

***Arabidopsis thaliana* Genomu Üzerinde LTR Retrotranspozonların Yerlerinin Belirlenmesi**

Mehmet GEZGİNCİ^{1*}Matej LEXA²¹ Gaziantep Üniversitesi Fen-Edebiyat Fakültesi Biyoloji Bölümü, Gaziantep² Masaryk University Computer Science Department, Brno-Czech Republic

*Sorumlu yazar:

E-mail: mehmetgezginci@gmail.com

Geliş Tarihi: Ocak 24, 2015

Kabul Tarihi: Mart 06, 2015

Özet

Transpozonlar DNA üzerinde bulunan ve hareket halinde olan genetik elementlerdir. Bu elementler organizmanın hastalanmasına neden olabileceği gibi istenilen özelliklerin daha da çok salgılanmasını sağlayabilir. *Arabidopsis thaliana* yaygın yaşam alanına sahip, genomu küçük ve tek yıllık bir bitkidir. Bu özelliklerinden dolayı tüm genomunun genetik sekans yapılmıştır ve bu genetik sekansa internet aracılığı ile erişmek mümkündür. Bu çalışmanın amacı *Arabidopsis thaliana* bitkisi üzerinde bulunan LTR retrotranspozonların yerlerinin belirlenmesi ve görüntülenmesidir. Çalışma için *Arabidopsis thaliana*'ya ait olan genetik sekanslar ve bilinen transpozonlar internet kullanılarak indirilmiş, bunlar arasında eşleşme olup olmadığını belirlemek için BLAST'lama işlemi gerçekleştirilmiştir. BLAST'lama sonucu elde edilen veriler, *Arabidopsis thaliana* üzerinde transpozonların varlığını göstermektedir. BLAST'lama sonuçları Sequel Pro programında oluşturulmuş veritabanına aktarılmıştır. Bu veritabanı kullanılarak Python programlama dilinde algoritma yazılmıştır. Elde edilen çıktı Integrative Genomics Viewer programına aktarılmıştır. Integrative Genomics Viewer sonuçları ile *Arabidopsis thaliana*'ya ait beş kromozom üzerinde de transpozonların olduğu görüntülenmiştir. Elde edilen görüntüler sağlıklı bir *Arabidopsis thaliana*'ya ait olduğu için bu görüntülerden yola çıkılarak, transpozon kaynaklı hastalıklar belirlenebilir, gereken tedavi yöntemleri de geliştirilebilir.

Anahtar Kelimeler: *Arabidopsis thaliana*, Biyoinformatik, LTR Retrotranspozon, Python, Integrative Genomics Viewer

Finding and Visualization of LTR Retrotransposons in *Arabidopsis thaliana*

Abstract

Transposons are mobile genetics elements found in DNA. These elements could contribute to cause diseases or produce desired genes more. *Arabidopsis thaliana* has a widespread, small genome and it is an annual plant. Thanks to these properties scientists sequenced all genome of *A. thaliana*. It is easy to access the sequences via internet. In this study we aimed to find and visualize LTR retrotransposons in *A. thaliana*. For the study, all genome sequences belonging to *A. thaliana*, and known LTR retrotransposons downloaded using Internet. To prove existence of transposons in *A. thaliana*, BLAST was done between the genome sequences and the LTR retrotransposons. Outputs of BLAST were transferred to database created in Sequel Pro application. Using the database algorithms was written in Python. Output of the algorithms was used in Integrative Genomics Viewer application to visualize retrotransposon in *A. thaliana*. As a result it was proven all chromosomes of *A. thaliana* had retrotransposons LTR. This information could allow scientists to develop new treatment techniques.

Keywords: *Arabidopsis thaliana*, Bioinformatics, LTR Retrotransposons, Python, Integrative Genomics Viewer

GİRİŞ

Arabidopsis thaliana, sistematikte lahağiller (*Brassicaceae*) ailesinde yer alan, fare kulağı teresi olarak bilinen tek yıllık, çiçekli bir bitkidir (1) (Resim 1.). İnsan Genom Projesi (İGP) çalışmaları sırasında farklı organizmaların da genom projeleri yapılmıştır. *Arabidopsis thaliana* da, geniş yetiştirme sahasına sahip, çok tohum veren, mutant çeşitleri olan, az kromozomlu bir bitki olduğundan genom sekansı yapılmış ve model bir organizmalardan birisi olmuştur (2). Bütün genomu sekanslanmış olup, internet üzerinden bu verilere ulaşmak hızlı ve de ücretsizdir.

