

Matematiksel Evrim Çalıştayı Ders Notları: Fizik yasalarından DNA evrimine

Murat Tuğrul

IST-Austria, Am Campus 1, A3400, Klosterneuburg (Viyana ormanları), Avusturya and
Nesin Matematik Köyü, Şirince, Selçuk, İzmir, Türkiye*

(Dated: 11 Eylül 2013)

Abstract

ÖZET: Bu derste moleküler evrim hakkında nasıl daha iyi fikir edinebileceğimiz üzerine kuramsal bir çerçeve çizmeye uğraşacağız. Ele alacağımız konu gen ifadesinin düzenlenimi olacak. DNA'daki düzenlenim bölgelerinin evrimini anlamak için ayrıklı iki teorik disiplin olan popülasyon genetiği ve biyofizik arasında bir köprü kurmayı deneyeceğiz.

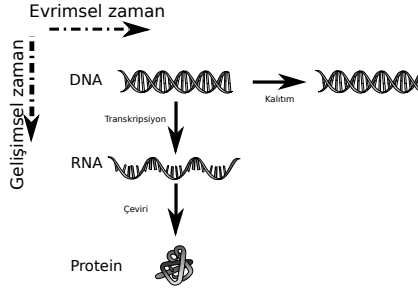


Figure 1: Moleküler boyutta temel yaşam faaliyetlerini DNA'nın i-) okunarak işlevsel moleküllerin oluşturulmasına (gelişimsel zaman ölçeği) ve ii-) çoğaltılarak gelecek kuşaklara aktarılmasına (evrimsel zaman ölçeği) indirgeyebiliriz. Çizim için Wikimedia Commons'tan yararlanılmıştır (Bknz bilgilendirme notları.)

Gen düzenleniminin evrimi

Kalıtım materyali olan DNA'daki gen adı verilen kodlar önce RNA ve sonrasında protein molekülüne çevrilir (Bknz Şekil-1). RNA ve proteinler (bundan böyle *gen ifadesi*) hücre içinde işlevsel açıdan en önemli molekül gruplarını oluştururlar. Evrimsel biyolojiyi ilgilendiren önemi ise, gen ifadesinin bireyler arası kalıtsal karakter farklılıkları yaratmalarıdır. Çok uzun bir dönem (1970'lere kadar) popülasyon içi ve arası görülen çeşitliliğin kaynağının genlerimizdeki kod farkı ve dolayısıyla gen ifadesinin yapısal (işlevsel) farklılığından ötürü olduğu düşünülüyordu. 1970'ler sonrası moleküler boyutta deneylerin mümkün hale gelmesi ve 2000'lerdeki teknolojik gelişmelerin doğurduğu genom projeleri, tür içi ve hatta yakın türler arası protein benzerliğinin şaşırtıcı derecede fazla olduğunu gösterdi. Çarpıcı bir örnek olarak bir insan ile bir şempanzenin genlerinin büyük oranda aynı (~%98) olduğunu gösterebiliriz.

Peki o zaman, doğadaki kalıtsal çeşitliliğe yol açan nedir? Bugün kabul gören hipoteze göre, kalıtsal çeşitlilik gen ifadesi **miktarlarındaki** farklılıklardan kaynaklanmaktadır. Gen ifadesini belirleyen ise belki

daha önce çöp DNA olarak nitelenen gen-dışı DNA bölgeleridir (bundan sonra *gen düzenleyici bölgeleri*).

Bildiğiniz gibi DNA 4 harfli (A, C, G, T) zincir bir moleküldür. Gen düzenleyici bölgelerindeki nükleotid değişiklikleri yol açtıkları farklılıklarla bu bölgeleri içeren ve gen ifadesi miktarını değiştirecek mekanizmaları etkiliyorlar. En başta gelen mekanizmalar içinde DNA-protein ilişkilerini göstermek mümkün. Mesela, RNA polimerazın, geni öncülleyen DNA bölgesinde (promotor) spesifik bağlanma bölgesini bulması önemlidir. Bağlanma enerjisi uygun ise, RNA polimeraz bu bölgeye oturur ve akabinde gen üzerinden geçerek koda göre mRNA moleküllerini oluşturur (*transkripsiyon*).

Belirli bir yaklaşım dahilinde, RNA polimerazın promotor bölgedeki kendisine has bölgeyi ne sıklıkla bulduğu ve bağlandığı ne kadar mRNA üretildiği ile ilişkilidir diyebiliriz. Bu spesifik bağlanmayı etkileyen bir çok mekanizma var. En öncelikli unsur RNA polimerazın DNA ile etkileşimi. Yine bunun dışında promotor üzerine bağlanan farklı proteinler de RNA polimerazın aktivitesini etkiliyor. Bazı proteinler daha fazla gen ifadesine yol açarken bazıları gen ifadesini azaltıcı etki edebilir.

