



Evrimsel Perspektifte KARŞILAŞTIRMALI GENOMİK

Fatih Dikmen

"Nothing in biology makes sense except in the light of evolution"

Theodosius Dobzhansky

Santral Dogma kavramı 1958'de Francis Crick tarafından ortaya atılmış ve ardından 1970 yılında *Nature*'da yayınlanmıştır. DNA'dan RNA'ya ve RNA'dan proteine genetik bilgi akışını özetleyen bu kavrama (1) göre DNA'nın yani genomun şifresinin çözülmesi ile bu genden oluşabilecek proteinler de bilinebilecekti. İnsan Genom Projesi'nin motivasyon kaynağı belki de bu kavrama dayanmıştır diyebiliriz. Bu klasik genetik yaklaşım ile hücredeki "santral dogma"ya, birbirinin kodları gibi görünen bu sisteme baktığınızda, şifrelerden birinin bilinmesi halinde diğerlerinin de "şifre çözümü" ile elde edilebileceğini düşünmeniz doğaldır. İnsan genomu çözülmeden önce insanların 100.000 adet gene sahip olduğu tahmin edilmekteydi. Genom sekansının tamamlandığının söylendiği 2001 yılında bu tahmin 25.000'e düştü. Şimdi ise toplam sayının yaklaşık 22.000 olduğunu ama her an 19.000'e bile düşebileceğini bilmekteyiz. Santral dogma mekanizmasının o kadar katı biçimde yürümediği gerçeği büyük umutlar bağlanan İnsan Genom Projesi'nin son aşamalarına gelindiğinde ve gen sayısının düşünülüğünden az miktarda olduğunun keşfi ile gün yüzüne çıkmıştır. Hatta bunun ötesinde Stein (2004)'e göre aslında 2001 yılında İnsan Genom Projesi Konsorsiyum'unun "insan genomu çözüldü" şeklindeki açıklaması da aslında tam gerçeği yansıtmamaktaydı. Çünkü hem sekans datalarındaki hatalar tam olarak ayıklanmamıştı hem de heterokromatin bölgeler dahil edilmemişti. Stein'e göre 2010 yılında birilerinin çıkıp tekrar "insan genomunu çözdük" diye

açıklama yapması mümkün, çünkü belki de anca o tarihe kadar eldeki veriler tam olarak derlenmiş ve doğrulanmış olabilecektir (2). Ancak artık biliyoruz ki asıl mesele genomu çözmek değil, onun fonksiyonları belirleyebilmektir (3). Bu esasında “santal dogma” denilen gen’den proteine giden yolun tamamıyla keşfi anlamına gelmektedir. Karşılaştırmalı genomik çalışmaları bu çalışmaların temel bir parçasını oluşturur. Bu temelde insanoğlunun kendini, kökenini arama, hayatı anlama çabasının ve hastalıklarla mücadelede doğaya karşı öne geçme mücadelesinin bir ürünüdür. Tüm bu mücadelenin merkezinin insana dayanması bir bakımdan sevindirici bir bakımdan da düşündürücüdür.

İşte doğanın gizli şifreleri olan genomu anlama çalışmaları bu açıdan farklı alt disiplinlerde çalışır. “Yapısal genomik” temelde genomum (DNA, RNA... vs) dizisinin kabaca çıkartılması ile ilgilenirken, “fonksiyonel genomik” çalışmaları mikroyarray çalışmaları ve metabolik yolların analizleri ile santral dogmanın ürünlerinin hücrede ne işe yaradığını anlamaya çalışır. Bu disiplinler arasında tam bir ara bulucu olan karşılaştırmalı genomik (comparative genomics) ise farklı organizmaların genomları arasındaki ilişkiyi (farklılık ve benzerlikleri) inceleyen yeni bir bilim dalıdır (Tablo.1). Organizmaların fenotiplerindeki farklılığın nedeni genomdan kaynaklandığına göre, bizler türleri birbirinden farklı kılan bu fenotipik özelliklere ait RNA ve proteinleri kodlayan genlerin sekansını bilirse bu türlerin birbirinden nasıl farklılaştığını anlayabiliriz (3). Bu karşılaştırmalı genomik’in temel çalışma alanını oluşturur. Bu karşılaştırmalardan çıkan veriler, yeni genlerin belirlenmesinde, protein çalışmalarında ve evrimsel çalışmalarda kullanılabilir.

