

GAZİ ÜNİVERSİTESİ
MÜHENDİSLİK FAKÜLTESİ
BİLGİSAYAR MÜHENDİSLİĞİ



BIOINFORMATICS (CENG 461)

HOMEWORK 2

SELİN CANSU AKBAŞ

191180005

GİRİŞ

Araştırmacıların var olan biyolojik verilere kolayca erişebilmeleri ve elde ettikleri yeni sonuçları hızla paylaşabilmeleri için birçok kuruluş internet üzerinden ücretsiz erişilebilecek veri tabanları kurmuştur. Bu veri tabanları, bilimsel keşifleri teşvik etmek ve bilgi paylaşımını kolaylaştırmak amacıyla geniş kapsamlı biyolojik verilere erişim sağlamaktadır.

Bu veritabanlarına erişim, bilim insanlarının araştırma süreçlerini hızlandırmakta ve veri tabanlı kararlar almalarını kolaylaştırmaktadır. Araştırmacılar, genomik, proteomik, metagenomik ve diğer biyolojik verileri bu veritabanlarında araştırarak yeni hipotezler geliştirebilmekte ve bu hipotezleri test etmek için veri tabanlarında sunulan araçlardan yararlanabilmektedirler.

Bu veritabanlarının internet üzerinden erişilebilir olması, bilimsel topluluğun geniş bir kesimine hizmet etmekte ve disiplinler arası iş birliğini teşvik etmektedir. Örneğin, bir mikrobiyolog bir proteinin yapısal özelliklerini inceleyebilir ve bu bilgiyi bir biyoinformatikçiye ileterek bu proteinin işlevini daha iyi anlamalarına yardımcı olabilir.

Bununla birlikte, bu veritabanları büyük miktarda veri barındırdığından, verilerin etkin bir şekilde sorgulanması ve yorumlanması için kullanıcıların biyoinformatik becerilerine ve araçlarına erişim sağlamaları gerekmektedir. Bu nedenle, bilimsel eğitim ve kapasite geliştirme çabaları, araştırmacıların bu veritabanlarını en verimli şekilde kullanabilmelerini sağlamak için önem taşımaktadır.

Sonuç olarak, ücretsiz olarak erişilebilen biyoinformatik veritabanları, bilimsel araştırmaları hızlandırmak ve geniş bir bilimsel topluluğun ortak çabalarını teşvik etmek için kritik öneme sahiptir. Bu veritabanlarının sürekli olarak güncellenmesi ve geliştirilmesi, bilimdeki ilerlemelerin önündeki engelleri azaltacak ve yeni keşiflerin yapılmasını kolaylaştıracaktır.

1. BİYOİNFORMATİK VERİ KAYNAKLARI VE BANKALARI

Günümüzde, biyoinformatik alanında aktif olarak bilinen kuruluşlar arasında Avrupa merkezli European Bioinformatics Institute (EBI), Amerika merkezli National Center for Biotechnology Information (NCBI) ve Japonya merkezli DNA Data Bank of Japan (DDBJ) öne çıkmaktadır. Bu kuruluşlar, nükleotid dizilerinin toplanması, depolanması ve paylaşılmasında birbirleriyle iş birliği yaparak önemli elektronik veritabanlarını oluşturmuşlardır. En bilinen ve en çok kullanılanları sırasıyla ENA (European Nucleotide Archive), GenBank ve DDBJ'dir.

NCBI ve EBI gibi kuruluşlar, bilim insanlarının hizmetine internet üzerinden erişilebilen birçok biyoinformatik araç sunmaktadırlar. Bu araçlar, genomik verilerin analizi, yapısal biyoloji, gen ifadesi analizi, protein fonksiyonları ve ilaç keşfi gibi çeşitli araştırma alanlarında kullanılmaktadır. Bu raporda da, NCBI ve EBI'nın sunduğu bazı önemli biyoinformatik araçlardan bahsedilecektir.

Nükleotid dizilim veritabanları, genetik bilgiyi depolamak, yönetmek ve araştırmacıların erişimine sunmak amacıyla oluşturulmuş elektronik platformlardır. Bu veritabanları, DNA ve RNA dizileri ile ilgili çeşitli bilgileri içerir ve genellikle biyoinformatik alanındaki araştırmalarda ve genomik çalışmalarda önemli bir rol oynar.

