

BİOPYTHON İLE DNA DİZİLERİNİN KARŞILAŞTIRILMASI VE ADLİ BİLİMLERDE SUÇLU SOY TAHMİNİ

Azra BAYKUŞ,Dr. Öğr. Emre GÜNGÖR

Kütahya Sağlık Bilimleri Üniversitesi Bilgisayar
Bilimleri Öğrencisi,Kütahya
Kütahya Sağlık bilimleri Üniversitesi Bilgisayar
Bilimleri Enstitüsü
azrabaykus04@gmail.com

ÖZET

Suç mahalinde bulunmuş veya kimliği belirsiz bir şahsın DNA'sının genbank[1] veri tabanıyla veya ikinci bir kişinin dna'sıyla karşılaştırılmasına izin veren ve akrabalık soya tahmini yapan bir arayüz . Biopython kütüphanesini kullanarak belirli bir benzerlik oranına göre ilişkileri değerlendirmek ve sonuçları ekrana yazdırmak amacıyla kullanılır.

Keywords:Adli bilimler,biophyton,genbank,soy tahmini.

Giriş

Locard'ın “Her temas bir iz bırakır”[2] sözünden yola çıkarak her suç mahalinde iz bırakıldığını varsayabiliriz. Adli bilimlerde, suç mahallinde bulunan biyolojik izlerin analizi, suçluların belirlenmesi ve adli süreçlerde bilimsel delil sunumu oldukça önemlidir. Adli bilimler, hukuk sistemine bilimsel ve teknolojik yöntemleri entegre eden bir alan olarak tanımlanabilir. DNA analizi, bu alanda en yaygın kullanılan ve güvenilir metodlardan biridir. DNA analizi, biyolojik materyallerin genetik özelliklerini inceleyerek suç mahallinde bulunan izleri çözümleme sürecidir. Bu izler, tükürük, kan, saç telidir ve suçlunun kimliğini belirlemek veya şüpheliler arasında akrabalık ilişkilerini saptamak için kullanılır.Fakat insanların bu analizleri yapması neredeyse imkansız ve uzun sürecek bir uğraş olması işlenen suçun kayıtsız kalmasına sebep olabilir.Bu sebeple analiz için bu uygulama şarttır.

Bu algortima Python dilinde Biopython kütüphanesi kullanılarak DNA dizileri arasında benzerlik oranını hesaplamak ve GenBank veritabanındaki DNA dizileri veya şüpheli ikinci bir dna'yı karşılaştırmak için geliştirilmiştir.

BİOPYTHON KÜTÜPHANESİ

Biopython, biyoinformatik uygulamalar için geliştirilmiş bir Python kütüphanesidir. DNA, RNA ve protein dizileri üzerinde çalışmak, bu dizileri analiz etmek ve karşılaştırmak için çeşitli araçlar sunar. Biyolojik verilerin bilgisayarla işlenmesi, adli bilimlerde genetik analizlerin hızlı ve güvenilir bir şekilde gerçekleştirilmesine olanak tanır.

Biopython'un içerisinde bulunan Seq modülü Biopython'da biyolojik dizileri (örneğin DNA, RNA, protein dizileri) temsil etmek için kullanılır, SeqIO modülü, Biopython'da çeşitli dosya formatlarından biyolojik dizileri okumak ve yazmak için kullanılır, Align modülü bu dizileri en doğru şekilde hizalayarak hızlı bir şekilde benzerlik analizi yapmayı mümkün kılar, Entrez modülü genetik, genomik ve biyomedikal bilgi alanlarında çeşitli veritabanlarına erişim imkanı sunar.

```
from Bio.Seq import Seq
from Bio import SeqIO
from Bio import Align
from Bio import Entrez
```

Resim 1. Algorimaya eklenen Biopython kütüphanesi ve kullanım şekli

2) BENZERLİK HESAPLAMA FONKSİYONU

Biopython'un Align modülü, DNA dizileri arasında hızlı bir şekilde hizalama yapmak için kullanılır. Bu modül içerisindeki PairwiseAligner sınıfı, iki dizinin benzerlik analizini gerçekleştirmek için kullanılır. Programımızda, bu sınıfı kullanarak iki DNA dizisi arasındaki benzerlik oranını hesaplayan bir fonksiyon bulunmaktadır.

```
8 def benzerlik_hesapla(dna1, dna2):
9     aligner = Align.PairwiseAligner()
10    alignments = aligner.align(dna1, dna2)
11    best_alignment = alignments[0]
12    similarity_score = best_alignment.score / max(len(dna1), len(dna2))
13
14    return similarity_score
```

Resim 2. benzerlik hesaplama fonksiyonu

“Align.PairwiseAligner()” ile bir aligner nesnesi oluşturulur. Bu nesne, iki diziyi hizalama ve benzerlik analizi yapma işlevselliğine sahiptir. “aligner.align(dna1, dna2)” ifadesi, iki DNA dizisini hizalamak için kullanılır. Bu, tüm olası hizalamaları içeren bir liste (alignments) döndürür. “alignments” listesinden en iyi hizalamayı seçmek için “best_alignment = alignments[0]” ifadesi kullanılır. Burada, en iyi hizalama, hizalamalar listesindeki ilk elemandır. Benzerlik skoru, seçilen en iyi hizalama için hesaplanır. Bu, hizalama skorunun iki DNA dizisinin maksimum uzunluğuna bölünmesi ile elde edilir. Bu sayede benzerlik, iki dizinin uzunluğuna bağlı olarak normalize edilmiş bir skor olarak elde edilir. Hesaplanan benzerlik skoru fonksiyon tarafından geri döndürülür. Bu fonksiyon, iki DNA dizisi arasındaki benzerliği sayısal bir değer olarak verir. Dönen benzerlik skoru 0 ile 1 arasında bir değer olup, 1'e yaklaştıkça benzerlik artar. Bu skor, genetik benzerlik analizinde kullanılarak DNA dizileri arasındaki ilişkilerin değerlendirilmesinde kullanılabilir.