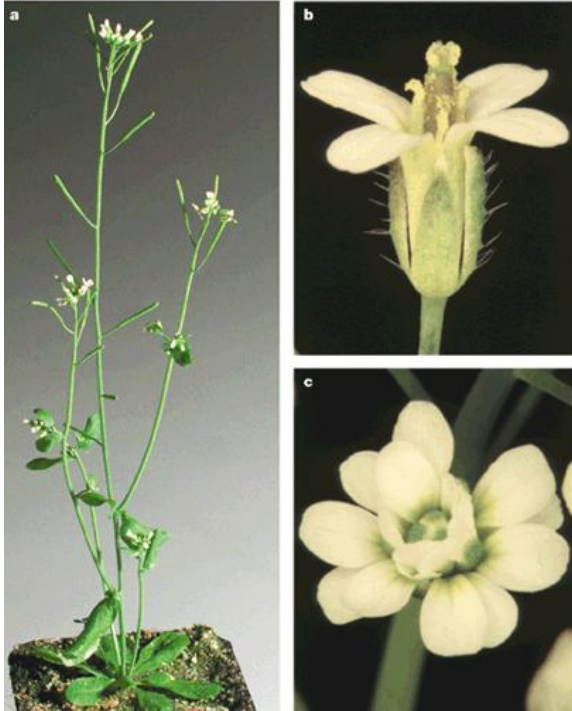
Transpozabl elementler (TE's) bir hücrenin genomunda hareket edebilen "sıçrayıcı genler" olarak bilinirler. Bu elementler özellikle gen aktivitesini kontrol ederek genomun evrimleşmesi üzerine önemli etkileri vardır (3). Çalışılan bütün ökaryotlarda hemen hemen varlığı gösterilmiştir (4). Gelişmiş organizmaların genomlarında oldukça fazla bulunan TE's, tekrar eden DNA dizisine sahiptirler. İnsan, pirinç, fare genomunda %40'dan fazla

TE's bulunurken, bakteri ve basit ökaryotlarda bu oran %5'in altındadır (5, 6). Transpozonlar tipik olarak DNA Transpozonları ve Retroelementler olarak sınıflandırılabilirler (Resim 2. (7)).

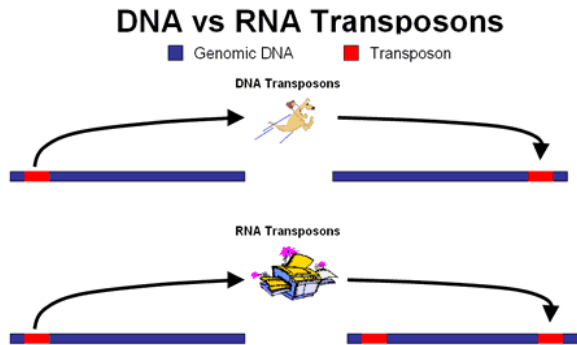
Retroelementler "kopyala-yapıştır" mekanizması ile bir yerden başka yere hareket ederlerken, DNA transpozonları "kes-yapıştır" mekanizmasıyla yer değiştirirler (Resim 3.). Bundan dolayı DNA transpozonları genom üzerinde daha az miktarda bulunurlar.

1953 yılında DNA'nın molekül yapısının belirlenmesinden sonra moleküler biyoloji alanında büyük gelişmeler yaşanmıştır. Yeni nesil dizilime teknolojileri sayesinde birçok canlıya ait bütün genom verileri kolayca dizilenebilmektedir (9). Dizilene sonucunda elde edilen bilgilerin yorumlanması büyük sorun haline gelmiştir (Resim 4.)(10). Bu sorunu çözebilmek için biyoloji bilimi, bilişim teknolojilerine ihtiyaç duymuştur. Bu iki bilimin birleşmesi sonucu biyoinformatik adı verilen yeni bir bilim dalı doğmuştur (11). Biyoinformatiğin temel hedefleri, artan bilgiye kolayca ulaşabilmek için bu bilgileri toplamak, yönetmek ve biyoloji biliminde yaşanan problemleri çözmektir (12).

