



BURSA TEKNİK ÜNİVERSİTESİ

MÜHENDİSLİK VE DOĞA BİLİMLERİ FAKÜLTESİ
BİLGİSAYAR MÜHENDİSLİĞİ

SEMİNER DERSİ PROJE RAPORU

Biopython ile Genetik Veri Analizi
DNA Dizilerinden Anlamlı Bilgilere Yolculuk

21360859022
SANEM COŞKUN

2023 – 2024 BAHAR DÖNEMİ

ÖNSÖZ

Bu rapor, 2024 Bahar Dönemi'nde, biyoinformatik alanında gerçekleştirdiğim seminer dersi projesi kapsamında hazırlanmıştır.

Rapor, biyoinformatik alanının tarihçesini, temel kavramlarını ve güncel teknolojilerini kapsamlı bir şekilde ele almaktadır. Bu çalışma, biyolojik veri tabanları ve analiz araçları hakkında detaylı bilgiler sunmakta, ayrıca BioPython kütüphanesi ve filogenetik ağaç oluşturma süreçlerini açıklamaktadır. Bu raporun, biyoinformatik alanında temel bilgi edinmek isteyen araştırmacılara faydalı bir kaynak olmasını umuyorum.

Bu çalışmanın hazırlanmasında emeği geçen herkese teşekkür eder, gelecekte yapacağım çalışmalarda bu deneyimlerin bana ışık tutacağını belirtmek isterim.

SANEM COŞKUN

Bursa 2024

İÇİNDEKİLER

ÖNSÖZ	2
İÇİNDEKİLER	3
ÖZET	4
BIYOİNFORMATİĞİN TARİHİ	5
1. Biyoinformatik Temel Bilgiler	6
1.1 Biyoinformatik nedir?.....	6
1.2 Biyolojik Terimler.....	6
2. Biyolojik Veri Tabanları ve Kullanılan Teknolojiler	9
2.1 Biyolojik Veri Tabanı Türleri	9
2.1.1. Nükleotid Dizilimli Veri Tabanları.....	9
2.1.2. Protein Veri Tabanları	12
2.2 Kullanılan Teknolojiler	15
2.2.1. Veri Madenciliği: BioMart.....	15
2.2.2. Dizi Hizalama: CLUSTAL	16
2.2.3. Dizi Karşılaştırma: BLAST	17
3. BioPython.....	19
2.2 BioPython kurulumu.....	20
2.2 Filogenetik Ağaç Oluşturma	22
4. Uygulama	23
4.1 Sekansları NCBI'den İndirme.....	23
4.2 Python ile Sekansları Birleştirme	24
4.3 Sekansları Hizalama ve Filogenetik Ağaç Oluşturma	24
4.4 Ağacın Görünümü	25
SONUÇ	26
5. KAYNAKÇA	27

ÖZET

Bu çalışma, biyoinformatik alanında moleküler biyoloji ve bilgisayar bilimlerinin kesişim noktasında yer alan genetik ve biyolojik verilerin toplanması, analizi ve yorumlanması üzerine odaklanmaktadır.

Biyoinformatiğin tarihçesi, temelleri ve güncel teknolojileri incelenmiş; veri tabanları ve kullanılan analiz araçları ayrıntılı olarak ele alınmıştır.

Biyoinformatik, 1950'lerde Linus Pauling ve Robert Corey'in proteinlerin ikincil yapılarının doğru tahmin edilmesi için geliştirdikleri metodolojilerle ve James Watson ile Francis Crick'in DNA'nın çift sarmal yapısını keşfetmeleriyle temelleri atılmıştır. 1960'lar ve 1970'lerde bilgisayar teknolojileri ve moleküler biyoloji metodolojilerinin gelişmesiyle, protein ve genler üzerinde daha kapsamlı analizler yapılmaya başlanmıştır. Brookhaven Protein Veri Bankası (PDB) gibi veri tabanlarının kurulması, biyoinformatikte dönüm noktaları olmuştur.

1977'de Frederick Sanger'ın DNA dizileme tekniklerini geliştirmesi, gen dizileme çalışmalarını büyük ölçüde artırmıştır. 1990 yılında başlatılan İnsan Genom Projesi, biyoinformatik alanında büyük ilerlemelere yol açmış ve bu proje, 2003 yılında başarıyla tamamlanarak, biyoinformatiğin genel kabul görmesine katkı sağlamıştır. 2000'li yıllarda, Yeni Nesil Dizileme teknolojileriyle biyoinformatik daha da ivme kazanmış ve biyolojik verilerin daha hızlı ve maliyet etkin bir şekilde toplanması mümkün hale gelmiştir.

Bu rapor, biyoinformatik alanında kullanılan veri tabanları ve teknolojileri kapsamlı bir şekilde ele almış; GenBank, ENSEMBL, UniProt ve PDB gibi veri tabanları detaylandırılmıştır. Ayrıca, veri madenciliği araçları (BioMart), dizi hizalama yazılımları (CLUSTAL) ve dizi karşılaştırma araçları (BLAST) gibi teknolojilerin kullanımı açıklanmıştır.

BioPython kütüphanesinin kurulumu ve biyoinformatik analizlerde kullanımı da raporun önemli bir kısmını oluşturmuştur. BioPython, DNA, RNA ve protein dizilerinin okunması, yazılması ve işlenmesi gibi biyoinformatik işlemleri kolaylaştıran modüller ve araçlar sağlar. Bu kütüphane, genetik veri tabanlarından veri çekme, sekans hizalama, moleküler dinamikler ve filogenetik analizler gibi çeşitli biyolojik veri işleme görevlerini destekler.

Filogenetik ağaç oluşturma süreci, biyoinformatik alanında önemli bir uygulama olarak ele alınmıştır. Bu süreç, ilgili türlerin veya genlerin DNA, RNA veya protein dizilerinin toplanması, hizalanması, evrimsel model seçimi, ağaç oluşturma ve ağacın güvenilirliğinin değerlendirilmesi aşamalarını içerir. Filogenetik ağaçlar, türlerin evrimsel geçmişi anlamak, türlerin sınıflandırılması ve genetik çeşitlilik gibi konuların incelenmesinde önemli bir araçtır.

Sonuç olarak, bu rapor biyoinformatik alanında temel bilgi edinmek isteyen araştırmacılar için kapsamlı bir kaynak niteliğindedir. Biyoinformatik, hastalıkların daha iyi anlaşılmasından yeni terapötik yöntemlerin geliştirilmesine kadar geniş bir yelpazede kullanılmakta ve bilimsel araştırmalarda temel bir araç olarak görülmektedir.

BİYOİNFORMATİĞİN TARİHİ

Biyoinformatik, moleküler biyoloji ve bilgisayar bilimlerinin kesişim noktasında doğan, genetik ve biyolojik verilerin toplanmasını, analizini ve yorumlanmasını sağlayan disiplinler arası bir alan olarak ortaya çıkmıştır.

