



# DNA'da Veri Depolama

**1** 956 yılında 1.5 ton ağırlığında ve 5 MB kapasiteye sahip dünyanın ilk hard diskinin üretilmesinden bu yana 60 yıldan uzun bir zaman geçti. Bu süre içerisinde önce taşınabilir hard diskler, sonrasında giderek ufalan flaş bellekler sayesinde gigabaytlarca, hatta terabaytlarca veri aktarabilecek teknolojiye sahip hale geldik.

Ancak dijital dünyada veri depolama ihtiyacı öyle hızlı arttı ki, bilim insanları son yıllarda geleceğin veri depolama çözümü olarak gösterdikleri yeni bir alanda çalışmalarını hızlandırdı. Dijital bilgiyi DNA'ya kaydetmek, son yıllarda genetik bilimcilerin en ilginç araştırma konularından biri oldu.

DNA tüm evrende fizik kurallarına uygun, bilinen en yoğun veri depolama mekanizması olarak kabul ediliyor. Bir milimetreküp DNA'nın, 1 kentilyon (1.000.000.000.000.000.000) byte veri tutabileceği hesaplanmış. *MIT Technology Review* dergisi bunun teknolojik gelişme açısından ne kadar önemli olduğunu ifade etmek için, "Bu teknoloji şu anda mümkün olsaydı, çekilmiş tüm sinema filmleri bir küp şeker büyüklüğünde bir yapıya depolanabilirdi" örneğini veriyor.

## **DNA'yı Saklama Cihazı Olarak Kullanmanın Yeni Yolu**

İlk olarak bakteriye veri depolama fikriyle başlayan süreçte DNA hakkında yapılan araştırmalar ve elde edilen sonuçlar bilim dünyasını da bambaşka bir boyuta taşıdı.

Genetikçi Seth Shipman ve Jeff Nivala tarafından yönetilen bir Harvard araştırma ekibinin, canlı bakteri hücrelerini mikroskopik sabit disklere dönüştürmeyi başarması dikkat çekici bir gelişme oldu. Daha önce bilim insanları bu yolla sadece 11 bit veriye erişebiliyordu ancak ekip bakterinin bağışıklık tepkisini, canlı hücrelere yaklaşık 100 baytlık veri yüklemek için kullanmayı sağlayan yeni bir yöntem keşfetti. Ayrıca bakterilerin bu bilgileri sonraki nesillere de aktardığı görüldü.

2017'de Harvard Üniversitesi Biyoloji Tabanlı Mühendislik Enstitüsünden araştırmacı George Church yönetimindeki ekibin CRISPR gen düzenleme tekniğini kullanarak ilk kez canlı hücrelere bir video kaydedip daha sonra oynatmayı başararak, veriler canlı hücrelere daha sonra erişilebilecek şekilde kaydedilebildi.

Washington Üniversitesinden bir ekip ise, DNA molekülü üzerine dijital fotoğraf, video, ses ve yazılar kaydedip bu verilere tekrar ulaşmayı başardı. DNA üzerine depolama yönteminin gelişimi için atılan en büyük adım olan bu araştırma için Washington Üniversitesi Bilgisayar Bilimi ve Mühendisliği Doçenti Luis Ceze, "Aslında yaptığımız şey resim, video, yazı gibi dijital verileri yüzlerce yıl boyunca kullanılacak şekilde saklamak için DNA'yı yeniden düzenlemek" dedi.

Washington Üniversitesinin Microsoft ile birlikte yürüttüğü çalışma neticesinde DNA ipliklerine 200 MB veri yüklendi. DNA'ya kodlanan bu verinin içinde İnsan Hakları Evrensel Beyanname'sinin 100 farklı çevirisi,

Project Gutenberg’de bulunan en popüler 100 kitap, Crop Trust’ın tohum veri tabanı ve OK Go grubunun This Too Shall Pass adlı şarkısının yüksek çözünürlüklü video klipi kaydedildi. DNA üzerine kaydedilen verilere, hiçbir veri kaybı olmadan, rastgele erişim sağlanabildi.

### **DNA’ya Veri Depolamada Rastgele Erişim Gelişiyor**

Yapılan tüm bu araştırmalar sonucunda DNA’da veri saklama alanı giderek daha büyüdü ve iyileşti. Bilim insanları da DNA üzerinde depolanan dosyaları bulmak ve seçerek almak için daha etkili bir yol arayışına girdi.