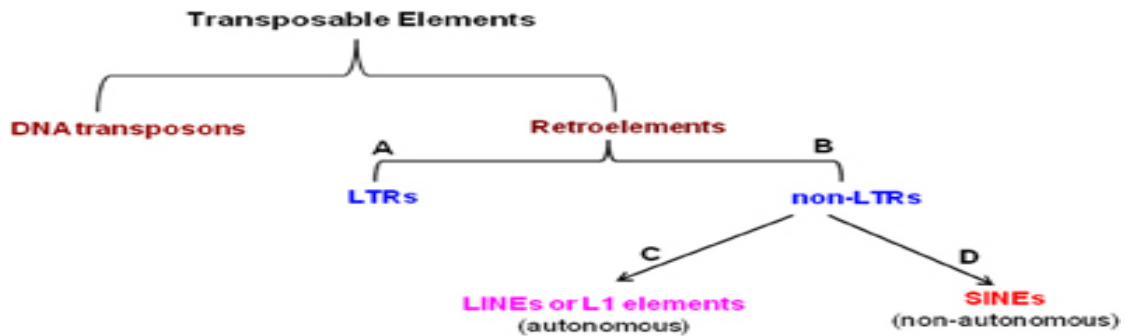
Bu çalışmada amaç, tüm genomu sekanslanmış olan *Arabidopsis thaliana*'ya ait olan LTR retrotranspozonların, Python yazılım dili kullanılarak yerlerinin belirlenmesi ve Integrative Genomics Viewer (IGV) programı ile yerlerinin görüntülenmesidir.



Resim 1. *Arabidopsis thaliana*'ya ait görüntüler



Resim 2. DNA ve RNA transpozonlarının (Retrotranspozonlar) transpozisyon mekanizması. DNA transpozisyon sırasında kendini çoğaltmaz (kes-yapıştır), sayısı sabit kalırken; Retrotranspozonlar kendilerini çoğaltırlar (kopyala-yapıştır), sayıları artar.(8)



Resim 4. Transpozabl Elementlerin sınıflandırılması.

veri tabanı	saklanan veri tipi	dizi sayısı	1986	2001
EMBL-GENBANK	Nükleotid dizileri	9978	14976310	
SWISSPROT	Protein dizileri	4060	114033	
PDB	Protein yapısı	296	16973	

Resim 3. Yeni nesil dizileme araçlarının artması sonucu elde edilen dizi sayısı miktarının artışı

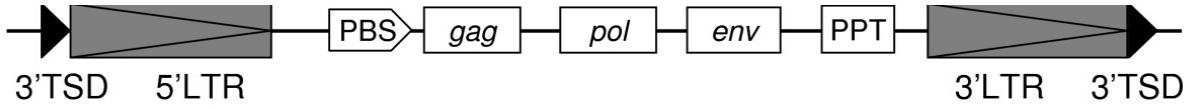
MATERYAL METOT

Bu çalışmada kullanılan *Arabidopsis thaliana*'ya ait olan tüm genom dizileri (ftp://ftp.ncbi.nlm.nih.gov/genomes/Arabidopsis_thaliana/) sitesinden indirilmiştir. LTR retrotranspozonları ise Gypsy Database 2 veritabanından (<http://gydb.org/index.php/Ty3/Gypsy>) indirilmiştir.

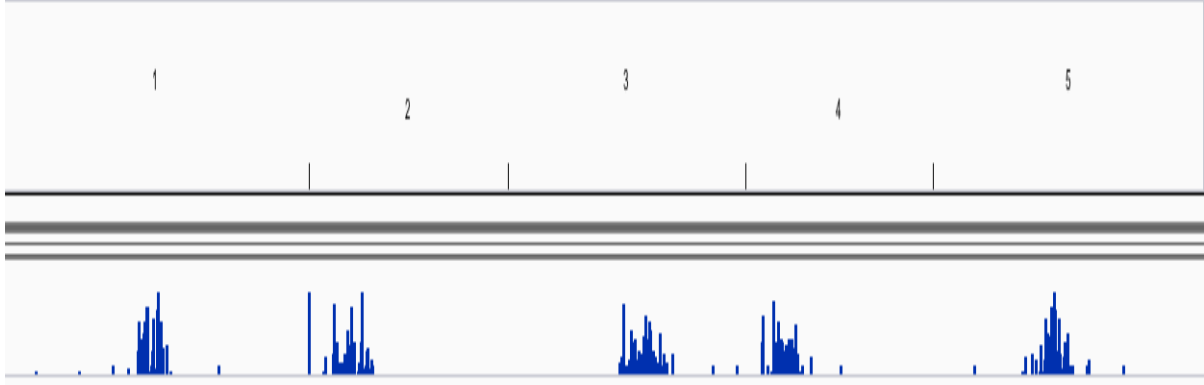
LTR retrotranspozonların *Arabidopsis thaliana* 'da varlığını gösterebilmek için BLAST'lama işlemi gerçekleştirilmiştir. BLAST sonuçları Sequel Pro veri tabanı programına aktarılmıştır. LTR retrotranspozonları 5' LTR ve 3' LTR uçları ile bu uçlar arasında yer alan proteinlerden oluşmaktadır (Resim 5.(13).) Bu yapıdan dolayı Python programlama dilinde algoritma yazılrlen önce LTR retrotranspozonlara ait 5' LTR uçları bulunmuştur. Her bir 5' LTR uçları için aynı isimli aynı kromozomda yer alan 3' LTR uçları bulunmuştur. En son olarak 5' LTR uçları ile 3' LTR uçları arasında kalan aynı isimli LTR proteinleri bulunmuştur. Bu üç kısım birleştirilerek algoritmanın çıktısı alınmıştır. IGV programı gff3 uzantılı çıktılarını kabul ettiğinde elde edilen Python çıktısının uzantısı (<http://www.sequenceontology.org/resources/gff3.html>) internet adresindeki yönergerler kullanılarak gff3 formatına dönüştürülmüştür. Elde edilen bu sonuçlar IGV programına aktarılmıştır.