1950'lerde, Linus Pauling ve Robert Corey'in proteinlerin ikincil yapılarının doğru tahmin edilmesi için geliştirdikleri metodolojilerle bu bilimin temelleri atılmıştır. Aynı dönemde, 1953 yılında James Watson ve Francis Crick, DNA'nın çift sarmal yapısını keşfetmişler ve moleküler biyoloji alanında devrimsel bir adım atmışlardır. [2]

1960'lar ve 1970'lerde, bilgisayar teknolojileri ve moleküler biyoloji metodolojilerinin gelişmesiyle, protein ve genler üzerinde daha kapsamlı analizler yapılmaya başlanmıştır. Brookhaven Protein Veri Bankası (PDB), proteinlerin kristalografi verilerini depolamak için bir veri tabanı olarak hizmet vermeye başlamış ve bu, biyoinformatikte bir dönüm noktası olmuştur.[2-5]

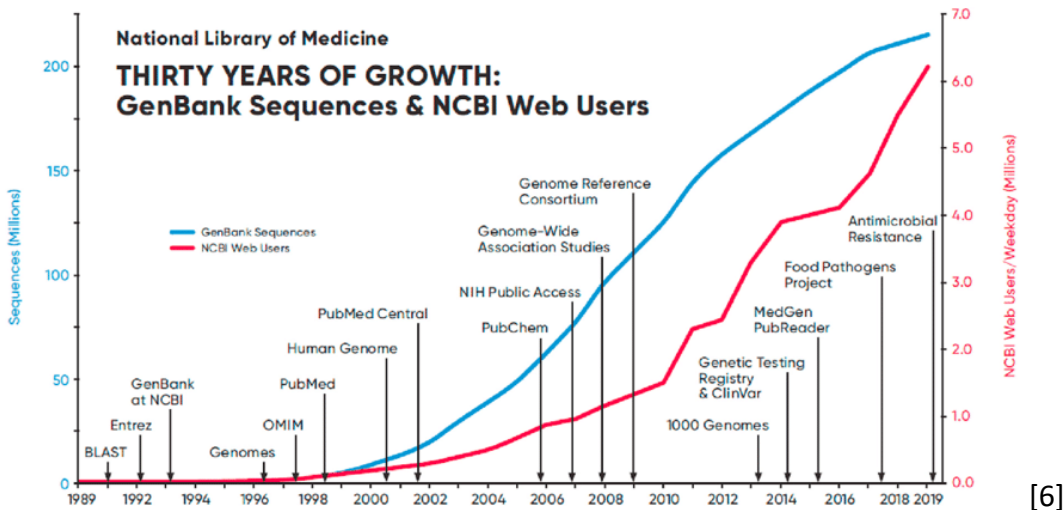
1977'de Frederick Sanger'ın DNA dizileme tekniklerini geliştirmesi, bu alanda önemli ilerlemelere yol açmıştır. Bu teknikler, 1980'lerde ve 1990'larda gen dizileme çalışmalarını büyük ölçüde artırmıştır. Biyoinformatik terimi 1980'li yılların ortalarından sonra kullanılmaya başlanmıştır.

1990 yılında başlatılan İnsan Genom Projesi, biyoinformatik alanında bir dönüm noktası olmuştur. Bu proje, insan DNA'sındaki yaklaşık 3 milyar baz çiftinin dizilimini saptamayı amaçlamış ve 2003 yılında başarıyla tamamlanmıştır. Bu başarı, biyoinformatiğin genel kabul görmesine ve genetik bilgilerin toplanmasında, saklanmasında ve analiz edilmesinde önemli bir gelişme sağlamıştır.[3]

2000'li yıllarda, biyoinformatik, Yeni Nesil Dizileme teknolojileri ile daha da ivme kazanmıştır. Bu teknolojiler, biyolojik verilerin daha hızlı ve maliyet etkin bir şekilde toplanmasını sağlayarak, araştırmacılara daha önce erişilemeyen bilgilere ulaşma imkanı sunmuştur. 454 dizileme ve daha sonra Illumina'nın Solexa dizileme platformları, biyolojik veri toplama ve analiz yöntemlerinde devrim yaratmıştır. 2008 yılında tanıtılan RNA-Seq yöntemi, transkriptom analizini ve gen ifade düzeylerinin belirlenmesini kolaylaştırmıştır.

Biyoinformatiğin evrimi, Linus Pauling gibi öncülerin geniş perspektifleri ve disiplinler arası çalışmalarıyla şekillenmiş, bu bilim dalı moleküler saat kavramını geliştirerek türler arası evrimsel ilişkilerin daha iyi anlaşılmasına olanak tanımıştır.

Günümüzde biyoinformatik, hastalıkların daha iyi anlaşılmasından yeni terapötik yöntemlerin geliştirilmesine kadar geniş bir yelpazede kullanılmakta ve bilimsel araştırmalarda temel bir araç olarak görülmektedir.



1. Biyoinformatik Temel Bilgiler

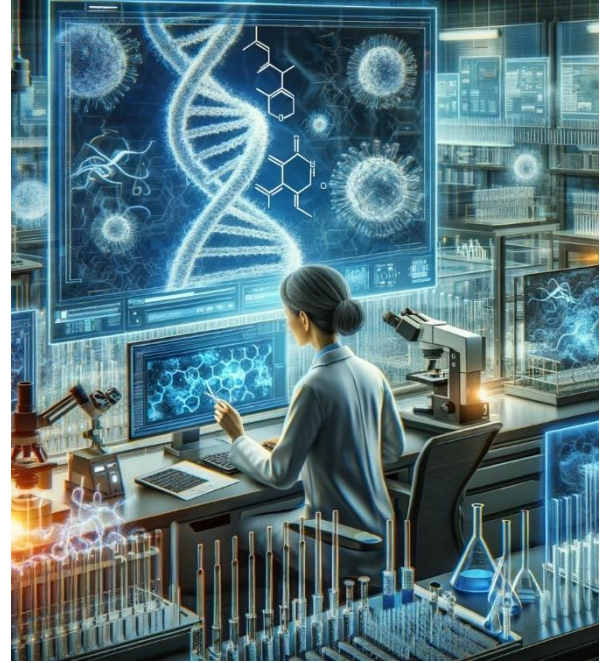
1.1 Biyoinformatik nedir?

Biyoinformatik, bilgisayar bilimlerini biyolojik bilimlerle birleştiren biyolojiye dair soruları çözme disiplini. Bu terim, ilk kez 1968 yılında "hesaplamalı biyoloji" olarak da bilinen biyoinformatik için kullanılmıştır. Özellikle moleküler genetik ve genomik alanlarına uygulanır ve biyokimyasal ile biyolojik verilerin bilgisayar yardımıyla toplanmasını, sınıflandırılmasını, depolanmasını, analiz edilmesini ve yorumlanmasını içerir. [6]

İnsan Genom Projesiyle artan biyoinformatik disiplini, moleküler çalışmalardan biyoteknolojiye, ilaç tasarımından endüstriyel tasarıma kadar bir çok alanda kullanılmaya başlanmıştır.

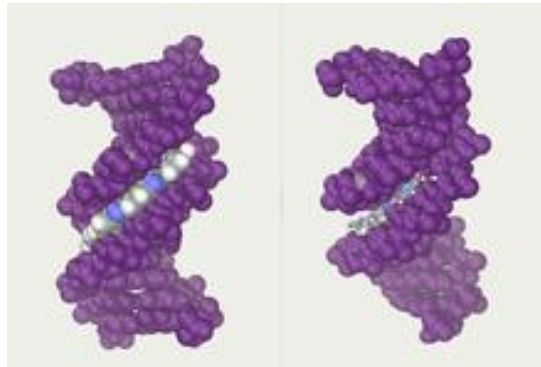
Biyoinformatik çalışmaları iki temel prensibe dayalıdır. İlki, biyolojik olarak anlamlı benzerliklere göre verilerin kıyaslanması ve gruplanması, ikincisi ise veri türünün analiz edilmesidir.

Yazılım, veri tabanları ve web kaynakları biyoinformatikle ilgili çalışmalarda sıklıkla kullanılır. Bu kaynaklar verilerin analizi ve karşılaştırılmasında büyük rol oynar



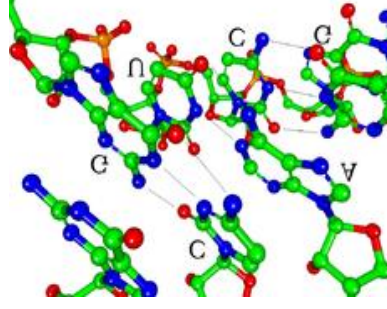
1.2 Biyolojik Terimler

DNA: Deoksiriboz nükleik asit(DNA) tüm organizmaların ve bazı virüslerin canlılık işlevleri ve biyolojik gelişmeleri için gerekli olan genetik talimatları taşıyan nükleik asittir.[7]Nükleotit olarak adlandırılan birimlerden oluşan polimerdir ve genetik bilgileri taşır.