Genomların karşılaştırılması işleminde organizmaların filogenetik mesafeleri ile ilgili pek çok soru gündeme gelebilir (Şekil.1). Eğer geniş ölçekli değerlendirmeler yapılacaksa uzak filogenetik ilişkide bulunan canlıların genomlarının kıyaslanması daha doğru olur. Mesela Rubin et. al (2000)’a göre maya, solucanlar ve sineklerin genomlarının kıyaslandığı çalışmada ökaryotik proteinlerin çoğu ortak olarak bulunurken filogenetik olarak daha yakın akraba olan sinek ve solucanların bazı temel proteinleri (sinyal yolu proteinleri), mayalara göre iki kat daha fazla taşıdıkları bulunmuştur. Maya ile

sinek+solucan grubu arasındaki evrimsel zaman farkı gibi geniş zaman dilimlerinde çalıştığınızda gen tertibinin ve işlevi düzenleyen sekans dizilerinin genelde korunmamış olduğunu (yani kaybedildiğini) görürsünüz (3). Ancak 70-100 milyon yıllık bir zaman dilimleri içinde birbirinden ayrılan organizmalarla çalıştığınızda hem fonksiyonel hem de fonksiyonsuz DNA'ların korunabildiğini görebilirsiniz. Bu tür vakalarda fonksiyonel DNA sekanslarının durulaşma ve negatif seleksiyon izleri gösterir ki bu da, fonksiyonel DNA'nın fonksiyonsuz veya nötral DNA'dan daha az değişeceğinden kaynaklanır (3).

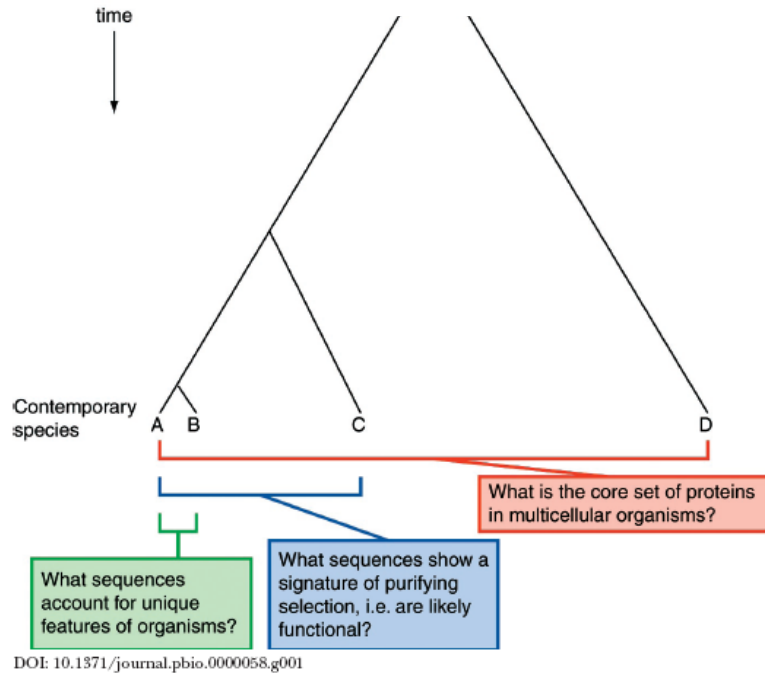
Tablo.1- Sekansı yapılmış bazı organizmalara ait genom verileri