Son olarak, 2018’in Şubat ayında Washington Üniversitesi ve Microsoft araştırmacıları, dijital veri için DNA tabanlı depolama sistemi geliştirmeye yönelik önemli bir adım attıklarını açıklayarak DNA’ya kodlanmış 200 megabayttan fazla dijital bilgiden, bağımsız veri dosyalarını kurtarabilecekleri ilk rastgele erişimli depolama sistemini raporladı. “Rastgele Erişim”, belirli bir ögeye fiziksel olarak veri bloğunun neresinde olduğuna bakılmaksızın, eşit zamanda erişebilme anlamına geliyor. *IEEE Spectrum*’dan Prachi Patel’in haberine göre; yapılması gereken sıralama miktarını azaltacak olan bu uygulama, pratik bir DNA tabanlı belleğin anahtarı. Ancak şimdiye kadar araştırmacılar sadece 0.15 megabayta kadar veride başarılı sonuçlar elde edebildi.

### **DNA 400 Megabayttan Fazla Veri Saklayabiliyor**

Geçtiğimiz günlerde *Nature Biotechnology*’de yayınlanan bir araştırmaya göre, Washington Üniversitesinden Luis Ceze ile birlikte yeni yapılan çalışmaya liderlik eden Microsoft Research’ten Karin Strauss, depolama sistemlerinin hatasız olarak DNA’da kodlanmış 400 megabayt’lık veriye kadar rastgele erişim sağladığını belirtti ([www.nature.com/articles/nbt.4079](http://www.nature.com/articles/nbt.4079)).

Strauss’a göre, mevcut veri depolama teknolojileri, dijital içerik üretme hızımıza yetişemiyor. Sentetik DNA cazip bir depolama ortamı çünkü teoride aynı hacimdeki bir sentetik DNA, manyetik bantın 10 milyon katı veriyi depolayabilir ve binlerce yıl hayatta kalabilir. Technology Review raporları da, Microsoft Research’ün 2010’ların sonlarına doğru veri merkezi içerisinde DNA tabanlı bir depolama sistemi bulundurmayı amaçladığını gösteriyor.

Günümüz bilgisayar sistemlerinde veriler 0 ve 1’lerden oluşan bit dizileri kullanılarak depolanıyor. Çünkü kullanmış olduğumuz işlemciler için sadece 0 ve 1’lerden oluşan bit dizileri bir anlam ifade ediyor. DNA’da ise dört temel yapı taşı bulunuyor. Bunlar Adenin (A), Guanin (G), Sitozin (S) ve Timin (T). DNA’da veri saklama, dijital verilerin ikili “0 ve 1’lerini”, DNA’yı oluşturan A, C, G ve T dört tabanının dizilerine çevirir. Bu sayede çok küçük alan kaplayan bir DNA dizisine çok büyük boyutlarda veriler kodlanabilir. Kodlanmış diziler sentezlenir ve tüplerde depolanır. Bir DNA dizilim makinesi daha sonra DNA moleküllerinden dizileri ayırıştırarak verileri çözer. Fakat belirli veri dosyalarına erişmek oldukça zordur.

Verileri sayısal formuna geri getirmek için DNA dizilir ve dosyalar 0 ve 1’lere dek tekrar çözülür. Bu süreç, veri miktarı arttıkça daha da korkutucu hale gelir; rastgele erişim olanağı olmadan, belirli dosyaların bulunması için tüm veri kümesinin seri halinde sıralanması ve kodun çözülmesi gerekir. Buna ek olarak, DNA sentezi ve dizilme işlemlerindeki hata olasılığı veri kaybına neden olabilir.

### **DNA’dan Veri Almak Kolaylaşıyor**

Şimdiye kadar yapılan araştırmaların çoğu, bir tüp içinde depolanan bilgilerin tümünün dizilmesi ve çözülmesiyle ilgiliydi. Strauss ise bu konuda yaptığı açıklamada, “Depo içerisinde sadece belirli bir bölümü okumak istediğinizde, tüm verilerin dizilmesi ekonomik değildir” diyerek gidilmesi gereken noktaya işaret etti.