BENZERLİK ORANINA GÖRE SOY TAHMİNİ

A ve B'nin alel frekansları A ve B 0.05 ve Hardy-Weinberg dengesine[3] göre çocuk B alelini annesinden miras almış ise A aleli babadan miras almıştır. Buna göre her bir çocuk, her bir ebeveyninden yarıymış set genetik materyal alır. Toplamda, bir çocuk, anne ve babasından aldığı genetik materyalle birlikte tam bir sete sahiptir. Yani bir çocuğun DNA benzerlik oranı her ebeveynle %50'dir.

Kardeşler, aynı anne ve babadan gelir ve genetik materyalin bir kısmını paylaşır. Her kardeş, aynı anne ve babadan farklı kombinasyonlarda genetik materyal alır. Kardeşler arasındaki DNA benzerlik oranı genellikle %50'dir, ancak tam olarak aynı değildir çünkü farklı kombinasyonlar olabilir. Ayrıca tek bir ortak ebeveynine sahip kardeşler arasında DNA benzerlik oranı %25'tir.[4]

Bir birey ile amca/teyze arasındaki genetik benzerlik oranı genellikle %25'tir. Amca/teyze, bireyin bir ebeveyninin kardeşi olduğu için genetik materyalin bir kısmını paylaşır. Kuzenler genetik olarak daha uzak akrabalar ve

genetik benzerlik oranları genellikle %12.5'tir.Kuzenler, iki ayrı kardeş grubunun birleşimi olarak düşünülebilir ve genetik materyali birbirinden farklı ebeveynlerden alırlar.

Algoritma içindeki benzerlik oranı bulma kodu açıklamak gerekirse; Tam Uyuşma (1.0): DNA dizileri tam olarak uyuşuyorsa, bu kişiler aynı olabilir.Birinci Dereceden Akrabalık (0.5 ve üzeri): Benzerlik oranı 0.5 veya daha fazla ise, bu kişiler birinci dereceden akraba olabilir.İkinci Dereceden Akrabalık (0.25 ve üzeri): Benzerlik oranı 0.25 veya daha fazla ise, bu kişiler ikinci dereceden akraba olabilir (teyze/amca-yeğen, dede/nine-torun, üvey kardeş ilişkisi).Kuzen İlişkisi (0.06 ve üzeri): Benzerlik oranı 0.06 veya daha fazla ise, bu kişiler kuzen olabilir.

PROGRAMIN İŞLEYİŞİ

Kullanıcıya iki seçenek sunulur: GenBank'ten DNA karşılaştırma veya iki kişinin DNA'sını karşılaştırma. GenBank'ten DNA karşılaştırma seçeneğinde, kullanıcının kendi DNA'sı belirli GenBank veritabanındaki DNA dizileri ile karşılaştırılır.İki kişinin DNA'sını karşılaştırma seçeneğinde, kullanıcıdan iki farklı kişinin DNA'sı alınır ve benzerlik oranı hesaplanır.Benzerlik oranına göre çeşitli kriterlere dayalı olarak sonuçlar kullanıcıya sunulur.

SONUÇLAR VE ÖNEMİ

Bu program, adli bilimlerde DNA analizi için güçlü bir araç olabilir. Suçlu soy tahmini, suç mahallinden alınan DNA örnekleri ile geniş bir veri tabanındaki potansiyel şüphelilerin DNA dizileri arasındaki benzerlikleri değerlendirerek gerçekleştirilebilir. Bu, adli bir durumu aydınlatmak ve suçluların belirlenmesine yardımcı olabilir.

Bu makalede incelenen Python programı, adli bilim uzmanlarına, kriminal vakalarda DNA analizi için hızlı ve etkili bir araç sağlamaktadır. Bu tür araçlar, adli bilimlerdeki teknolojik gelişmelerin ve bilgisayar tabanlı analizlerin güçlü bir örneğini oluşturmaktadır.

REFERANSLAR

1. [1] <https://www.ncbi.nlm.nih.gov/genbank/>
2. https://cms.galenos.com.tr/Uploads/Article_44036/TBLM-24-131.pdf
3. <http://earsiv.hitit.edu.tr/xmlui/handle/11491/8510>
4. <https://dergipark.org.tr/en/pub/ashd/issue/40351/482210>
5. [2] "Ölümler Sır Saklamaz Adli Bilimler Tarihi" Edmond Locard
6. [3] <https://familias.name/Files/manualFamilias.pdf>
7. [4]<https://sarkac.org/2022/11/2022-nobel-tip-fizyoloji-odulu-antik-dna-ve-insan-evrimi/>