Bu derste yapmaya çalıştığımız, kuramcı bakış açısı ile gen düzenleyici bölgelerinin nasıl evrildiğini anlamaya çalışmak olacak. Yaşam bilimlerinde genel kuramlar/modeller üretmek pek mümkün gözüküyor. Dolayısıyla, hangi olgunun ve organizmanın konu edildiği önem kazanıyor. Dersin devamında bir bakteri (ör: *E.coli*) organizmasını konu edineceğiz. Bilinmesi gereken önemli husus, bu organizmanın haploid olduğu yani aynı genden bir kopya taşıdığı ve eşeysiz üreme ile çoğaldığıdır.

Araştırma: Genel olarak ökaryot ve prokaryot transkripsiyonunu ve mekanizmalarını karşılaştırabilir misiniz?

Popülasyon Genetiği: Basit Bir Mutasyon-Seçilim modeli

Basit bir evrim modeli ile bir bakteri popülasyonunu ele alalım. Amaçlayacağımız, popülasyon içerisindeki genotiplerin frekansının evrimsel dinamiğini matematikçe betimlemek olacak.

Popülasyon büyüklüğü N 'in çok büyük olduğunu

*Electronic address: mtugrul@ist.ac.at; URL: <http://pub.ist.ac.at/~mtugrul/>

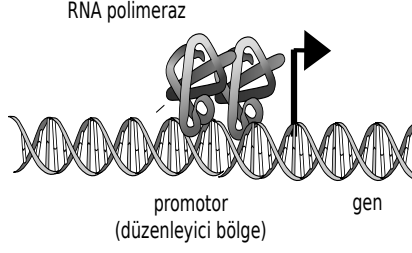


Figure 2: Genin ifade edilmesinde gen dışı birçok DNA bölgesi aktif rol oynar. Şekilde transkripsiyonu gerçekleştirecek RNA polimerazın promotor bölgesine bağlanımı gösterilmektedir. Promotor dizisindeki farklılıklar bağlanma enerjilerini değiştirerek gen ifadesi miktarlarına etki edebilir. Çizim için Wikimedia Commons'tan yararlanılmıştır (Bknz bilgilendirme notları.)

kabul edelim (*E.coli*'nin efektif popülasyon büyüklüğü 10^6 civarında), böylelikle *genetik sürüklenme* olarak adlandırılan örneklenim salınımlarını önemsememiş olacağız.

Yine aynı basitleştirici modelleme ile bu toplulukta promotor açısından iki genetik türün olduğunu farz edelim. Birincisine popülasyonda görülme olasılığı daha fazla olduğu için **asıl** tip diyelim. İkincisi tip nadir görülen olduğu ve asıldan mutasyonla oluştuğu için **mutant** diye adlandıracağız. Asıldan mutant tipe mutasyon oranının (hızının) sabit olduğunu ve kuşak ve birey başına μ olduğunu düşünelim. Bugün nükleotid ya da noktasal mutasyon hızının yaklaşık $10^{-7} - 10^{-9}$ mertebesinde olduğunu biliyoruz. Biz burada promotor başına bir orandan bağsettığımız için $\mu \sim 10^{-4} - 10^{-6}$ diye düşünebiliriz. Ters mutasyonun görülmesi çok nadir olduğu için önemsemeyeceğiz.

Mutant tipin bir nevi hastalıklı olduğunu ve asıl tipe göre daha az sayıda bölünme geçirdiğini düşünelim. Yani oransal olarak, mutant tip 1 çocuk veriyorsa, asıl tip $1 + s$ çocuk versin. s burada pozitif küçük bir sayı ve asıl tipin mutanta göre seçilim avantajının büyüklüğünü nicelliyor. Burada dikkat etmeniz gereken seçilimin bir çoğalım avantajından kaynaklanan oransal değişimden ibaret olduğudur.

Bu tariflere göre popülasyondaki asıl promotorların oranı seçilim sayesinde artarken, mutasyonlar sonucu azalacaktır. Gelin bu değişimleri bulalım. Şu andaki asıl tipin popülasyondaki oranı p olsun. Dolayısıyla mutant tip $1 - p$ olacaktır. Gelecek kuşakta seçilim aritmetiğinden dolayı $Np(1 + s)$ asıl tipte çocuk olacaktır, aynı şekilde $N(1 - p)$ mutant. Asıl tipin gelecek kuşaktaki seçimden dolayı oranına p' dersek bu oranı rahatlıkla ifade edebiliriz:

$$p' = \frac{Np(1 + s)}{Np(1 + s) + N(1 - p)} = \frac{p(1 + s)}{ps + 1} \quad (1)$$