Çalışmalarını bu konuda yoğunlaştıran Strauss, Ceze ve meslektaşları, rastgele erişimli bir sistem oluşturmak için, akıllı kodlama algoritmaları geliştirdiler. Bu algoritmaları, adına “DNA Yükseltme” denilen DNA zincirinin binlerce kopyasını üretmek için kullanılan ve iyi bilinen bir laboratuvar tekniği olan “Polimeraz Zincirleme Reaksiyonu”na çevirdiler.

Araştırmacılar, daha önce DNA’da depoladıkları 29 kilobayttan 44 MB’a kadar büyüklükte 35 farklı dosyayla çalıştı. Her dosyayı 150 baz uzunluktaki çok sayıda DNA parçacıklarına kodladılar. Tüm bu çalışmalarını yaparken “Hata düzeltici Reed-Solomon Kodu”nu kullanan araştırmacılar, daha önceki çalışmalarının aksine daha uzun veri dizilerini DNA dizilerine dönüştüren bir kodlama şeması kullanmayı tercih ettiler.

Araştırmalar sonucunda, 13 milyondan fazla, 150 baz uzunluğunda DNA dizileri olan eşsiz bir DNA kütüphanesi ortaya çıktı. Her bir parçacığın, dosyadaki konumunu gösteren kodlanmış bir adresle başladığı bu sistemde aynı dosyaya ait parçacıkların her iki ucu, polimeraz zincir reaksiyonu başlangıç noktası olan kısa DNA zincirleri (Primer Hedef) ile çevrelendi. Strauss’a göre dizilerin nasıl dizayn edilecekleri konusunda çok dikkatli olunması gerekiyordu. Bu konuda en çok üzerinde durulan, kodlanmış veri veya adres dizileri ile çakışmayan primer hedefleri hazırlayan yaratıcı algoritmalar oldu.

Sıra DNA’yı sıralayarak verilerin okunmasına gelince, araştırmacılar, sadece seçilen bir dosyaya ait DNA parçacıklarını genişleten polimeraz zincir reaksiyonu için olan primerleri kullandı. Çoğaltılan DNA’nın tümü dizildi. Son olarak, ekibin geliştirdiği yeni bir kod çözme algoritması, benzer görünümlü dizileri birlikte kümeledi ve istatistiksel teknikler ile hata düzeltmeleri kullanarak özgün dizileri yeniden yapılandırdı; sonrasında ise dijital verileri elde etmek için çözülecek dizilimleri sağladı. Washington Üniversitesinin internet sitesinde yayınlanan makalede yer alan açıklamaya göre, Microsoft Araştırmacısı Sergey Yekhanin, çalışmanın hem sıralanan kapasite hem de işleme açısından çabayı azalttığını ve DNA’daki bilgiyi tamamen kurtarabildiğini belirtti.

ZDNet’ten Liam Tung’ın haberine göre; bu araştırma, DNA üzerinde yapılan ilk rastgele erişim çalışması olmasa da bu ölçeklerde yapılan ilk çalışma olması açısından önem taşıyor. Bu yeni kilometre taşıyla, mevcut dijital depolama teknolojilerinden çok daha yoğun ve daha dayanıklı olduğu bilinen DNA tabanlı veri depolamasının nasıl pratik olabileceği gösterilmiş oldu. Ayrıca birden fazla disiplin ve organizasyonda görev yapan araştırmacıları da bir araya getirdi.

Biyoteknoloji ve bilgisayar mimarisi arasındaki kesişmenin son derece umut verici olduğunu belirten Ceze, “Bu araştırma yayınlanmak üzere gönderildiğinden beri 400 megabayta ulaştık ve hâlâ büyüyor, büyük ölçekli DNA veri saklama alanı hakkında her gün daha fazla şey öğreniyoruz” dedi.

“Sürecin otomatikleştirilmesine odaklanıyoruz çünkü sürecin bir kısmı halen insanlar ya da pahalı makineler tarafından yürütülüyor” diyen Strauss, sistemi daha sağlam, otomatik ve ucuz yapmak istediklerini ancak bunun için yapılması gereken çok şey olduğuna işaret etti.

Sentetik DNA’yı, dünyanın daha fazla veri depolama gereksinimini giderebilecek uzun vadeli bir arşiv aracı olarak görmek isteyen Microsoft Araştırmacısı Karin Strauss da, “DNA’da veri saklama inanılmaz heyecan verici bir alan ve araştırmamızdaki ilerlemelerin *Nature Biotechnology* gibi saygın bir yayın tarafından kabul gördüğünü görmek çok güzel” dedi. 