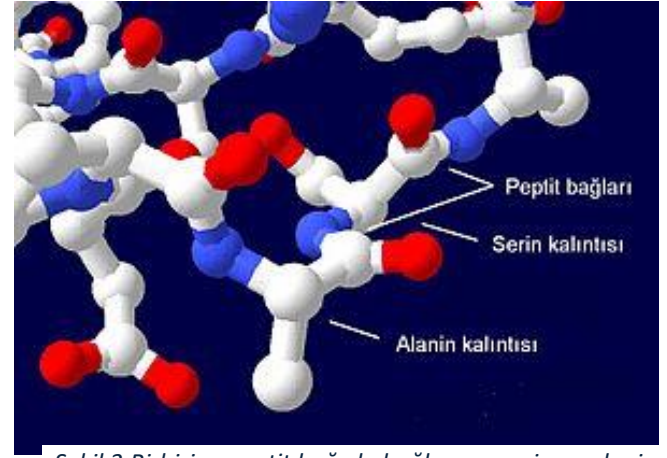
Şekil 1:Deoksiriboz nükleik asit

RNA: Ribonükleik asid (RNA), bir nükleik asittir, nükleotitlerden oluşan bir polimerdir. Yaşamın hücresel süreçlerini kontrol eden proteinlerin biyosentezine katılır. RNA DNA'dan farklı olarak tek zincirlidir, Riboz şekeri içerir ve DNA'da bulunan timin bazı yerine urasil içerir. [8]



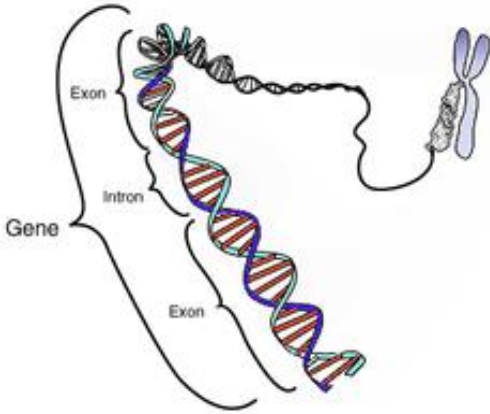
Şekil 2: Ribonükleik asit

Protein: Proteinler, bir veya daha fazla uzun amino asit artık zincirini içeren büyük biyomoleküller ve makromoleküllerdir. Proteinler organizmalar içinde, hücrelere yapı ve organizmalar sağlayarak ve molekülleri bir konumdan diğerine taşıyarak metabolik reaksiyonları katalizleme, DNA kopyalama, uyarılara yanıt verme dahil olmak üzere çok çeşitli işlevler gerçekleştirir. Proteinler, genlerinin nükleotit dizisi tarafından dikte edilen ve genellikle faaliyetini belirleyen özel 3D yapıya protein katlanmasıyla sonuçlanan amino asit dizilimlerinde birbirlerinden farklıdır.



Şekil 3: Birbirine peptit bağıyla bağlanmış serin ve alanin birimlerini gösteren bir protein yapı kesiti

Proteinler, 20 farklı L-alfa-amino asitten oluşmuş lineer polimerlerdir. Tüm amino asitler, bir alfa karbonuna birer karboksil ve amino grubu ve bir yan zincirin bağlanıyor olması gibi ortak yapısal özelliklere sahiptir.[9]



Gen: bir kalıtım birimidir. DNA'nın belli bir kısmını oluşturan nükleotid dizisidir. Genler birbirinden farklı işlevlerde ve büyüklüklerde olabilir. Gen, transkripsiyonu yapılan, düzenleyici ve fonksiyonel bölgeleri olan bir bölgedir. Kalıtımın temel fiziksel ve işlevsel birimidir.[10]

Şekil 4: Kromozom üzerinde genlerin görünüşü

Santral dogma: Biyolojik bir sistem içindeki genetik bilgi akışının bir açıklamasıdır. Genellikle “DNA RNA’yı, RNA proteini yapar.” şeklinde ifade edilir. Tek yönlü gerçekleşir.

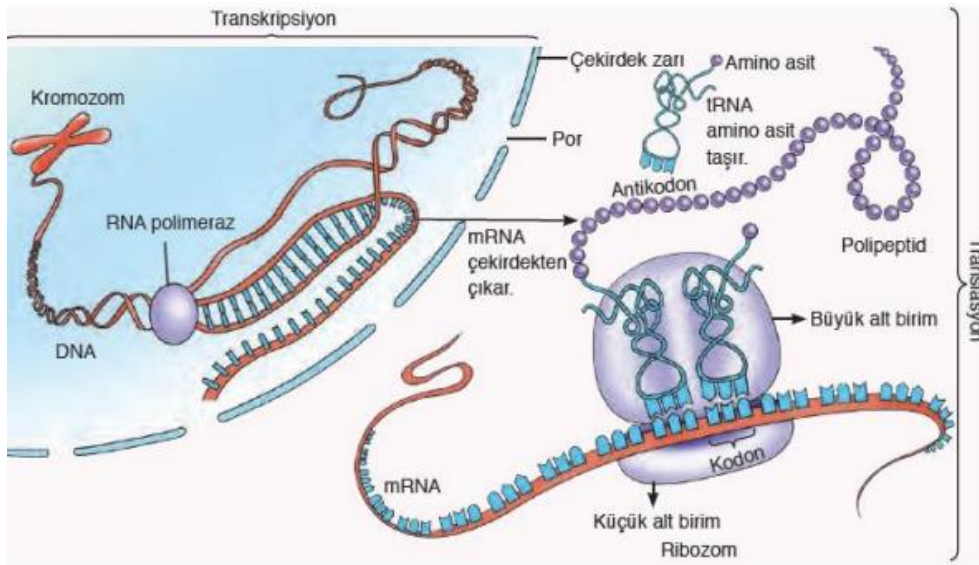


Şekil 5:Santral dogma aşamaları

Replikasyon(I): DNA’nın kendini eşlemesi olarak da bilinir. Replikasyon esnasında DNA sarmalını birbirine bağlayan zayıf hidrojen bağları Helikaz enzimi ile fermuar gibi açılır. Pürin ve pirimidin uçları açıkta kalır. Bu uçlara uygun bazları DNA polimeraz III ile tamamlayarak kendini eşleme başlar. Sürekli bir zincir oluşturmak için iki Okazaki fragmanını fosfodiester bağlarıyla Ligaz enzimi yardımıyla birleştirir. [11]

Transkripsiyon(II): DNA’ dan RNA oluşumu aşamasıdır. elde edilen RNA molekülü genellikle haberci RNA (mRNA) olarak işlev görür ve protein sentezinin gerçekleşeceği ribozoma taşınır. Bazı RNA molekülleri ise, ribozomal RNA (rRNA), transfer RNA (tRNA) gibi protein sentezinde veya hücrenin diğer işlevlerinde görev alır.