Asıl tipin yeni frekansını değişim cinsinden ifade eder-

popülasyon

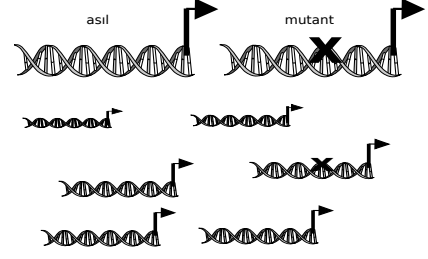


Figure 3: Popülasyon genetiği çerçevesinde bireyleri promotor dizileri olarak düşünüp, topluluk içindeki frekanslarının değişimini inceliyoruz. Çizim için Wikimedia Commons'tan yararlanılmıştır (Bknz bilgilendirme notları.)

sek, $p' = \Delta p^{(s)} + p$. Seçilim etkisini böyle özetleyebiliriz:

$$\Delta p^{(s)} = \frac{ps(1 - p)}{ps + 1} \approx ps(1 - p). \quad (2)$$

Son yaklaşımı $ps \ll 1$ olduğu için yapabildik. Mutasyon dolayısıyla frekans değişiminin de ($\Delta p^{(\mu)}$) şu andaki frekansa etki ettiğini düşünelim ($\Delta p^{(s)}$ küçük bir değer olduğu için bunu yapmamız sorun yaratmayacaktır). Asıl tipin gelecek kuşaktaki mutasyondan dolayı oranına p'' dersek,

$$p'' = \Delta p^{(\mu)} + p = (1 - \mu)p. \quad (3)$$

Dolayısıyla,

$$\Delta p^{(\mu)} = -\mu p. \quad (4)$$

Bu şekilde toplam frekans değişimini ifade edebiliriz.

$$\Delta p^{(T)} = \Delta p^{(s)} + \Delta p^{(\mu)} = ps(1 - p) - \mu p. \quad (5)$$

Bu evrimsel dinamiğimizi özetleyen matematiksel ifadeydi. Evrimsel dinamiğin bir süre sonra değişimin olmadığı faza girmesini bekleyebiliriz, yani $\Delta p^{(T)} = 0$. Durağan halde, mutant promotorun oranı

$$1 - \hat{p} = \mu/s \quad (6)$$

olarak tahmin edilir (“ ^ ” işaretini durağan hal için kullanacağız).

Alıştırma: Yukarıdaki dinamik ifadeyi $\Delta p^{(T)} = ps(1 - p) - \mu p$ ve yaklaşımlar yapılmamış halini farklı p değerlerinden başlayarak bilgisayarda kontrol edin. $1 - p(t)$ grafiğini çizdiğinizde her zaman μ/s değerinde mi son buluyor? Hangi μ ve s değerlerinde bu modelleme yaklaşımı anlamlıdır?

Biyofizik: Genotip-Fenotip-Uyumluluk Eşlenimi için Termodinamik bir Model

Yukarıdaki modelde iki genotip olduğunu varsaydık, ve dolayısıyla seçilim etkisini genel bir parametre (s) ile belirtebildik. Daha gerçekçi bir modelde birden çok genotipi işin içine katmamız gerekebilir. Böyle bir durumda sadece genel seçilim şiddetini değil, genotipler arası uyumluluk değişimini bilmemiz de elzem olabilir. Başka bir deyişle, uyumluluk sahası diye adlandırılan genotip-uyum eşlenimini (fonksiyonunu) modellememiz gerekecektir.

İzleyeceğimiz strateji

$$X \rightarrow g(X) \rightarrow r(X) \quad (7)$$

olacaktır. Burada X genotipi (promotor), g fenotipi (gen ifadesi miktarı) ve r ise uyum değerini belirtir.

İkinci kısımda ($g \rightarrow r$) modellenecek gene bağlı olarak fenotip üzerine etkiyen seçilim için farklı hipotezler düşünebiliriz. En basit örnekler olarak

yönlü seçilim: $r(g) \propto g$

ve dengeleyici seçilim: $r(g) \propto \text{Exp}[-(g - g_{opt})^2]$

modellerini gösterebiliriz.

Birinci kısım ($X \rightarrow g$) için promotorumuzun sadece RNA polimeraz spesifik bağlanım bölgesi olduğunu düşünelim. Basit bir yaklaşımla polimerazın promotora bağlanma olasılığı $p_{on}(X)$ 'un durağan gen ifadesi miktarı ile orantılı olduğunu söyleyebiliriz. Aşağıdaki denklemde bu söylediklerimiz doğrultusunda bir formülasyon veriyoruz.

$$\frac{d}{dt}g = -\gamma_{boz}g + \gamma_{üre}p_{on}(X) \quad (8)$$

Gen ifadesi miktarının gelişimsel zaman ölçeğindeki dinamiğini betimleyen bu türevsel denklemde; birinci ifade hücredeki RNA ya da protein bozunmasını, ikinci ifade ise gen ifadesi üretimini içeriyor. γ 'ların sabit olduğu bu betimlemede durağan halde ($\frac{d}{dt}g = 0$) gen ifadesi miktarı

$$\hat{g} = \frac{\gamma_{üre}}{\gamma_{boz}}p_{on}(X) \quad (9)$$

olacaktır.

