çalışma sonuçlarının patlıcan markör destekli ıslah çalışmalarına ileri de adapte edilebileceği düşünülmektedir.

KAYNAKLAR

- Afgan, Enis, Dannon Baker, Bérénice Batut, Marius Van Den Beek, Dave Bouvier, Martin Ech, John Chilton, et al. The Galaxy Platform for Accessible, Reproducible and Collaborative Biomedical Analyses.”Nucleic Acids Research 46 (W1), . 2018; W537–44.
- Arvas, Y. E., M. M. Abed, Q. A. Zaki, Kocacalışkan, and E. K. Haji. 2021. “The Potential Role of Transposable Elements as Molecular Markers.” IOP Conference Series: Earth and Environmental Science 761 (1).
- Bennetzen, Jeffrey L. 2000. “Transposable Element Contributions to Plant Gene and Genome Evolution Jeffrey.” Plant Molecular Biology 42: 251–69.
- Bhat;, R. S., K. Shirasawa, Y. Monden;, H. Yamashita, and M. Tahara. 2020. Legume Genomics Methods and Protocols Methods in Molecular Biology 2107. Legume Genomics. Methods in Molecular Biology.
- Classification, Transposable Elements. 2019.
- Makałowski2019_Protocol_TransposableElementsClassifica.Pdf. Vol. 1910.
- Dahariya, Swati, Indira Paddibhatla, Santosh Kumar, Sanjeev Raghuwanshi, Adithya Pallepati, and Ravi Kumar Gutti. 2019. “Long Non-Coding RNA: Classification, Biogenesis and Functions in Blood Cells.” Molecular Immunology 112 (January).
- Doganlar, Sami, Anne Frary, Marie Christine Daunay, Richard N. Lester, and Steven D. Tanksley. 2002. “A Comparative Genetic Linkage Map of Eggplant (*Solanum Melongena*) and Its Implications for Genome Evolution in the Solanaceae.” Genetics 161 (4): 1697–1711.
- Fattash, I., R. Rooke, A. Wong, C. Hui, T. Luu, P. Bhardwaj, and G. Yang. 2013. “Miniature Inverted-Repeat Transposable Elements: Discovery, Distribution, and Activity1.” Genome 56 (9): 475–86
- Flynn, Jullien M., Robert Hubley, Clément Goubert, Jeb Rosen, Andrew G. Clark, Cédric Feschotte, and Arian F. Smit. 2020. “RepeatModeler2 for Automated Genomic Discovery of Transposable Element Families.” Proceedings of the National Academy of Sciences of the United States of America 117 (17): 9451–57.
- Fort, Victoire, Gabriel Khelifi, and Samer M I Hussein. 2021. “BBA - Molecular Cell Research Long Non-Coding RNAs and Transposable Elements : A Functional Relationship.” BBA - Molecular Cell Research 1868 (1): 118837.

- Guidotti, G. 1972. "Membrane Proteins." *Annual Review of Biochemistry* 41: 731–52.
- Han, Jeffrey S. 2010. "Non-Long Terminal Repeat (Non-LTR) Retrotransposons: Mechanisms, Recent Developments, and Unanswered Questions." *Mobile DNA* 1
- Hirakawa, Hideki, Kenta Shirasawa, Koji Miyatake, Tsukasa Nunome, Satomi Negoro, Akio Ohyama, Hirotaka Yamaguchi, et al. 2014. "Draft Genome Sequence of Eggplant (*Solanum Melongena* L.): The Representative *Solanum* Species Indigenous to the Old World." *DNA Research* 21 (6): 649–60.
- Hou, Jinna, Dandan Lu, Annaliese S. Mason, Baoquan Li, Meili Xiao, Sufang An, and Donghui Fu. 2019. "Non-Coding RNAs and Transposable Elements in Plant Genomes: Emergence, Regulatory Mechanisms and Roles in Plant Development and Stress Responses." *Planta* 250 (1): 23–40.
- Jones, David T. 2007. "Improving the Accuracy of Transmembrane Protein Topology Prediction Using Evolutionary Information." *Bioinformatics* 23 (5): 538–44.
- Käll, Lukas. 2010. "Prediction of Transmembrane Topology and Signal Peptide given a Protein's Amino Acid Sequence." *Methods in Molecular Biology* (Clifton, N.J.) 673: 53–62.
- KALLOO, G. 1993. *Eggplant. Genetic Improvement of Vegetable Crops*. Pergamon Press Ltd.
- Kidwell, Margaret G. 2005. "Transposable Elements." *The Evolution of the Genome*, 165–221.
- Kozomara, Ana, and Sam Griffiths-Jones. 2014. "MiRBase: Annotating High Confidence MicroRNAs Using Deep Sequencing Data." *Nucleic Acids Research* 42 (D1): 68–73.
- Krogh, Anders, Björn Larsson, Gunnar Von Heijne, and Erik L.L. Sonnhammer. 2001. "Predicting Transmembrane Protein Topology with a Hidden Markov Model: Application to Complete Genomes." *Journal of Molecular Biology* 305 (3): 567–80.
- Kumar, Ashish, Vinay Sharma, Bharat Taindu Jain, and Prashant Kaushik. 2020. "Heterosis Breeding in Eggplant (*Solanum Melongena* L.): Gains and Provocations." *Plants* 9
- Lee, Sung-Il, and Nam-Soo Kim. 2014. "Transposable Elements and Genome Size Variations in Plants." *Genomics & Informatics* 12 (3): 87.