Organizma	Tahmini genom büyüklüğü	Tahmini gen sayısı	Kromozom Sayısı
<i>Homo sapiens</i> (insan)	2,900 milyon baz	~30,000-20.000	46
<i>Rattus norvegicus</i> (sıçan)	2,750 milyon baz	~30,000	42
<i>Mus musculus</i> (fare)	2500 milyon baz	~30,000	40
<i>Drosophila melanogaster</i> (sirke sineği)	180 milyon baz	13.600	8
<i>Arabidopsis thaliana</i> (bitki)	125 milyon baz	25.500	5
<i>Caenorhabditis elegans</i> (nematod)	97 milyon baz	19.100	6
<i>Saccharomyces cerevisiae</i> (maya)	12 milyon baz	6300	16
<i>Escherichia coli</i> (bakteri)	4.7 milyon baz	3200	1
<i>H. influenzae</i> (bakteri)	1.8 milyon baz	1700	1

Karşılaştırmalı genomik bunun gibi sadece filogenetik olarak uzak veya yakın akrabalığı ya da fonksiyonlu-fonksiyonsuz DNA'yı ayırt etmeye yaramaz. Bunların yanında kodlanan exonlar, kodlanmayan RNA'lar, gen düzenleyici bölgeler gibi genel DNA dizisi sınıflarının fonksiyonlarını da anlamamıza yardımcı olur. Bu konuda Fare Genom Konsorsiyumu'nun 2002 yılında fare ve insan genomu arasında yaptıkları karşılaştırmalar, enterik bakterileri türlerinin ve maya türlerinin kendi aralarındaki ve çapraz kıyaslamaları örnek

gösterilebilir. Yine yakınlarda Stein et. al (2003) tarafından *Caenorhabditis briggsae* ve *C. elegans* genomları arasında yaptıkları analizler de bu kapsamda ele alınabilir (3). Bunlara zıt olarak çok yakın tarihte (5 myö) birbirinden ayrılmış olan insan ve şempanze'nin oldukça benzer genomlarının kıyaslanması ile bu iki yakın akraba arasındaki farklılığı yaratan gen dizileri araştırılmaktadır. Bunlar pozitif seleksiyon sonucu oluşan değişikliklerdir (3). Bu açıdan bakıldığında, sekans datalarının birikimi ile karşılaştırmalı genomik çok daha bilgi verici güçlü bir disiplin olarak kalacaktır (3).

Şekil.1 (Hardison,2003)

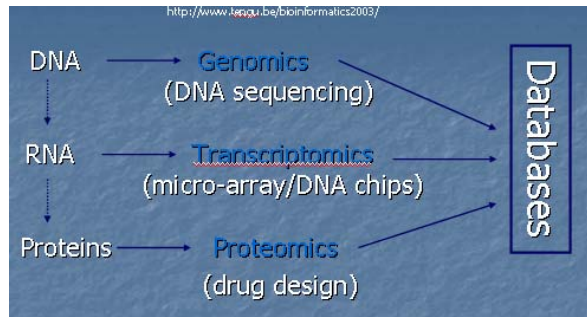


Bilgisayar Programları ve Veri Tabanları:

Farklı DNA dizilerinin hizalanması karşılaştırmalı genomik çalışmalarındaki ana işlemdir. Hizalama işlemi, bir dizideki nükleotid ile diğer dizideki nükleotidin karşılaştırılarak iki dizi arasındaki uyumlu alanların belirlenmesidir. Bu yüzden en iyi şekilde hizalama yapılmalıdır ki ona göre en iyi uyumluluk sonuçları alınabilsin. Bu konuda çok güçlü algoritmalar geliştirilmiştir. Karşılaştırmalı genomik çalışmalarının en zor kısmı işte bu gibi milyarlarca bazlık genom dizileri arasında hizalama yapmak ve ondan çıkan verileri yorumlamaktır (3). Bunun yanında