1.1. Nükleotid Veri Kaynakları

Nükleotid dizilim veritabanları, farklı organizmalardan elde edilen DNA ve RNA dizilerini depolar. Bu veritabanları, genom projeleri, tek gen dizilimleri, transkriptomik veriler ve diğer genetik kayıtları içerebilir. Veritabanları, büyük miktarda genetik veriyi organize eder, indeksler oluşturur ve araştırmacıların verilere kolayca erişimini sağlar.

Araştırmacılar, nükleotid dizilim veritabanları aracılığıyla genetik verileri paylaşabilir, diğer bilim insanlarıyla iş birliği yapabilir ve bilgi alışverişinde bulunabilir.

GenBank: Uluslararası apta genetik veri deposu olan GenBank, DNA ve RNA dizilerini ieren en byk nkleotid dizilim veritabanıdır. GenBank, NCBI (National Center for Biotechnology Information) tarafından iletilir ve arařtırmacılar tarafından geniř apta kullanılmaktadır.

European Nucleotide Archive (ENA): Avrupa kkenli genetik verilerin saklandıėı ve paylařıldıėı bir veritabanıdır. ENA, EMBL-EBI (European Bioinformatics Institute) tarafından iletilir ve dnya apındaki arařtırmacılar tarafından eriřilebilir.

DNA Data Bank of Japan (DDBJ): Japonya merkezli nkleotid dizilim veritabanı olan DDBJ, Japonya'dan ve diėer lkelerden elde edilen genetik verileri ierir. DDBJ, National Institute of Genetics (NIG) tarafından ynetilir.

RefSeq: NCBI tarafından iletilen RefSeq veritabanı, referans genomik ve transkriptomik dizileri ierir. Bu veritabanı, yksek kaliteli ve doėrulanmıř genetik bilgi saėlamak iin kullanılır.

UniProt: Protein dizileri ve iliřkili bilgilerin depolandıėı kapsamlı bir veritabanı olan UniProt, aynı zamanda nkleotid dizilerini de ierir. UniProt, proteinlerin genomik kaynaklarını arařtırmak ve analiz etmek iin nemli bir kaynaktır.

Kullanım Alanları

- *Genomik Arařtırmalar:* Nkleotid dizilim veritabanları, genom analizi, gen ifadesi alıřmaları, poplasyon genetiėi ve evrimsel biyoloji gibi alanlarda kullanılır.
- *Tıbbi Arařtırmalar:* Genetik hastalıkların nedenlerini anlamak, kiřiselleřtirilmiř tıbbi uygulamalar geliřtirmek ve ila keřfi iin nkleotid dizilim veritabanları nemlidir.
- *Tarımsal Biyoteknoloji:* Bitki ve hayvan genomları zerine yapılan arařtırmalarda nkleotid dizilim veritabanlarından faydalanılır.

Nükleotid dizilim veritabanları, bilimsel arařtırmalarda kullanılan temel kaynaklardan biridir ve genetik bilginin küresel ölçekte paylaşılmasını ve kullanılmasını sağlar. Bu veritabanları, biyoinformatik alanındaki ilerlemelerin temelini oluşturur ve genetik biliminin daha derinlemesine anlaşılmasına katkı sağlar. Arařtırmacılar, bu veritabanlarındaki genetik verileri analiz ederek yeni bilimsel keşifler yapar ve sağlık, tarım ve çevre gibi alanlarda uygulamalı çözümler geliştirirler.

1.2. Protein Veri Kaynakları

Protein veritabanları, biyoinformatik alanında önemli kaynaklar olarak hizmet verir ve arařtırmacıların proteinlerle ilgili bilgilere erişimini sağlar. Bu veritabanları, protein dizileri, yapıları, fonksiyonları, etkileşimleri ve diğerk özellikleri içeren geniş veri setlerini barındırır. Bilimsel keşiflerin desteklenmesi, ilaç tasarımı, hastalık mekanizmalarının anlaşılması ve biyolojik süreçlerin analizi gibi birçok alanda kullanılırlar. Protein veritabanlarından bahsedecek olursak bunlardan bazıları şunlardır:

UniProt: UniProt, proteinlerle ilgili kapsamlı bilgileri içeren en büyük ve en kapsamlı protein veritabanıdır. Dünya çapındaki protein dizilerini, yapılarını, fonksiyonlarını, etkileşimlerini ve diğerk özelliklerini içerir.