Peki verilen bir X promotoru için $p_{on}(X)$ 'u nasıl tahmin edebiliriz? Bu soruya cevap vermek için termodinamik bir akıl yürütme yapacağız. RNA polimerazın promotora bağlanım ve çözünüm kinetiği eğer dengede ise *serbest promotor* ve *RNA polimeraz bağlanmış promotor* hallerinin görünme olasılığı Boltzmann ağırlıkları ile orantılı olacaktır. Yani sırasıyla 1 ve $e^{-E(X)/k_B T}$. Promotor çevresinde efektif olarak c tane polimeraz varsa,

$$p_{on}(X) = \frac{c e^{-E(X)/k_B T}}{1 + c e^{-E(X)/k_B T}} \quad (10)$$

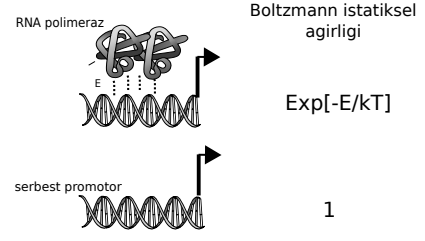


Figure 4: Termodinamik yaklaşım altında RNA polimerazın bağlanım olasılığını promotor durumlarını sıralayarak ve Boltzmann istatistiksel ağırlıklarını kullanarak hesaplamamız mümkün. Çizim için Wikimedia Commons'tan yararlanılmıştır (Bknz bilgilendirme notları.)

olarak ifade edilir. En son olarak, gelişen deneysel teknikler sayesinde verilen bir promotor için bağlanım enerjisi $E(X)$ 'i daha güvenilir şekilde modelleyebildiğimizi ekleyelim.

Alıştırma: Birden fazla bağlanım bölgelerini dikkate alarak, RNA polimeraz dışında transkripsiyon faktörlerinin de etkilerini modelleyebiliriz. Bunun için ilk adım olarak, iki bağlanım bölgesi için bütün promotor durumlarını belirteip, karşılık gelen Boltzmann ağırlıklarını yazabilir misiniz?

Bilgilendirmeler ve Kaynaklar

Öncelikle bu dersin içeriğini ve akışını şekillendirmemde bana yardımcı olan Hande Acar, Şebnem Keşaf ve Srdjan Sarikas'a çok teşekkür ederim. Şebnem ayrıca bu ders notlarının kritik okumasını yaparak çok önemli düzeltmeler ve önermeler yaptı.

Çizim için Wikimedia Commons'tan yararlandım. Dosya adı: Central_dogma_of_molecular_biology.svg ve orijinal atıf: Barillot vd., Computational Systems Biology of Cancer Chapman & Hall/CRC Mathematical & Computational Biology , 2012.

Böylesi bir disiplinlerarası etkileşimi henüz flört aşamasındaki bir ilişkiye benzetmek günümüz bilimi için daha doğru bir tasvir olur. Zira, iki disiplin etkileşimi çok yeni ve henüz bir evliliğin getirdiği somut bir meyvesi var denilemez. Bu sebepten, dersin tüm içeriğini tek bir kitap dahilinde bulmak henüz mümkün olamayabilir. Bu notları okuyan sizin böylesi bir dersin belki de bir ilk deneme olduğunu ve sizlerden alacağım geribildirimlerden memnuniyet duyacağımı bilmenizi isterim.

Konuyla ilgili olabilecek kaynakları buraya not ediyorum. Daha fazlasına ihtiyaç duyarsanız benimle iletişime geçin.

- Transkripsiyon tasviri içi: Essential Cell Biology,

3rd ed., Alberts et al. (2010), pp. 231-246

- Transkripsiyon için termodinamik model: William Bialek, pp.270-279, Biophysics, 2012, Princeton Universtiy Press.
- Transkripsiyon için termodinamik model: Philips et al., pp. 244-248, Physical Biology of the Cell, 2013, Gerland Science Press
- Bu dersin ruhuna yakın bir makale: Gerland & Hwa, J. Mol. Evol. (2002), "On the selection and evolution of regulatory DNA motifs",
- Bağlanım dizisinden bağlanım enerjisine nasıl gideriz için: Kinney et al., PNAS (2010), "Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence."