



Some Software's Used in DNA Sequencing Based Haplotype Analysis[#]

Emel Tüten Sevim^{1,a}, Kemal Karabağ^{1,b,*}

¹Department of Agricultural Biotechnology, Faculty of Agriculture, Akdeniz University, 07070 Antalya, Turkey

*Corresponding author

ARTICLE INFO	ABSTRACT
<p>[#]This study was presented as an oral presentation at the 4th International Anatolian Agriculture, Food, Environment and Biology Congress (Afyonkarahisar, TARGID 2019)</p> <p><i>Review Article</i></p> <p>Received : 13/06/2019 Accepted : 20/11/2019</p> <p>Keywords: DNA sequencing SNP Haplotypes Software Phylogenetic</p>	<p>Determination of genetic resources and their diversity constitutes the initial step of the breeding programs. Before starting the selection of the desired properties, many molecular methods are used to determine the existing genetic potential and determine the genotypic values. Many molecular methods (RAPD, RFLP, AFLP, SSR and DNA sequencing etc.) based on PCR technology are used to identify populations, identify genes related to yield and resistance, and demonstrate phylogenetic relationships. Many software's for the analysis of molecular datasets obtained from molecular methods have been developed and the field of bioinformatics was born. In this study, after DNA sequence analysis; it is aimed to give information about haplotype analysis by determining SNPs (Single Nucleotide Polymorphism's) in individuals of populations, calculation of the genetic distance among populations and within populations, and software and analysis used for phylogenetic tree drawing.</p>

Türk Tarım – Gıda Bilim ve Teknoloji Dergisi 7(sp1): 107-109, 2019

DNA Dizileme Temelli Haplotip Analizlerinde Kullanılan Bazı Yazılımlar

MAKALE BİLGİSİ	ÖZ
<p><i>Derleme Makale</i></p> <p>Geliş : 13/06/2019 Kabul : 20/11/2019</p> <p>Anahtar Kelimeler: DNA dizileme SNP Haplotip Yazılım Filogenetik</p>	<p>Genetik kaynakların ve çeşitliliğin belirlenmesi ıslah programlarının ilk adımını teşkil eder. İstenilen özelliklerin seleksiyonuna başlamadan önce mevcut genetik potansiyelin belirlenmesini ve genotipik değerlerin tespitini sağlayan birçok moleküler yöntem kullanılmaktadır. PCR teknolojisine dayalı birçok moleküler yöntem (RAPD, RFLP, AFLP, SSR ve DNA dizi analizi vb.) kullanılarak popülasyonlarının tanımlanması, verim ve direnç ile ilgili genlerin belirlenmesi, filogenetik ilişkilerin ortaya konulması gibi konularda yoğun çalışmalar sürdürülmektedir. Buna paralel olarak moleküler yöntemlerden sağlanan veri setlerinin analizleri için birçok yazılım geliştirilmiş ve biyoinformatik bilim alanı doğmuştur. Bu çalışmada, DNA dizi analizleri sayesinde SNP'lerin (Single Nucleotide Polymorphism's=Tek Nükleotid Polimorfizm) belirlenerek haplotip analizlerinin yapılması, popülasyonlar arası ve popülasyonlar içi genetik uzaklıkların hesaplanması ve filogenetik ilişkilerin belirlenmesi için kullanılan bazı yazılımlar ve analizleri hakkında bilgi verilmesi amaçlanmıştır.</p>

^a tutensevim@akdeniz.edu.tr ^b <https://orcid.org/0000-0002-3105-6416> | karabag@akdeniz.edu.tr ^c <https://orcid.org/0000-0002-4516-6480>



Giriş

Hayvan biyoteknolojisinde filogenetik ilişkilerin belirlenmesi sırasında popülasyonlardaki bireylere ait ilgili DNA bölgesindeki dizilerde görülen tek nükleotid polimorfizmlerden (SNP) yararlanılabilmektedir. Polimorfizmlerin belirlenmesinde “FinchTV 1.4.0”, “ChromasPro v2.1.2” gibi yazılımlar kullanılır. Sonrasında popülasyonlar arasındaki ilişkilerin belirlenmesi için farklı analizler yapılmaktadır. Bu analizler için “DnaSP v5.10” (Rozas ve ark., 2010), “Arlequin v3.11” (Excoffier ve ark., 2006), “SplitsTree v4.14.6” (Huson ve Bryant, 2006), “NETWORK v5.0.1.1” (Bandelt ve ark., 1999) yazılımları kullanılabilmektedir. Yapılan bu çalışma da filogenetik ilişkilerin belirlenmesi için kullanılan bu yazılımlar hakkında bilgi verilmesi amaçlanmıştır.

DNA Nükleotid Dizi Analizleri

İlgili DNA bölgesine spesifik primerler ile gerçekleştirilen polimeraz zincir reaksiyonu (PCR) sonucunda yükseltgenen ürünler öncelikle reaksiyonun başarısını kontrol etmek amacıyla jel elektroforezine tabi tutulur. Jel elektroforezinde DNA fragment büyüklüklerinin tespiti için DNA ladder (büyüklük markeri) kullanılır. Doğru bölgenin yükselttiği belirlendikten sonra DNA dizisini oluşturan nükleotidlerin sırası oluşturulur. Genellikle Sanger dizilime yöntemine dayalı “Applied Biosystems 3500XL Genetic Analyzer” gibi cihazlar kullanılarak yapılan dizileme ile çalışılan gen bölgesindeki nükleotid dizilimlerini gösteren pikler elde edilir. Bu dizilere ait pik görüntüleri “FinchTV 1.4.0” veya “ChromasPro v2.1.2” yazılımında okunarak analiz başarısı kontrol edilir. Burada, dizileme cihazından kaynaklanan hatalı okunan bölgeler varsa düzeltilir. Değerlendirilmeye alınacak ve alınmayacak dizileme sonuçları tespit edilir. Kötü DNA dizileri analizden çıkartılır. Değerlendirilebilir sonuçlar veren dizilerin çalışılan gen bölgelerine ait olup olmadığı “BLAST 2.2.20” (Zhang ve ark., 2000) yazılımı ile kontrol edilir. PCR çalışmalarında kullanılan özgün primerlerin doğru sonuçlar verdiğinin kesinleştirilmesinden sonra her bir gen bölgesi için tüm dizi bilgileri “MEGA 6” (Tamura ve ark., 2013) ve benzeri yazılımlara yüklenerek DNA dizi hizalaması (alignment) yapılır. Hizalama yapıldıktan sonra oluşturulan dosya sonraki analizlerde kullanılabilir formatta “.meg” uzantılı olarak kaydedilir.

Haplotip Analizleri

Hizalaması yapılan “.meg” uzantılı dosya formatındaki diziler “DnaSP v5.10” (Rozas ve ark., 2010) yazılımına yüklenerek SNP bölgeleri belirlenir ve haplotipler oluşturulur. Bu yazılımda yapılan analiz sonucunda 3 farklı formatta dosya elde edilir. Bunlar sırası ile; “.nex” (nexus haplotip dosyası), “.arp” (Arlequin dosyası) ve “.rdf” (roehl dosyası) uzantılı dosyalardır. Bunlardan Arlequin dosyası bir popülasyonda hangi haplotipte bireyler görüldüğü ve her bir haplotipte o popülasyondan kaç birey olduğu bilgilerini içerir ve moleküler varyans analizi (AMOVA) için gereklidir. Bu Arlequin dosyanın düzenlenmesi için her bir haplotipteki bireyleri gösteren nexus haplotip dosyası gereklidir. Roehl dosyası ise;

haplotiplerde görülen SNP bölgelerinin kaçınıcı bazda olduğunu gösterir ve haplotip dağılımı görselleştirmek istenirse “NETWORK v5.0.1.1” yazılımında median-joining ağacı çizmek için gerekli formattadır. Ayrıca bu yazılım ile popülasyonların haplotip çeşitliliği (h), nükleotid çeşitliliği (π) ile popülasyonlar arasındaki genetik uzaklık (γ_{st} , Gammast) hesaplanabilmektedir.

Genetik Varyasyon Parametrelerinin Hesaplanması

Genetik varyasyon parametrelerinin hesaplanmasında “Arlequin v3.11” yazılımı kullanılmaktadır. Bu yazılımda “.arp” uzantılı dosya kullanılarak istenilen analizler yapılabilmektedir. Bu analizler sonucunda moleküler varyans analizi (AMOVA: Analysis of Molecular Variance) yapılır; popülasyonlara ait gen akış değerleri (F_{st}), popülasyonlar arası ve popülasyonlar içi genetik varyasyon ile popülasyonların ikiyeşerli gen akış değerleri (Pairwise F_{st}) matris tablosu elde edilir. Bunlardan Pairwise F_{st} matris tablosu filogenetik ağaç çizimi için gerekmektedir.

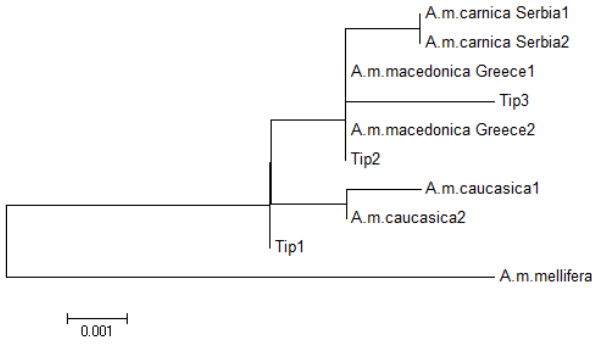
Filogenetik İlişkilerin Belirlenmesi

Çalışılan popülasyonlara ait filogenetik ilişkileri belirlenmesinde “SplitsTree v4.14.6” yazılımı kullanılmaktadır. Bu yazılım için popülasyon isimlerini ve popülasyonlara ait ikiyeşerli gen akış değerleri matris tablosunu içeren metin belgesi hazırlanması gerekmektedir. Metin belgesi içerisinde öncelikle popülasyonlar tanımlanmalıdır. Popülasyon isimleri sırasıyla alt alta yazılmalı ve numaralandırılmalıdır. Sıralamanın AMOVA analizinde elde edilen matris tablosu ile uyumlu olması gerekmektedir. Popülasyon numaraları köşeli parantez içerisinde belirtilmelidir (Ör: [1]). Popülasyon isimleri ise tek tırnak (‘) içerisinde yazılmalıdır (Ör: ‘DEMRE’).

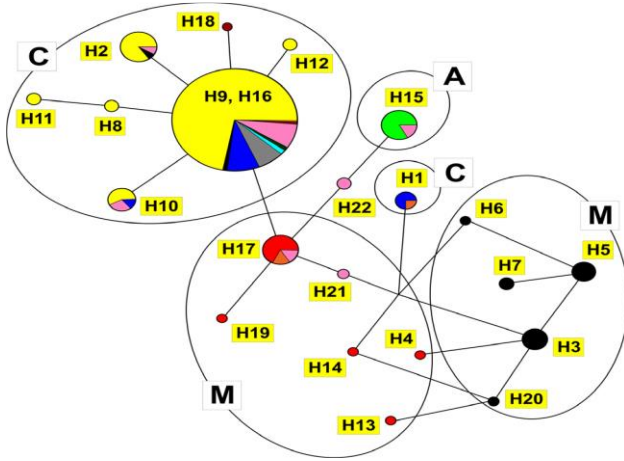
Popülasyonların ikiyeşerli gen akış değerleri matris tablosu yazılırken popülasyon isimleri tanımlanırken olduğu gibi her bir değer önüne ilgili popülasyonun ismi ve numarası yazılmalıdır (Ör: [1] ‘DEMRE’). Burada dikkat edilmesi gereken nokta metin belgesinde popülasyonlar tanımlanırken isimleri nasıl yazıldı ise matris tablosuna da o şekilde yazılması gerektiğidir. Eğer üst kısımda popülasyon ismi küçük harfle yazıldı ise alt kısımda da küçük harfle yazılmalı ve popülasyon sıralamaları her iki kısımda da aynı olmalıdır.

“SplitsTree v4.14.6” yazılımına uygun olarak hazırlanan metin belgesi kullanılarak popülasyonların filogenetik ilişkilerini ortaya koyan dendogram çizilebilmektedir. Bu yazılımda farklı metodlarla (UPGMA, Neighbour-Joining gibi) farklı yapıda (rooted, unrooted gibi) filogenetik ağaç oluşturulabilmektedir (Şekil 1).

Ayrıca haplotip dağılımını gösteren Median-Joining Tree çizilmek istenirse “NETWORK v5.0.1.1” yazılımı kullanılmaktadır (Şekil 2). Özetle hayvan biyoteknolojisinde SNP haplotip analizleri ile popülasyonlar arasındaki filogenetik ilişkilerin belirlenmesinde elde edilmek istenen bilgiye göre çeşitli yazılımlar kullanılmaktadır.



Şekil 1 SplitsTree yazılımında Neighbour-Joining metoduna göre çizilen filogenetik ağaç (Ünal, 2016)



Şekil 2 NETWORK yazılımında çizilen Median-Joining Tree (Péntek Zakar ve ark., 2015)

Sonuç

Hayvan biyoteknolojisinde SNP'ler popülasyonlar arasındaki filogenetik ilişkilerin belirlenmesi için sığır (Goud ve ark., 2019), manda (Li ve ark., 2018), bal arısı (Özdil ve İlhan, 2009), tavuk (Godinez ve ark., 2019), keçi (do Prado Paim ve ark.), koyun (Zhou ve ark., 2019) gibi birçok türde yaygın olarak kullanılmaktadır. Çalışılan bölgeye spesifik primerler ile polimeraz zincir reaksiyonları gerçekleştirilen ürünlerden dizileme analizi yapılarak ilgili bölgedeki tek nükleotid polimorfizmleri belirlenebilmektedir. Bunların belirlenmesinde "FinchTV 1.4.0" yazılımı kullanılabileceği gibi "ChromasPro v2.1.2" yazılımı da kullanılabilmektedir.

Tek nükleotid polimorfizmleri ile belirlenen haplotiplerden yararlanılarak popülasyonlara ait genetik varyasyonlar ve popülasyonlar arası filogenetik ilişkiler belirlenebilmektedir. Bu amaçla elde edilmek istenen bilgiye göre kullanılabilecek birçok yazılım bulunmaktadır. Örnek olarak popülasyonlara ait Fst matris tablosu oluşturulması için "Arlequin v3.11" yazılımı kullanılırken, haplotip çeşitliliği gibi parametreler için "DNASP v5.10" yazılımı kullanılmaktadır. Bu yazılımlardan elde edilen

sonuçlar görselleştirilmek istenirse "SplitsTree v4.14.6" yazılımı ile popülasyonlar arasındaki ilişkiyi gösteren dendrogram oluşturulabilmektedir. Ayrıca haplotip dağılımını gösteren Median-Joining Tree çizilmek istenirse "NETWORK v5.0.1.1" yazılımı kullanılmaktadır. Özetle hayvan biyoteknolojisinde SNP haplotip analizleri ile popülasyonlar arasındaki filogenetik ilişkilerin belirlenmesinde elde edilmek istenen bilgiye göre çeşitli yazılımlar kullanılmaktadır.

Kaynaklar

- Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, 16: 37-48.
- do Prado Paim T, Faria DA, McManus C, Lanari MR, Esquivel LC, Cascante MI, Mezzadra CA. 2019. New world goat populations are a genetically diverse reservoir for future use. *Scientific reports*, 9(1): 1476.
- Excoffier L, Laval G, Schneider S. 2006. ARLEQUIN version 3.01: an integrated software package for population genetics data analysis. University of Bern, Institute of Zoology, Switzerland. <http://cmpg.unibe.ch/software/Arlequin3>
- Godinez CJP, Nishibori M, Matsunaga M, Espina DM. 2019. Phylogenetic Studies on Red Junglefowl (*Gallus gallus*) and Native Chicken (*Gallus gallus domesticus*) in Samar Island, Philippines using the Mitochondrial DNA D-Loop Region. *The Journal of Poultry Science*, 0180131.
- Huson DH, Bryant D. 2006. Application of Phylogenetic Networks in Evolutionary Studies, *Molecular Biology Evolution*, 23(2): 254-267.
- Goud TS, Upadhyay RC, Onteru SK, Pichili VBR, Chadipiralla K. 2019. Identification and sequence characterization of melanocortin 1 receptor gene (MC1R) in *Bos indicus* versus (*Bos taurus X Bos indicus*). *Animal Biotechnology*, 1-12
- Li J, Liu S, Li Z, Zhang S, Hua G, Salzano A, ... & Yang L. 2018. DGAT1 polymorphism in Riverine buffalo, Swamp buffalo and crossbred buffalo. *Journal of Dairy Research*, 85(4), 412-415.
- Özdil F, İlhan F. 2012. Diversity of *Apis mellifera* Subspecies from Turkey Revealed by Sequence Analysis of Mitochondrial 16s rDNA Region, *Biochemical Genetics*, vol. 50, pp. 748-760
- Péntek-Zakar E, Oleksa A, Borowik T, Kusza S. 2015. Population structure of honey bees in the Carpathian Basin (Hungary) confirms introgression from surrounding subspecies. *Ecology and Evolution*, 5(23): 5456-5467.
- Rozas J, Librado P, Sanchez-Delbarrio JC, Messeguer X, Rozas R. 2010. *Universitat de Barcelona Current Released Version: 5.10.1*
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30: 2725-2729.
- Ünal G. 2016. "Trakya bölgesindeki bal arılarında (*Apis Mellifera* L.) MtDNA sitokrom c oksidaz altbirim I (COI) geni analizi". YL Tezi, Fen Bilimleri Enstitüsü, NKÜ, Tekirdağ
- Zhang Z, Schwartz S, Wagner L, Miller W. 2000. A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology*, 7: 203-214.
- Zhou S, Cai B, He C, Wang Y, Ding Q, Liu J, Li C. 2019. Programmable base editing of the sheep genome revealed no genome-wide off-target mutations. *Frontiers in Genetics*, 10.