Translasyon(III): Genetik bilginin haberci RNA (mRNA) molekülünden proteinlere çevrildiği süreçtir. Bu süreç sitoplazmada, ribozomlar adı verilen hücre organellerinde gerçekleşir. Translasyon süreci, başlama, uzama ve sonlanma olmak üzere üç ana aşamadan oluşur. Başlama aşamasında, mRNA ribozoma bağlanır ve başlangıç kodonu AUG, başlangıç tRNA’sı tarafından tanınarak ribozom kompleksi oluşur. Uzama aşamasında, ribozom mRNA üzerinde ilerlerken, her kodona uygun tRNA amino asitleri getirir ve bunlar peptit bağlarıyla birbirine bağlanarak polipeptit zinciri oluşturur. Sonlanma ise, ribozomun bir durdurma kodonuna ulaşmasıyla gerçekleşir; bu noktada polipeptit zinciri ribozomdan ayrılır ve translasyon süreci tamamlanır. Bu süreçler sayesinde genetik bilgi, işlevsel proteinlere dönüştürülür.



Şekil 6:Santral dogma şema

2. Biyolojik Veri Tabanları ve Kullanılan Teknolojiler

Biyoloji, yaşamın karmaşık yapılarını, süreçlerini ve sistemlerini anlamamızı sağlayan temel bir bilim dalıdır. Modern biyoloji, genetikten hücre biyolojisine, ekolojiden moleküler biyolojiye kadar geniş bir yelpazede ilerlemektedir. Bu ilerlemelerin arkasında, biyolojik veri tabanları ve ileri teknolojiler yatmaktadır. Biyolojik veri tabanları, DNA ve RNA dizileri, protein yapıları, metabolik yollar gibi kritik bilgileri barındırır ve bu bilgiler, bilim insanlarına hastalıkların daha iyi anlaşılması, yeni ilaçların geliştirilmesi ve çevresel etkilerin değerlendirilmesi gibi konularda derinlemesine analizler yapma imkanı tanır. Ayrıca, biyoinformatik, bu veri tabanlarını daha etkin kullanabilmek için bilgisayar teknolojileri ve algoritmaları devreye sokar.

2.1 Biyolojik Veri Tabanı Türleri

2.1.1. Nükleotid Dizimli Veri Tabanları

GenBank: Amerika kaynaklı NCBI tarafından oluşturulmuş olan bir veri tabanıdır. Tüm kamuya açıktır. Belirli bir gen için arama ve DNA dizilimi yapar. Gen ile ilgili genel bilgiler dışında, gen ile ilgili makaleler ve güncellemeleri de kullanıcıya sunar.



Şekil 7: [GenBank](#)

Siteyi ilk ziyaret ettiğimizde sol kısımda veri tabanı listesini görürüz. Bu listede çeşitli veri tabanları yer alır. Örnek olarak Pubmed; biyolojik sistemlerde bulunan tüm etkileşimlere ait makalelerin bulunduğu kısımdır. Protein; bu kısımda bir protein hakkında bilgi almak için kullanabileceğimiz bir veri tabanı bulunur. Nucleotide; Özel bir genle ilgileniliyorsa bu genin ya da bu gene ait bir tek bir virüsün tüm genomunun vs. sekansını bulmak için bu veri tabanı kullanılır

Şekil 8: Neanderthal genome

ID ile de arama yapılabilir. Örnek olarak AM948965 sekans dizisi gibi. Arama kısmına ID yazdıktan sonra çıkan ekranda genomla ilgili bilgiler görünecektir.

Homo sapiens neanderthalensis complete mitochondrial genome

GenBank: AM948965.1

[FASTA](#) [Graphics](#)

LOCUS AM948965 16565 bp DNA circular PRI 26-JUL-2016
DEFINITION Homo sapiens neanderthalensis complete mitochondrial genome.
ACCESSION AM948965
VERSION AM948965.1
KEYWORDS complete genome.

Şekil 9: AM948965 bilgileri

Burada FASTA yazısına tıklayıp mitokondriyal DNA sekansını görüntüleyip projelerinizde kullanabilirsiniz.

ENSEMBL: Avrupa Biyoinformatik Enstitüsü EBI tarafından oluşturulan ENSEMBL birçok canlı için genom bilgisi içeren ve internet üzerinden rahatlıkla ulaşılabilen biyolojik bir veri tabanıdır. Genler üzerinde ayrıntılı açıklamalar içerir ve ayrıca tüm dizilimler FASTA formatında görüntülenebilir. Canlıların kromozom ve genom bilgilerine ulaşılabilir, karşılaştırma yapılabilir. ENSEMBL veri tabanında bir gen adı, bir hastalık adı veya belirli bir tür için arama yapılabilir.

Bir hastalık arandığında o hastalıkla ilgili olabilecek genleri ve transkriptleri karşımıza çıkarır. Bir gen, adı ile arandığında farklı türlerdeki ilgili geni ve transkriptlerini listeler. Bir lokasyon ile arandığında ise o lokasyon için dizilim görüntüler. ENSEMBL veri tabanı üzerinde özelleştirilmiş analiz araçları mevcuttur. BioMart ve BLAST bu analiz araçlarındandır.[12]

Tools

[All tools](#)

BioMart >

Export custom datasets from Ensembl with this data-mining tool

BLAST/BLAT >

Search our genomes for your DNA or protein sequence

Variant Effect Predictor >

Analyse your own variants and predict the functional consequences of known and unknown variants

Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotate genes, computes multiple alignments, predicts regulatory function and collects disease data. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species.

Ensembl Release 112 (May 2024)

- Many new fish genomes have been added to Ensembl
- Population frequency data is available for chicken, dog, goat and sheep through VEP
- Update to our current regulation annotation. The promoters now align with the 5' ends of known transcripts
- VEP will be updated to use the dbNSFP commercial data release

[More release news](#) on our blog

Search


All species ▾ for

Go

e.g. BRCA2 or rat 5:62797383-63627669 or rs699 or coronary heart disease


All genomes

-- Select a species -- ▾

 **Pig breeds**
Pig reference genome and 12 additional breeds

[View full list of all species](#)


Favourite genomes ✎



Human


GRCh38.p14

[Still using GRCh37?](#)



Mouse

GRCh39



Zebrafish

Ensembl Rapid Release

New assemblies with gene and protein annotation every two weeks.

Note: species that already exist on this site will continue to be updated with the full range of annotations.

Go

The Ensembl Rapid Release website provides annotation for recently produced, publicly available vertebrate

Şekil 10: ENSEMBL arayüzü

Sayfada arama kısmına görüntülenmek istenen gen ismi veya ID yazıldığında bizi gen ile ilgili kısa bilgilerin yer aldığı, genin kromozom üzerindeki yeri, transkript sayısı gibi bilgilerin olduğu bir sayfa karşılar.

Show transcript table butonuna tıklayarak tabloyu görüntüleyebilirsiniz. Tabloda bir gen için tüm kesişim varyansları kodlanmamış transkriptlerle birlikte gösterilir.