- Li, Aimin, Junying Zhang, and Zhongyin Zhou. 2014. "PLEK: A Tool for Predicting Long Non-Coding RNAs and Messenger RNAs Based on an Improved k-Mer Scheme." *BMC Bioinformatics* 15 (1): 1–10.
- Li, Shitian, Muthusamy Ramakrishnan, Kunnummal Kurungara Vinod, Ruslan Kalendar, Kim Yrjälä, and Mingbing Zhou. 2020. "Development and Deployment of High-Throughput Retrotransposon-Based Markers Reveal Genetic Diversity and Population Structure of Asian Bamboo." *Forests* 11 (1): 1–25.
- Macher, Bruce A., and Ten Yang Yen. 2007. "Proteins at Membrane Surfaces - A Review of Approaches." *Molecular BioSystems* 3 (10): 705–13.
- Panni, Simona, Ruth C. Lovering, Pablo Porras, and Sandra Orchard. 2020. "Non-Coding RNA Regulatory Networks." *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms* 1863 (6): 194417.
- Patil, Prakash G., N. V. Singh, Shilpa Parashuram, Abhishek Bohra, Dhanajay M. Mundewadikar, Vipul R. Sangnure, K. Dhinesh Babu, and Jyotsana Sharma. 2020. "Genome Wide Identification, Characterization and Validation of Novel MiRNA-Based SSR Markers in Pomegranate (*Punica Granatum* L.)." *Physiology and Molecular Biology of Plants* 26 (4): 683–96.
- Rebollo, Rita, Mark T. Romanish, and Dixie L. Mager. 2012. "Transposable Elements: An Abundant and Natural Source of Regulatory Sequences for Host Genes." *Annual Review of Genetics* 46 (July): 21–42.
- Roy, Neha Samir, Ji Yeong Choi, Sung Il Lee, and Nam Soo Kim. 2014. "Marker Utility of Transposable Elements for Plant Genetics, Breeding, and Ecology: A Review." *Genes and Genomics* 37 (2): 141–51.
- Sharma, Shweta, Akhilesh Sharma, and Viveka Katoch. 2020. "Biotechnological Interventions in Eggplant (*Solanum Melongena* L.)." *Journal of Horticultural Science and Biotechnology* 95 (3): 273–85.
- Tan, Sandra, Tong Tan Hwee, and Maxey C.M. Chung. 2008. "Membrane Proteins and Membrane Proteomics." *Proteomics* 8 (19): 3924–32.
- Venkatesh, and B. Nandini. 2020. "Miniature Inverted-Repeat Transposable Elements (MITEs), Derived Insertional Polymorphism as a Tool of Marker Systems for Molecular Plant Breeding." *Molecular Biology Reports* 47 (4): 3155–67.
- Von Heljne, G. (1998). Life and death of a signal peptide. *Nature*, 111-113.

- Wang, Xuwen, and Le Wang. 2016. "GMATA: An Integrated Software Package for Genome-Scale SSR Mining, Marker Development and Viewing." *Frontiers in Plant Science* 7 (September): 1–11.
- Weerden, Gerard M. Van Der, and Gerard W.M. Barendse. 2007. "A Web-Based Searchable Database Developed for the EGGNET Project and Applied to the Radboud University Solanaceae Database." *Acta Horticulturae* 745 (June 2007): 503–6.
- Wicker, Thomas, François Sabot, Aurélie Hua-Van, Jeffrey L. Bennetzen, Pierre Capy, Boulos Chalhoub, Andrew Flavell, et al. 2007. "A Unified Classification System for Eukaryotic Transposable Elements." *Nature Reviews Genetics* 8 (12): 973–82.
- Yadav, Chandra Bhan, Venkata Suresh Bonthala, Mehanathan Muthamilarasan, Garima Pandey, Yusuf Khan, and Manoj Prasad. 2015. "Genome-Wide Development of Transposable Elements-Based Markers in Foxtail Millet and Construction of an Integrated Database." *DNA Research* 22 (1): 79–90.
- Yepuri, Vijay, Saakshi Jalali, Nagesh Kancharla, V. B. Reddy, and S. Arockiasamy. 2020. "Development of Genome Wide Transposable Elements Based Repeat Junction Markers in Jatropha (*Jatropha Curcas* L.)." *Molecular Biology Reports* 47 (7)
- Fu, D. M., & An, Z. (2013). MicroRNA-based molecular markers: a novel PCR-based genotyping technique in Brassica species. *Plant Breeding*,.
- Ganie, S. A., & Mondal, T. K. (2015). Genome-wide development of novel miRNA-based microsatellite markers of rice (*Oryza sativa*) for genotyping applications. *Molecular Breeding*.
- Xue, M., Zhuo, Y., & Shan, B. (2017). MicroRNAs, Long Noncoding RNAs, and Their Functions in Human Disease. *Methods in Molecular Biology*.
- You, F. M. (2010). RJPrimers: Unique transposable element insertion junction discovery and PCR primer design for marker development. *Nucleic Acids Research*, 38.
- Mazaheri, M. P. (2014). Transposable Element Junctions in Marker Development and Genomic Characterization of Barley. *The Plant Genome*, 1-8.