çeşitli yazılımlar, genomdaki genleri aramak için genellikle benzerlik araştırması ilkelerinin ile ilgili algoritmaları birleştirir. Tüm bu analizler çeşitli bilgisayar programları ve biyoinformatik sayesinde çözülebilmektedir. Günümüzde pek çok araştırma grubunun üzerinde çalıştığı genom dizileri ile ilgili veri tabanlarının bir kısmının listesi Tablo.2'de verilmiştir (3). Veri tabanlarının genomikteki önemi küçümsenmeyecek derecede çoktur. Bu veritabanlarının birçoğuna internet üzerinden serbestçe ulaşılabilmektedir. Bu sistemin işleyişine yönelik şematik anlatım Şekil.2'de gösterilmektedir.

Şekil.2



Primer veri tabanları orijinal nükleik asit ve protein dizisini içerirken, sekonder veritabanları ise yüksek oranda korunmuş protein ailelerinin profillerinin sıralanmasında primer veritabanlarındaki bilgiyi kullanır. BLAST, FASTA ve onların türevleri gibi arama algoritmaları yeni sekansların daha önce depolanmış tüm sekanslarla karşılaştırılmasına izin verir. Buradaki diğer bir önemli nokta sekans veri tabanlarının, sadece sekansı dizisi hakkında bilgi vermekle kalmaması aynı zamanda fonksiyonel gen taramaları ile ilişkili olabilecek bilgileri de verebilmesidir. Mesela bir genomdaki bilinmeyen genler ile ilgili olarak diğer organizmalar ile karşılaştırma yapıp muhtemel fonksiyonel gen bölgeleri hakkında bilgi verebilirler.

Tablo.2 (Hardison, 2003'den alınmıştır)

Table 1. URLs for Accessing Precomputed Whole-Genome Alignments and Their Analysis		
Server or Browser	Genomes Covered	URL
EnterIX	enteric bacteria	http://bio.cse.psu.edu/
VISTA Genome Browser	human, mouse, rat	http://pipeline.lbl.gov/
UCSC Genome Browser	mammals, worms, zoo*	http://genome.ucsc.edu/
Ensembl	mammals, fish, insects, worms	http://www.ensembl.org/
GALA	human, mouse, rat	http://bio.cse.psu.edu/

*Data from multiple alignments of 13 vertebrate genome sequences homologous to the human CFTR region (Thomas et al. 2003) are included.
DOI: 10.1371/journal.pbio.0000058.t001

Genom Evrimi

Karşılaştırmalı genomik kabaca bizlere,

- İnsan ile şempanze arasında gen farkının %1'den az olduğu,
- İnsan ile goriller veya orangutanlar arasındaki genetik benzerliğin %97'den fazla olduğu,
- İnsanın yüzlerce atasal (3,5 milyar yıllık) gen taşıdığı (mesela bakterilerle ortak 500 gen olduğu),
- *Drosophila* ile insan arasında ortak 1500 gen bulunduğu,
- İnsanlardaki genlerin %80'inin fare ile ortak olduğu,

gibi bilgilerin özetlenebilmesini sağlamıştır. Temelde bu yaklaşım ile yapılabilecek en basit değerlendirmelerden biri, bu şekilde genomlar arası uyumun analizidir. Mesela evrimsel süreçte insan kabaca 75-80 myö fareden ayrılmıştır fakat %90 oranındaki genetik yapı homoloji göstermektedir (3). Bu analizler ayrıca şunu da göstermektedir ki, sadece fonksiyonel DNA değil fonksiyonel olmayan DNA da karşılaştırılabilmektedir. Özellikle fare-insan kıyaslamasında memelilere özgü ortak atadan miras fonksiyonel olamayan mesela relik transpozonlar da analizlere katılabilmektedir. Teorik olarak fonksiyon olmayan atasal dizi tekrarı ile tüm genomun hizalanarak analizi bize fonksiyonel ya da fonksiyonel olamayan DNA dizileri hakkında oransal bir yorum hakkı tanır. Mesela böyle bir çalışmaya göre insan genomunun sadece %5'inin seçilim sonucunda fonksiyonel olduğunu göstermiştir. Bu oran protein kodlayan normal bölgeye göre 3 kat daha büyüktür (3).

İnsanda protein kodlayan genlerin hemen hemen tümünü hizalandığında fare ile %99 homoloji ve %80 1:1 ortholog'luk göstermektedir. Çoğu durumda intron-exon yapıları oldukça korunmuştur. Fare ile insan biyokimyası arasındaki benzerlik ve uyum düşünüldüğünde protein kodlayan bölgenin bu olağanüstü korunumu belki beklenebilir bir durumdur. Ancak bu durum aynı filogenetik ayrımı gösteren her kıyaslama vakasında da görülmemektedir (3).

Ortologlar; aynı fonksiyonu kodlayan ve vertikal alçalma ile yavaş evrimleşen farklı organizmalardaki homolog genlerdir.

- Ortologlar mutasyonların yavaş olarak birikmesi ile ortaya çıkar.

Paraloglar; benzer fakat özdeş olmayan fonksiyonları kodlayan bir organizmadaki homolog genlerdir.

- Paraloglar mutasyon birikiminin ardından gen duplikasyonu ile meydana gelir.

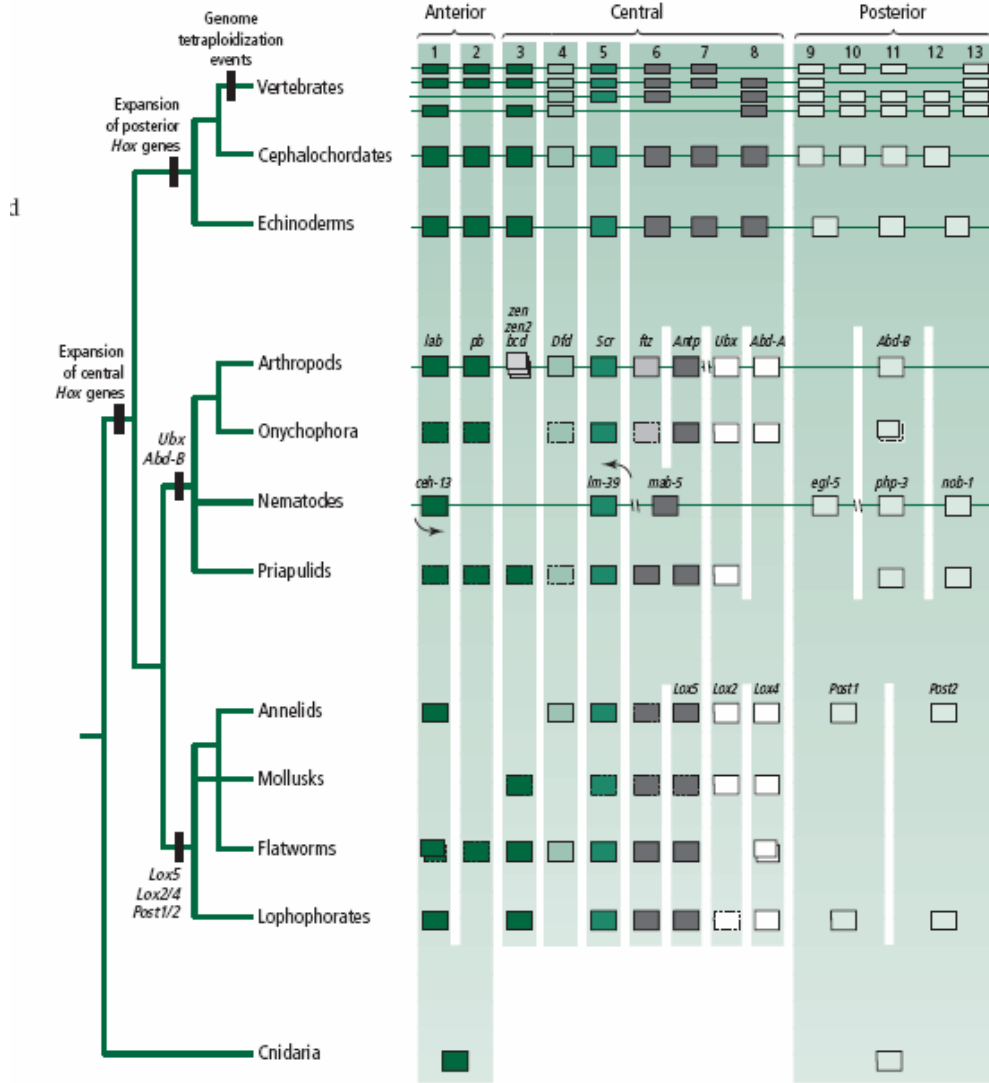
Rhesus makakları *Macaca mulatta*, insan ve şempanzelerden sonra, tüm genom dizisi çıkarılan üçüncü primat türüdür. Yapılan incelemeler, makak-insan genomlarının protein kodlayıcı dizi bakımından %97,5 aynı olduklarını ortaya çıkardı. Makaklarda, insanlarda ve şempanzelerde aynı olan 108 gen ailesinin farklı şekilde evrimleştikleri anlaşılmış oldu. Genler arasındaki bu farklılığın, türler arasındaki farklılığa açıklama getirebileceği, böylece türlerin doğal seçilim ve evrim süreçlerinin daha iyi kavranabileceği düşünülebilir. Bunun yanında gen yapılarındaki farklılıkların açığa çıkarılmasının tıbbi araştırmalar için de önemli olduğu kesindir. Mesela son çalışmalarla, insanlarda vücudun hastalıklara karşı savunmasını oluşturan HLA sisteminin makaklarda çok daha fazla sayıda gen tarafından oluşturulduğu ortaya çıkarıldı. Bu, makakların bağışıklık sisteminin bizdekinden farklı çalıştığına dair bir işaret olabilir. Makaklar bağışıklık sistemi araştırmalarında sıkça kullanılan model organizmalardan biri olduklarından, bu bilgi araştırmacılar için son derece önemlidir (4).