Protein Data Bank: PDB, protein kristalografisi ve nükleer manyetik rezonans (NMR) gibi yöntemlerle elde edilen protein yapılarını depolar. PDB'deki veriler, protein yapılarının görselleştirilmesi, analizi ve ilaç tasarımı için kullanılır.

InterPro: InterPro, protein ailesi, alanı ve motifleri gibi protein yapısal özelliklerini analiz eden bir veritabanıdır. Protein fonksiyonları, yapıları ve evrimsel ilişkileri hakkında bilgi sağlar.

SWISS-MODEL: SWISS-MODEL, protein yapı modellemesi ve homoloji modellemesi için kullanılan bir platformdur. SWISS-MODEL, protein yapılarının tahmin edilmesi ve analiz edilmesi için çeşitli araçlar sunar.

STRING Database: STRING, protein-protein etkileşim ağlarını ve ilişkili protein fonksiyonlarını içeren bir veritabanıdır. Bu veritabanı, proteinler arasındaki fiziksel ve fonksiyonel etkileşimleri anlamak için kullanılır.

Protein Information Resource (PIR): PIR, protein dizileri ve yapıları hakkında ayrıntılı bilgi sağlayan bir veritabanıdır. PIR, protein verilerinin analizi, karşılaştırılması ve sınıflandırılması için kullanılan bir kaynaktır.

Bu protein veritabanları, araştırmacıların proteinlerle ilgili bilgileri analiz etmelerine, karşılaştırmalar yapmalarına ve biyolojik süreçleri anlamalarına olanak tanır. Proteomik çalışmalar, ilaç tasarımı, hastalık mekanizmalarının araştırılması ve biyolojik ağların modellenmesi gibi birçok alanda bu veritabanlarından faydalanılır. Ayrıca, bu veritabanları, bilimsel topluluğun veri paylaşımını teşvik ederek bilimsel keşifleri hızlandırmaya yardımcı olur ve araştırmacıların geniş çapta iş birliği yapmasını sağlar.

1.3. Yapısal Veri Kaynakları

Yapısal veritabanları, proteinlerin üç boyutlu yapıları hakkında bilgi içeren ve bu yapıları depolayan veritabanlarıdır. Proteinlerin atomik düzeydeki yapılarını, kristalografik ve nükleer manyetik rezonans (NMR) gibi yöntemlerle belirlemek mümkündür. Yapısal veritabanları, bu tür protein yapılarıyla ilgili elde edilen verileri barındırır ve araştırmacıların bu yapıları incelemeleri, analiz etmeleri ve kullanmaları için bir kaynak sağlar.

Protein Data Bank (PDB):

- Proteinlerin kristalografi veya NMR yöntemleriyle belirlenmiş yapılarını içeren en büyük yapısal veritabanıdır.
- PDB, dünya genelindeki araştırmacılar için ücretsiz olarak erişilebilir ve çeşitli araştırma alanlarında yaygın olarak kullanılır.

Protein Structure-Function Linkage Database:

- Protein yapıları ve fonksiyonları arasındaki ilişkileri incelemek için oluşturulmuş bir veritabanıdır.
- PSD, yapısal verileri fonksiyonel ve evrimsel bilgilerle entegre ederek protein işlevlerini anlamak için kullanılır.

SWISS-MODEL Repository:

- SWISS-MODEL tarafından oluşturulan protein modellemelerini içeren bir veritabanıdır.
- Araştırmacılar, SWISS-MODEL'de bulunan yapıları kullanarak yeni protein modellemeleri oluşturabilir ve analiz edebilir.

CATH (Class, Architecture, Topology, Homology) Database:

- Protein yapılarını sınıflandırmak ve analiz etmek için kullanılan bir veritabanıdır.
- CATH, yapısal benzerlikleri temel alarak proteinler arasındaki evrimsel ilişkileri ve yapısal özellikleri incelemek için kullanılır.