Gene: BRCA2

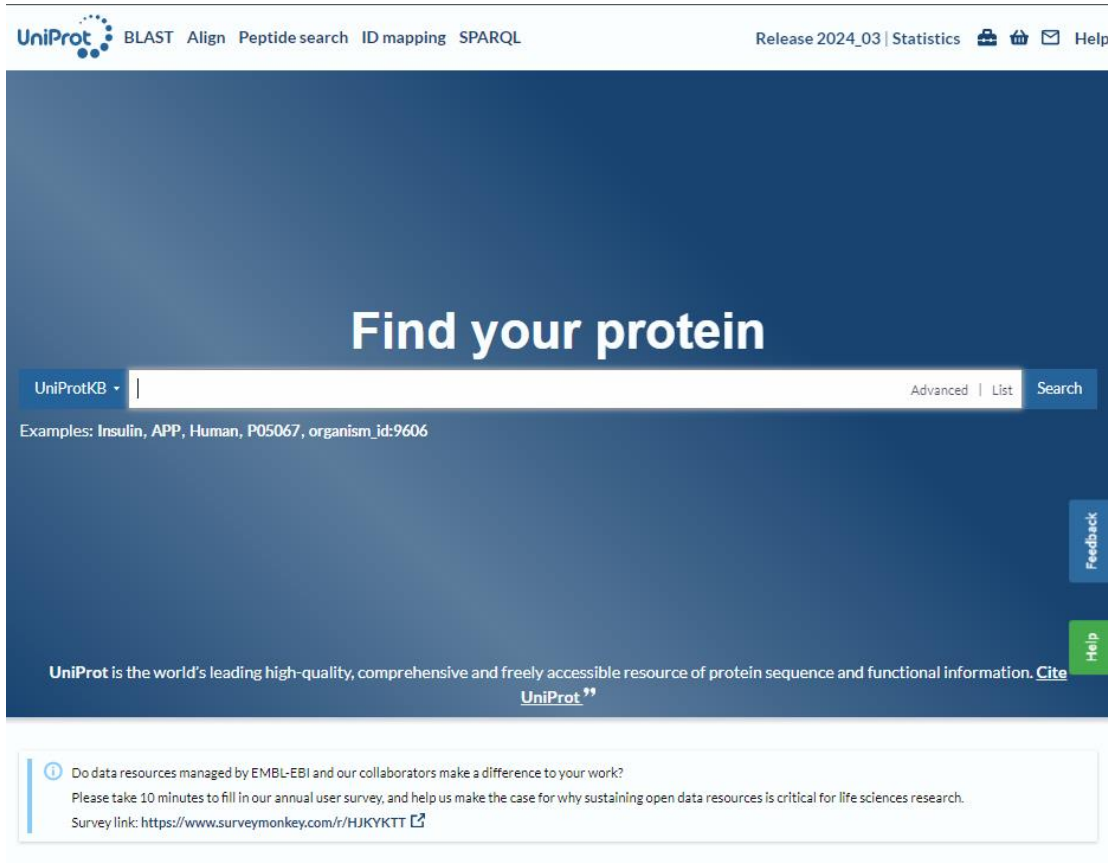
Gene: BRCA2 ENSG00000139618

<p>Description</p> <p>Gene Synonyms</p> <p>Location</p> <p>About this gene</p> <p>Transcripts</p> <p>Summary ⓘ</p> <p>Name</p> <p>MANE</p> <p>UniProtKB</p> <p>RefSeq</p> <p>CCDS</p> <p>LRG</p> <p>Ensembl version</p> <p>Other assemblies</p> <p>Gene type</p> <p>Annotation method</p> <p>Annotation Attributes</p>	<p>BRCA2 DNA repair associated [Source:HGNC Symbol;Acc:HGNC:1101]</p> <p>BRCC2, FAD, FAD1, FANCD, FANCD1, XRCC11</p> <p>Chromosome 13: 32,315,086-32,400,268 forward strand. GRCh38: CM000675.2</p> <p>This gene has 19 transcripts (splice variants), 173 orthologues and is associated with 123 phenotypes.</p> <p>Show transcript table</p> <p>BRCA2 ⓘ (HGNC Symbol)</p> <p>This gene contains MANE Select ENST00000380152, ENSP00000369497</p> <p>This gene has proteins that correspond to the following UniProtKB identifiers: P51587 ⓘ</p> <p>This Ensembl/Gencode gene contains transcript(s) for which we have selected identical RefSeq transcript(s). If there are other RefSeq transcripts available they will be in the External references table</p> <p>This gene is a member of the Human CCDS set: CCDS9344.1 ⓘ</p> <p>LRG_293 provides a stable genomic reference framework for describing sequence variants for this gene</p> <p>ENSG00000139618.19</p> <p>This gene maps to 32,889,223-32,974,405 in GRCh37 coordinates. View this locus in the GRCh37 archive: ENSG00000139618 ⓘ</p> <p>Protein coding</p> <p>Annotation for this gene includes both automatic annotation from Ensembl and Havana manual curation, see article.</p> <p>overlapping locus [Definitions] ⓘ</p>
---	--

2.1.2. Protein Veri Tabanları

Protein veri tabanı, proteinlerin amino asit dizilimleri, üç boyutlu yapıları, fonksiyonları ve diğer ilgili biyolojik bilgileri içeren, sistematik olarak düzenlenmiş bir bilgi havuzudur. Bu veri tabanları, bilim insanlarına ve araştırmacılara, proteinlerin yapısını ve işlevlerini detaylı bir şekilde inceleme imkanı sağlar. Protein dizilimleri de FASTA formatı ile gösterilir.

UniProt: protein dizilimleri ve fonksiyonu için kullanılan bir veri tabanıdır. European Bioinformatics Institute (EMBL-EBI), Swiss Institute of Bioinformatics (SIB) ve Protein Information Resource (PIR) enstitü ve kaynaklarının işbirliği sonucu oluşmuş olan bir veri tabanıdır. UniProt veri tabanı üç farklı kısma ayrılır; UniProt bilgi tabanı (UniProtKB), UniProt Reference Kümeleri (UniRef), ve UniProt Arşivi (UniParc). Bir proteine ait birçok detaylı bilgiyi araştırmacıya sunar. Her proteine ait spesifik bir kimlik numarası içerir.



Şekil 10: UniProt arayüzü

- i. Protein hakkında manüel olarak ek açıklamalar eklenen ve gözden geçirilmiş olan Swiss-Prot bölümü (sarı yıldız bölümü)
- ii. Protein hakkında ek açıklamaların otomatik olarak eklendiği ve gözden geçirilmemiş olan TrEMBL bölümü (gri yıldız bölümü) [15]

O95905 · ECD_HUMAN

Protein ⁱ	Protein ecdysoneless homolog	Amino acids	644 (go to sequence)
Gene ⁱ	ECD	Protein existence ⁱ	Evidence at protein level
Status ⁱ	UniProtKB reviewed (Swiss-Prot)	Annotation score ⁱ	
Organism ⁱ	Homo sapiens (Human)		

Entry Variant viewer [View](#) Feature viewer Genomic coordinates Publications External links History

BLAST Align [Download](#) [Add](#) Add a publication Entry feedback

Functionⁱ

Regulator of p53/TP53 stability and function. Inhibits MDM2-mediated degradation of p53/TP53 possibly by cooperating in part with TXNIP (PubMed:16849563, PubMed:23880345). May be involved transcriptional regulation. In vitro has intrinsic transactivation activity enhanced by EP300. May be a transcriptional activator required for the expression of glycolytic genes (PubMed:19919181, PubMed:9928932).

Involved in regulation of cell cycle progression. Proposed to disrupt Rb-E2F binding leading to transcriptional activation of E2F proteins (PubMed:19640839). The cell cycle-regulating function may depend on its RUVBL1-mediated association with the R2TP complex (PubMed:26711270).