Caenorhabditis briggsae ve *C. elegans* genomlarının protein kodlayan bölgeleri arasında yapılan analiz %60 homolojiyi göstermiştir. Bu iki nematod arasında morfolojik fark bulmak çok zordur. Muhtemeldir ki gelişim basamakları da aynıdır. Ancak bu iki benzer nematodun bu kadar farklılık göstermesi bize bunların benzer fenotipleri farklı genetik yollarla sağladıklarını düşündürmektedir. Bu konuda yapılacak detaylı benzerlik/farklılık analizler belki de gelişim biyolojisi konusunda çok yeni ufuklar açabilecektir(3).

Karşılaştırmalı genomik çalışmalarına bir diğer örnek de evrimsel gelişim biyolojisi (evo-devo) çalışmalarından verilebilir. Bu alandaki çalışmalar genelde farklı organizmalardaki canlı gelişiminin genetik temelleri üzerine yapılır. Bu alandaki en güzel örneklerden biri *Hox* gen kompleksinin çok hücreli canlılardaki gelişimidir. Bu konuda yapılan çalışmalar ve karşılaştırmalara göre, *hox* gen kompleksi 2 temel değişim göstermiştir (Şekil.3). Birincisi radyal simetrlili Cnidaria'dan bilateral simetrlili canlılara geçişte, ikincisi ise omurgalılara geçiştir. Cnidaria'da 2 *hox* gen seti bulunmaktayken omurgasızlarda bu

sayı 13'dür. Omurgalılarda ise bu 13'lük setten 4 adet bulunur. Omurgalılarda duplikasyon ile oluştuğu düşünülen bu setler daha komplike vücut yapısının gelişimine katkı sağlamıştır (5).

Şekil.3- Hox genlerinin geçmişi; 12 hayvan grubundaki Hox genlerinin karşılaştırılması (5)



Nükleotid seviyesindeki kıyaslamaya göre fare-insan genomları %40'lık bir uyum göstermektedir. %60 fark ise insersiyon, delesyon ve diğer mekanizmalar ile değişmiştir. Bu farklı kısım içinde sadece insana özgü bir dizi sınıfı (genomun %24'ünü oluşturur) vardır ki onlar transpozisyon ile ortaya çıkmış olan tekrar bölgeleridir. Bu bölge farelerde kesinlikle bulunmaz ve bu yüzden hizalama analizlerinde hiç uyum göstermez. Buna benzer biçimde “kemirgenler-fareler”e özgü retrotranspozonlar vardır ki fare genomunda %33 oranında yer kaplar ve bunlar da insanda bulunmaz. İnsandaki %36'lık öteki farklı bölge

konusunda pek bir yorum yapılamamıştır. Belki bu durum hizalama yapan programların algoritmalarındaki yetersizlikten kaynaklanıyor olabilir. Bunun yanında buralar, eskiden homolog olmalarına karşın sonradan evrimsel süreçte genomda kaybolmuş bölgeler de olabilir. Tüm bu yorumların daha doğru yapılabilmesi başka türlerinde bu analizlere eklenmesi ve tekrarsız dizileri ve hizalanmayan genomları içeren kapsamlı açıklamaların yapılabilmesi ile mümkündür (3). Ayrıca sadece evrimleşme hızının değişebilirliği değil, genomların da epeyce varyasyona açık olduğunun göz önünde tutulması gerekmektedir. Özellikle memeli kromozomlarının nükleotid değişimine, transpozon eklenmelerine ve rekombinasyonlara açık olması, primat genomunda segmental duplikasyonların değişikliklerin temel kaynağı olduğunun bilinmesi ve ona göre analizlerin gözden geçirilmesi gerekmektedir (3).

Genlerin İzinde

Genom dizilerinin benzerliği üzerine yapılan araştırmalar fonksiyonel gen bölgeleri belirlemede en önemli yöntemdir. Bu konudaki en iyi örnek protein kodlayıcı bölge tespitinde bu analizlerin katkılarıdır. Bunun için normalde türler arası benzerliği analiz eden yazılımlar büyük genomların analizine uyarlanmıştır. Bu sayede memelilerde 1000 yeni gen seti tespit edilmiştir. Nematod genomlarının kıyaslanması ile de 1275 adet iyi desteklenmiş gen bölgesi önerilmiştir. Artık günümüzde kodlanmayan RNA bölgelerinin tespiti biyoinformatik ve genomik çalışmaların en önemli hedefi olmuştur. Ayrıca şunu da belirtmeliyiz ki, kodlanan bölgeler için kullanılan türler arasındaki benzerliği analiz eden programlar bu yeni hedef için de umut vaat etmektedir (3).

Asgari Genom (Minimal Genom)

1995 yılında Craig Venter'in başkanlığını yaptığı TIGR (The Institute For Genomic Research) insan üriner kanalında yaşayan parazit bir bakteri olan *Mycoplasma genitalium*'un genomunun dizilimini çıkarmaya başladı. Yalnızca 580.000 baz çiftinden (580 kb) oluşan *M. genitalium*'un genomunda 517 gen bulundu. Daha sonra bunlar içinden protein kodlayan genlerin 480 olduğu tespit edildi. Bu bilinen bir canlıdaki en küçük gen sayısıydı (6).

Bu sonuçlar araştırmacıları, hücrenin yaşaması için gerekli olan asgari (minimum) gen sayısının ne kadar olabileceğini bulmaya teşvik etti. Acaba tüm canlılarda ortak asgari bir gen takımı mı vardır?" Canlılık ya da yaşamın sırrı bu asgari gen takımında olabilir mi?