Yapısal veritabanları, yapısal biyoloji, ilaç tasarımı, biyoteknoloji, moleküler modelleme ve biyolojik veri madenciliği gibi alanlarda önemli bir rol oynar. Araştırmacılar, bu veritabanlarından elde edilen yapısal bilgileri kullanarak protein fonksiyonlarını anlamak, ilaç geliştirme süreçlerini yönlendirmek ve biyolojik süreçlerin detaylarını çözmek için çalışmalarını ilerletirler.

Yapısal Veritabanlarının Özellikleri

Protein Yapıları: Yapısal veritabanları, proteinlerin üç boyutlu yapılarını içerir. Bu yapılar genellikle atomik çözünürlükte tanımlanır ve proteinin amino asit dizisindeki her bir atomun konumunu gösterir.

Çözünürlük Bilgisi: Yapısal veritabanları, çözünürlük derecesiyle belirtilmiş protein yapılarına sahiptir. Kristalografik yöntemlerle belirlenmiş yapılar genellikle yüksek çözünürlükte sunulur.

Yapısal Bilgiler: Her bir yapısal kayıt, proteinin topolojisi, dönüşler, heliksler, katlamalar ve diğer yapısal özellikleri gibi ayrıntılı bilgiler içerebilir.

Protein-Ligand Etkileşimleri: Yapısal veritabanları, proteinlerin doğal ligandlar veya yapay moleküllerle etkileşimlerini gösteren verileri içerebilir.

Yapısal Veri Analiz Araçları: Bazı yapısal veritabanları, yapısal verileri analiz etmek için çeşitli araçlar ve görselleştirme seçenekleri sunar.

SONUÇ

Biyoinformatik veri kaynakları ve bankaları, modern bilimsel arařtırmalarda hayati öneme sahip olan geniş veri setlerini depolamak, yönetmek ve paylaşmak için kritik bir rol oynar. Bu veritabanları, genomik dizilerden protein yapılarına, metabolit profillerinden gen ifade verilerine kadar çeşitli biyolojik verileri barındırır. Arařtırmacılar, bu kaynaklardan elde ettikleri verilerle genetik hastalıkların temelini anlayabilir, ilaç keşfi ve geliştirme süreçlerini yönlendirebilir, tarımsal biyoteknoloji projelerinde verimliliği artırabilir ve çevresel faktörlerin biyolojik sistemlere olan etkilerini değerlendirebilirler. Biyoinformatik veri bankaları, bilimsel topluluğun veri paylaşımını teşvik ederek farklı disiplinlerden arařtırmacıların iş birliği yapmasını sağlar. Ayrıca, bu kaynaklar, veri madenciliği, yapay zekâ ve makine öğrenimi gibi gelişen teknolojilerle entegre edilerek büyük veri kümelerinin analizini ve yorumlanmasını kolaylaştırır. Böylece, biyolojik sistemlerin karmaşıklığını anlamak ve insan sağlığıyla ilgili sorunlara çözüm bulmak için önemli bir araç haline gelirler. Sonuç olarak, biyoinformatik veri kaynakları ve bankaları, bilimsel keşiflerin hızlanmasına, kişiselleştirilmiş tıp uygulamalarının geliştirilmesine ve sürdürülebilir tarım ve çevre politikalarının oluşturulmasına katkı sağlar.

KAYNAKLAR

- 1- EMBL Outstation European Bioinformatics Institute. (2012). InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Research* .
- 2- *GenBank and WGS Statistics*. (n.d.). [Www.ncbi.nlm.nih.gov.
https://www.ncbi.nlm.nih.gov/genbank/statistics/](http://www.ncbi.nlm.nih.gov/genbank/statistics/)
- 3- Madden, T. L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Research*.
- 4- National Center for Biotechnology Information. (2019).
- 5- S. Velankar, G. J. (2013). PDBe: Protein Data Bank in Europe. *Nucleic Acid Research*
- 6- Suzek, B. E. (2007). UniRef: comprehensive and non-redundant UniProt reference clusters. *Bioinformatics*.
- 7- Atalay, R. Ç. (2002). Neden Biyoinformatik. *Avrasya Dosyası, Moleküler Biyoloji ve Gen Teknolojileri Özel, Sonbahar*, 8(3), 129-141.