May play a role in regulation of pre-mRNA splicing (PubMed:24722212). [5 Publications](#) [2 Publications](#)

GO annotationsⁱ

Access the complete set of GO annotations on QuickGO [↗](#)

Slimming set:

Cell color indicative of number of GO terms

ASPECT	TERM	Source/UniprotKB	Publications
Cellular Component	cytoplasm ↗	Source/UniprotKB	1 Publication
Cellular Component	cytosol ↗	Source/Ensembl	
Cellular Component	nucleoplasm ↗	Source/UniprotKB	
Cellular Component	nucleus ↗	Source/UniprotKB	1 Publication
Molecular Function	histone acetyltransferase binding ↗	Source/UniprotKB	1 Publication
Biological Process	fibroblast proliferation ↗	Source/Ensembl	
Biological Process	mRNA processing ↗	Source/UniprotKB-KW	
Biological Process	positive regulation of transcription by RNA polymerase II ↗	Source/UniprotKB	1 Publication
Biological Process	regulation of G1/S transition of mitotic cell cycle ↗	Source/Ensembl	
Biological Process	RNA splicing ↗	Source/UniprotKB-KW	

13

PDB: Protein Data Bank (PDB), büyük biyolojik moleküllerin, özellikle proteinler ve nükleik asitlerin üç boyutlu yapısal verilerini içeren bir veri tabanı olarak çalışır. Her bir yapı için ayrı ayrı detaylı analizler ve şematik sistemler içerir. UniProt protein veri tabanında araştırılan bir proteinin yapısı hakkında bilgi edinmek için, PDB veri tabanı UniProt' dan elde edilen kimlik numarası ile sorgulanabilir. [16]

RCSB PDB

MyPDB Contact us

Deposit Search Visualize Analyze Learn About Documentation Careers COVID-19

3D Structures ? Enter search term(s), Entry ID(s), or sequence Include CSM ? Help

Advanced Search | Browse Annotations

PDB-101 wwPDB EMDatResource NAKB wwPDB Foundation PDB-Dev

Access Computed Structure Models (CSMs) of available model organisms Learn more

Welcome Deposit Search Visualize Analyze Learn

RCSB Protein Data Bank (RCSB PDB) enables breakthroughs in science and education by providing access and tools for exploration, visualization, and analysis of:

- Experimentally-determined 3D structures from the **Protein Data Bank (PDB)** archive
- Computed Structure Models (CSM)** from AlphaFold DB and ModelArchive

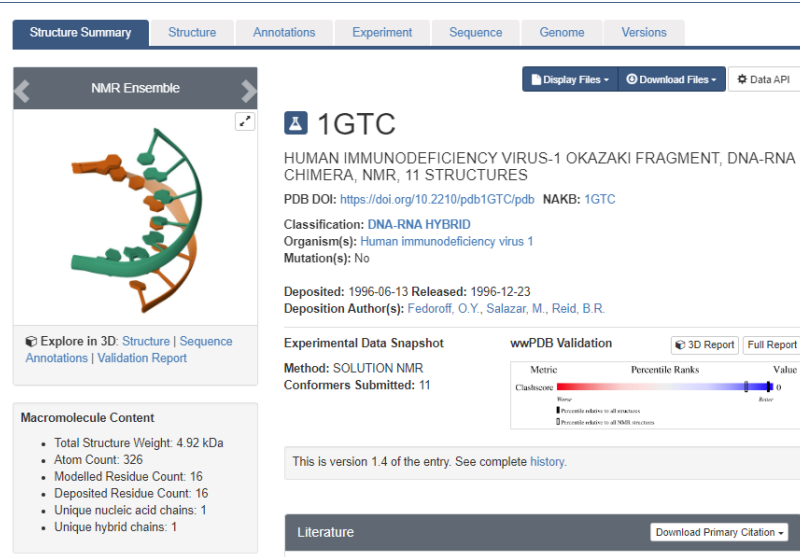
These data can be explored in context of external annotations providing a structural view of biology.

Explore NEW Features PDB-101 Training Resources

May Molecule of the Month

CFTR and Cystic Fibrosis

Şekil 11: PDB arayüzü



Şekil 12: PDB arama sonuç ekranı

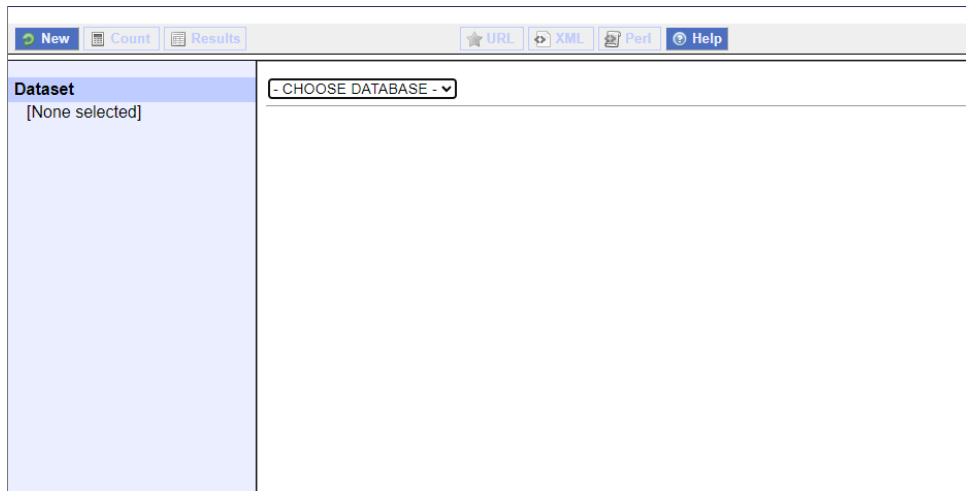
Bahsettiğimiz veri tabanları sayesinde nükleotidler, proteinler ve bu yapıların işlevleri, fonksiyonları ile ilgili verilere rahatlıkla ulaşabiliriz. Verileri filtreleme, veriler arasında bağlantı kurma gibi daha kapsamlı çalışmalar için analiz araçları geliştirilmiştir.

2.2 Kullanılan Teknolojiler

2.2.1. Veri Madenciliği: BioMart

Biomart biyolojik veri tabanları üzerinde detaylı sorgular oluşturabilmeye olanak sağlayan, açık kaynak kodlu bir biyoinformatik analiz aracıdır. 4 kıtada, 11 ülkede 46 adet veri tabanına sahiptir. [17]

Ensembl gibi veri tabanları, gen notasyonundan veri tabanı madenciliğine kadar bir dizi güçlü çevrimiçi sorguyu bioMart kullanıcılarına doğrudan sunar. bioMart, veri tabanı şemalarını veya karmaşık SQL sorgularını bilmeye gerek kalmadan büyük miktarda veri alımını sağlar. Bu sayede araştırmacılar, biyolojik verileri kolayca çekebilir ve analiz edebilir.



Şekil 12: BioMart Veritabanı Seçim Ekranı

2.2.2. Dizi Hizalama: CLUSTAL

Clustal, birden fazla sekansın hizalanmasını gerçekleştirmek için kullanılan popüler bir bilgisayar yazılımıdır. Genetik dizilimler arasındaki benzerlikleri ve farklılıkları belirlemek için geniş çapta kullanılan bu araç, biyolojik araştırmalarda ve filogenetik ağaç analizlerinde önemli bir role sahiptir. Clustal, kullanıcı dostu arayüzü ve verimli algoritmaları ile DNA, RNA ve protein dizilimlerini hızlı ve etkili bir şekilde hizalayabilir.

Clustal Omega
Multiple Sequence Alignment (MSA)

Job Dispatcher Help & Privacy Your Jobs **Input form** **Feedback**

Welcome to the new Job Dispatcher website. We'd love to hear your [feedback](#) about the new webpages!

Clustal Omega is a new multiple sequence alignment program that uses seeded guide trees and HMM profile-profile techniques to generate alignments between three or more sequences. For the alignment of two sequences please instead use our pairwise sequence alignment tools.