Aslında yaşam için gerekli olan asgari genom büyüklüğünü tahmine yönelik ilk çalışmalar 1990'lı yılların ortalarında başlar. Bakterilerin, bazı genlerinin içine kısa DNA dizileri sokularak işlevsiz hale getirilmesine tolerans gösterebilirler de, bazı gen bölgelerinin işlevini yitirmesine karşı son derece dayanıksız oldukları, mutasyonların doğasını anlamak üzere yapılan çalışmalardan bu yana bilinmektedir (6).

Canlılar, yaşamsal açıdan zorunlu "temel işlevlerine" ek olarak tamamen uyumsuz nedenlerle daha karmaşık ve lüks diyebileceğimiz metabolik yollara sahiptir (6). Bu da, yaşamsal genlerinin yanında ekstra genlere de sahip olmayı sağlamıştır. Evrimsel süreç içinde gerçekleşen, gen duplikasyonları, yatay gen taşınması (lateral gen transferi) gibi olaylar canlılarda genom büyüklüğünü artırmış, çevresel değişimlere karşı gen ifadesini düzenleyen kontrol mekanizmaları geliştirmiştir (6). Tüm bunlar canlıların yaşamını kolaylaştırıp, uyum yeteneğini artırır; onlara çok yönlü bir yaşam tarzı sağlar. Ancak, varlıkları zorunlu olmayan ve genomu şişiren bu fazladan genlerin elenmesiyle, bu asgari genoma yeniden ulaşılabileceği düşünülmektedir(6). Bu aynı zamanda yaşamın tekrar tanımlanması anlamına gelmektedir. Asgari genomu tanımlayıp sentetik olarak onu üretebilecek noktaya gelindiğinde belki yepyeni bir çağır açılmış olacaktır. Bu konuda Venter Enstitüsü çalışmalarına devam ettiği gibi, *Newsweek* 04 Haziran 2007 sayısında Venter "Tanrı rolünü oynamak" rolüyle bu konudaki çalışmalarından dolayı kapak olmuştur. Dergi

aynı zamanda Princeton Üniversitesi profesörlerinden Lee Silver imzalı bir derleme yazı ile sentetik yaşamı keşfetmeye çalışan “Life 2.0” projesini anlatmaktadır. Buna göre sentetik olarak amaca göre tasarlanabilecek küçük tek hücreli organizmalar ile mesela insanlığın besin ve enerji ihtiyacı karşılanmaya çalışılabilecektir. Bu konuya iş dünyasının da ilgisinin yoğun olduğunu söyleyen yazara göre Berkeley Üniversitesi araştırmacılarına, bu tarz araştırmaları yapmaları için Bill Gates tarafından 42 milyon dolar teşvik primi verilmiştir. Yine Venter da planladıkları “yakıt üreten hücreler” projesine pek çok gelişmemiş ülkeden çok ciddi destekler aldıklarını belirtmiştir (7).

Kaynaklar

1. Crick, F., 1970, Central Dogma of Molecular Biology. Nature 227, 561-563.
2. Stein, L.D., 2004, *Human genome: End of the beginning*, Nature, 431, 915-916
3. Hardison, R.C., 2003 Primer on Comparative Genomics. PLOS Biology, 1:156-160.
4. Nivart Taşçı, 2007, Makak maymunu genomu çalışmaları, Bilim ve Gelecek, Sayı 39 <http://www.bilimvegelecek.com.tr/?act=2&id=117&sayi=39>
5. Ridley, M., 2004, Evolution, pp.751, Blackwell Publishing, part 20, Evolutionary Developmental Biology, pp. 572-590.
6. Gülsaçan, M., 2004, Yapay Yaşam, Bilim ve Teknik, Nisan
7. <http://www.msnbc.msn.com/id/18882828/site/newsweek/page/0/>