SONUÇLAR

Python programlama dili kullanarak yazılan algoritma sonucu *Arabidopsis thaliana*'nın LTR retrotranspozonlara sahip olduğu bulunmuştur. Bu retrotranspozonların yerlerinin görüntülenmesi için kullanılan IGV sonuçları *Arabidopsis thaliana*'nın bütün kromozomlarında LTR retrotranspozonların varlığını göstermiştir. 1., 3., ve 5. kromozomlarda bu retrotranspozonlar daha çok sentromer bölgelerinde kümelenmiş olarak bulunmuşken, 2. ve 4. kromozomlarda telomer bölgelerinde buldukları görüntülenmiştir (Resim 6.).



Resim 5. LTR Retrotranspozonlarının yapısı.



Resim 6. IGV görüntü sonuçları.

KAYNAKLAR

- [1] Çalış Ö, Yazar Ç. 2011. Fare Kulağı Teresi, *Arabidopsis thaliana*'da Konukçu Dışı Dayanıklılığın Erken Yanıklık Hastalık Etmeni *Alternaria solani*'nin Kontrolü İçin Araştırılması. Gaziosmanpaşa Üniversitesi Ziraat Fakültesi Dergisi 2011.
- [2] Türk FM. 2006. Konukçu-Patojen İlişkisinde Model Bir Bitki: *Arabidopsis thaliana*/A Model Plant In Host-Pathogen Interaction: *Arabidopsis thaliana*. Journal of the Faculty of Agriculture 37.
- [3] Biémont C, Vieira C. 2006. Genetics: junk DNA as an evolutionary force. *Nature* 443:521-524.
- [4] Quesneville H, Bergman CM, Andrieu O, Autard D, Nouaud D, Ashburner M, Anxolabehere D. 2005. Combined evidence annotation of transposable elements in genome sequences. *PLoS Comput Biol* 1:166-175.
- [5] Curcio MJ, Derbyshire KM. 2003. The outs and ins of transposition: from mu to kangaroo. *Nature Reviews Molecular Cell Biology* 4:865-877.
- [6] Kalyanaraman A, Aluru S. 2006. Efficient algorithms and software for detection of full-length LTR

retrotransposons. *Journal of bioinformatics and computational biology* 4:197-216.

[7] Ramulu HG, Raoult D, Pontarotti P. 2012. The rhizome of life: what about metazoa? *Frontiers in cellular and infection microbiology* 2.

[8] Bacha J 2005, posting date. Transposons: Spam from the Dark Ages. [Online.]

[9] Üstek D, Abacı N, Sırma S, Çakiris A. 2011. Yeni Nesil DNA Dizileme (New Generation DNA Sequencing). *DeneySEL Tıp Araştırma Enstitüsü Dergisi* 1:11-18.

[10] Atalay RÇ. 2002. NEDEN BİYOİNFORMATİK? *Avrasya Dosyası, Moleküler Biyoloji ve Gen Teknolojileri Özel, Sonbahar* 8.

[11] Elen A, Turan MK. 2013. "Genetik algoritmalar ve çoklu dizi hizalama probleminin çözümü", 26-30 Ekim 2013, XIII Tıbbi Biyoloji ve Genetik Kongresi, Aydın

[12] Akın AC, Bürçe B, Çevirici B, Şahin B, Şahin E, Şahin Y. *Disiplinler Arası Bir Bilim Dalı: Biyoinformatik*.

[13] Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC bioinformatics* 9:18.