Input sequence ⓘ **Sequence Type**

☒ Protein ☐ DNA ☐ RNA

Paste your sequence here - or use the example sequence

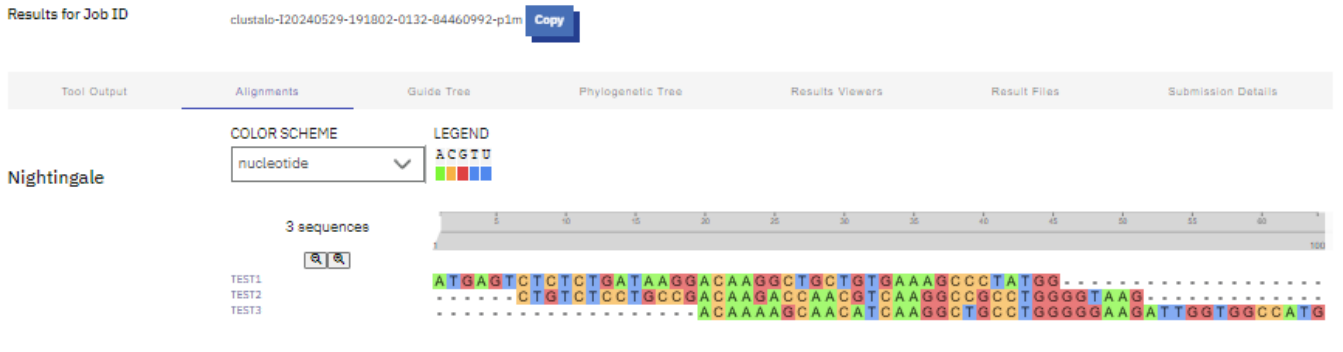
Dosya Seç **Dosya seçilmedi** **Use the example** **Clear sequence** More example inputs

Parameters **OUTPUT FORMAT ⓘ**

ClustalW with character counts

Şekil 13: CLUSTAL

Clustal FASTA formatını destekler. Varsayılan parametreler girildikten sonra “submit” butonu ile hizalama gerçekleştirilir. Show Result diyerek detaylı inceleme yapabilirsiniz.



If you use this service, please consider citing the following publication: Search and sequence analysis tools services from EMBL-EBI in 2022. More information about this bioinformatics application can be found in its [bio.tools](#) record.

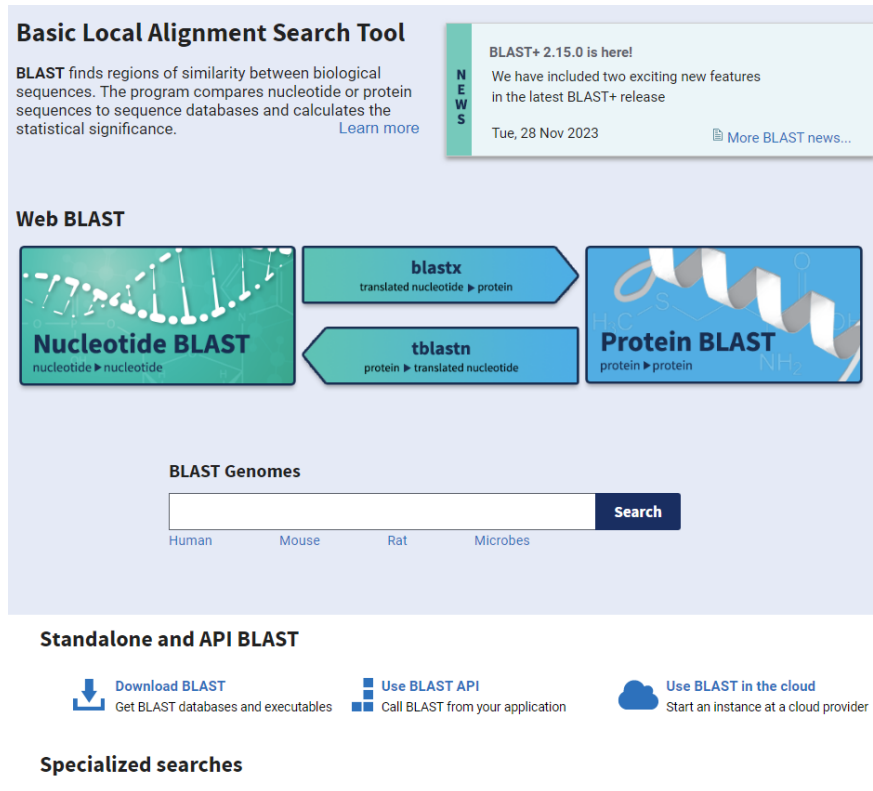
Please read the provided [Help & Privacy](#) before seeking help from our support staff. If you have any feedback or experienced any issues please let us know via [EMBL-EBI Support](#). Read our [Privacy Notice](#) if you are concerned with your privacy and how we handle personal information.

Şekil 14: Örnek hizalama

2.2.3. Dizi Karşılaştırma: BLAST

BLAST (Basic Local Alignment Search Tool), bilimsel araştırmalarda gen, protein sekanslarını veya benzer yapıları büyük veri tabanlarına karşı hızlı bir şekilde karşılaştırmak için kullanılan bir araçtır. BLAST, veri tabanlarında bulunan sekanslarla benzerlik gösteren sekansları belirlemek için geniş çapta kullanılır ve bu sayede bilim insanları yeni sekansların fonksiyonel ve evrimsel ilişkilerini daha iyi anlayabilirler.

Bir dizilim, veri tabanındaki diğer dizilimler ile ikili olarak karşılaştırılır ve iki dizilimin benzerlikleri istatistiksel olarak hesaplanır. BLAST sorgu dizilimini belli uzunluktaki kelimelere ayırır ve bu kelimeleri tüm veri tabanında bulunan dizilimler içinde arar. Sorgu kelimesi ile aynı olan bölgeler üzerinde hizalama genişletilir ve bu genişletme belli bir puanlama sistemi ile puanlandırılır. Bu puanlama sistemi sonucunda sorgu dizilimine en benzer dizilimler listelenir.[15]



Şekil 15: BLAST arayüzü

Şekil 15: BLAST sorgu ekranı

Enter Query Sequence: Bu alana, sorgulanacak DNA veya protein sekansı girilir. Sekans doğrudan yazılabilir veya bir Accession numarası, GI numarası ile referans verilebilir.

Or, upload file: Bu seçenek, bilgisayarınızdan bir FASTA formatındaki sekans dosyasını yüklemenizi sağlar.

Job Title: Arama işlemine bir başlık eklemek için kullanılır. Bu, sonuçları daha sonra kolayca tanımlamak için yararlıdır.

Choose Search Set: Kullanıcıların hangi veritabanında arama yapmak istediklerini seçmelerine olanak tanır. Bu örnekte, "Genome (all assemblies top-level)" seçeneği belirlenmiştir, bu da genel bir genom veri tabanında arama yapılacağını gösterir.

Program Selection: Farklı BLAST programlarından birini seçmek için kullanılır. Bu seçenekler, arama duyarlılığını ve performansını optimize etmek için değişir. Örneğin, "Highly similar sequences (megablast)" seçeneği, çok benzer dizilimler arasında arama yapmak için kullanılırken, "More dissimilar sequences (discontiguous megablast)" ve "Somewhat similar sequences (blastn)" seçenekleri daha az benzerlik gösteren sekanslar için uygundur.

Algorithm Parameters: Bu bölümde, BLAST aramasının daha detaylı parametrelerini ayarlayabilirsiniz. Bu ayarlar, eşleşme duyarlılığını, skrolama sistemini ve diğer teknik detayları içerebilir.

3. BioPython



BioPython, biyolojik hesaplamalar yapmak için geliştirilmiş, açık kaynaklı bir Python kütüphanesidir. Bu kütüphane, DNA, RNA ve protein sekanslarının okunması, yazılması ve işlenmesi gibi biyoinformatik işlemleri kolaylaştıran modüller ve araçlar sağlar. BioPython, genetik veri tabanlarından veri çekme, sekans hizalama, moleküler dinamikler ve filogenetik analizler gibi çeşitli biyolojik veri işleme görevlerini destekler. Biyologlar ve bilim insanları için pratik ve verimli bir araç olan BioPython, bilimsel araştırmalarda kullanılan genetik ve moleküler biyoloji verilerini işlemek ve analiz etmek amacıyla yaygın olarak kullanılmaktadır.

Biopython, FASTA , FASTQ , GenBank, Clustal, PHYLIP ve NEXUS dahil olmak üzere bir dizi ortak dizi formatını okuyabilir ve yazabilir . Dosyaları okurken, dosyadaki açıklayıcı bilgiler Biopython sınıflarının üyelerini doldurmak için SeqRecord kullanılır.

Özellikler

- Yorumlanmış, etkileşimli ve nesne odaklıdır.
- FASTA, PDB, GenBank, Blast, SCOP, PubMed/Medline, ExPASy ile ilgili formatları destekler.
- Dizi formatlarıyla ilgilenme seçeneği.
- Protein yapılarını yönetme araçları.
- BioSQL - Sekansları, özellikleri ve açıklamaları depolamak için standart bir SQL tablosu kümesi.
- NCBI hizmetlerine (Blast, Entrez, PubMed) ve ExPASy hizmetlerine (SwissProt, Prosite) erişim.
- Blast, Clustalw, EMBOSS gibi yerel hizmetlere erişim.

Hedefler

- Bioinformatik kaynaklarına standart erişim sağlamak.
- Yüksek kaliteli, tekrar kullanılabilir modüller ve betikler.
- Küme kodunda, PDB'de, NaiveBayes'te ve Markov Modeli'nde kullanılabilecek hızlı dizi işlemleri.
- Genom verilerinin analizi.

Avantajlar

- Yığın oluşturmada kullanılan mikrodizi veri türünü sağlar.
- Ağaç Görünümü tipi dosyaları okur ve yazar.
- PDB ayrıştırma, temsil ve analiz için kullanılan yapı verilerini destekler.
- Medline uygulamalarında kullanılan dergi verilerini destekler.
- Tüm biyoinformatik projeleri arasında yaygın olarak kullanılan standart bir veritabanı olan BioSQL veritabanını destekler.
- Bir biyoinformatik dosyasını belirli bir kayıt nesnesi veya dizin artı özelliklerin genel bir sınıfına ayrıştırmak için modüller sağlayarak ayrıştırıcı geliştirmeyi destekler.
- cookbook-style tarzında belgelerle açık belgeleme.

2.2 BioPython kurulumu

1. Python Kurulumunu Doğrulama

Biopython, Python 2.5 veya daha yüksek sürümlerle çalışacak şekilde tasarlanmıştır. Bu nedenle öncelikle Python'un kurulması zorunludur. Komut isteminizde aşağıdaki komutu çalıştırın.

```
C:\Users\sanem>python --version  
Python 3.10.7
```

Eğer Python düzgün yüklenmişse Python sürümünü gösterir. Aksi takdirde yeniden kurulum yapın ve komutu yeniden çalıştırın.

2. Biopython'u pip kullanarak kurmak

Biopython' u pip kullanarak yüklemek oldukça basittir. Aşağıdaki komutu komut isteminizde çalıştırarak yüklemeyi gerçekleştirebilirsiniz.

```
C:\Users\sanem>pip install biopython
```

Aynı zamanda biopython'u güncellemek için şu komutu çalıştırabilirsiniz.

```
C:\Users\sanem> pip install biopython --upgrade
```

Bu komutu çalıştırdıktan sonra, Biopython'un eski sürümleri ve NumPy (Biopython ona bağımlıdır) en son sürümleri yüklemeyi gerçekleştirmeden önce kaldırılacaktır.

3. Biopython Kurulumunu Doğrulama

Eğer Biopython başarılı bir şekilde yüklendiyse aşağıda yer alan komuta yanıt verecektir. Örneği siz de kendi bilgisayarınızda deneyebilirsiniz.

```
C:\Users\sanem>python
Python 3.10.7 (tags/v3.10.7:6cc6b13, Sep 5 2022, 14:08:36) [MSC v.1933 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> import Bio
>>> print(Bio.__version__)
1.83
>>>
```

Küçük Bir Örnek []

İlk önce fasta formatında bir dosya oluşturun ve içerisinde aşağıdaki sekansları yapıştırın.

```
>sp|P25730|FMS1_ECOLI CS1 fimbrial subunit A precursor (CS1 pilin)
MKLKKKTIGAMALATL FATMGASAVEKTISVTASVDPTVDLLQSDGSALPNSVALTYSPAV
NNFEAHTINTVVHTNDSKGVVVKLSADPVLSNVLNPTLQIPVSVNFAGKPLSTTGITID
SNDLNFASSGVNKKVSSSTQKLSIHADATRVTTGGALTAGQYQGLVSIILTKSTTTTTTKGT

>sp|P15488|FMS3_ECOLI CS3 fimbrial subunit A precursor (CS3 pilin)
MLKIKYLLIGLSL SAMSSYSLAAAGPTLTKEALNVLSPAALDATWAPQDNLTLNLTGVS
NTLVGVLTLNLTSDTVSIASNTVSDTSKNGTVTF AHETNNSASFATTISTDNANITLDK
NAGNTIVKTTNGSQLPTNLPLKFITTEGNEHLVSGNYRANITITSTIKGGGTTKKTDDK
```

Daha sonra yeni bir Python scripti oluşturun, aşağıdaki kodu yapıştırın ve kaydedin.

```
from Bio.SeqIO import parse
from Bio.SeqRecord import SeqRecord
from Bio.Seq import Seq

file = open("example.fasta")

records = parse(file, "fasta")
for record in records:
    print("Id: %s" % record.id)
    print("Name: %s" % record.name)
    print("Description: %s" % record.description)
    print("Annotations: %s" % record.annotations)
    print("Sequence Data: %s" % record.seq)
    print("Sequence Alphabet: %s" % record.seq.alphabet)
```

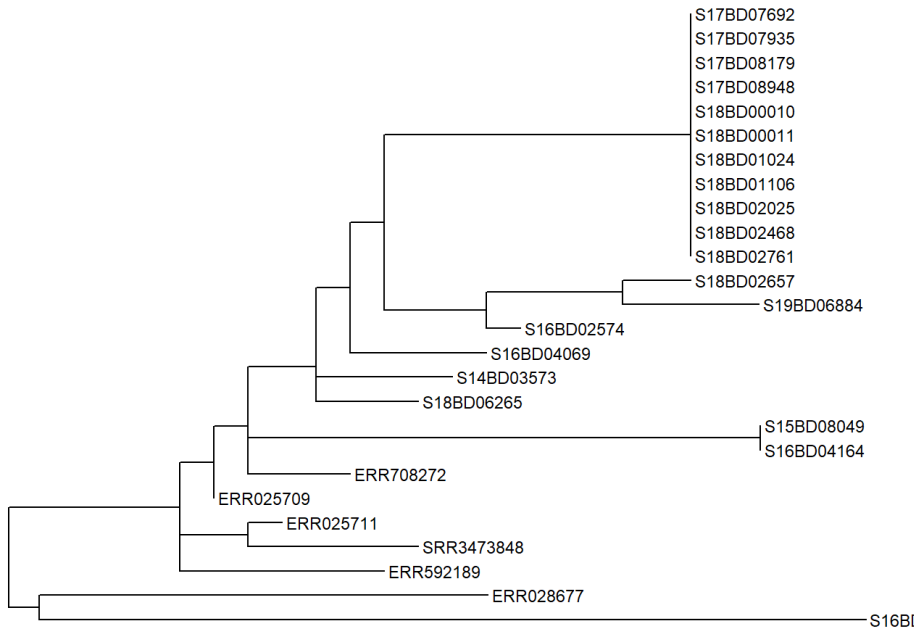
Bununla birlikte bilgileri görüntüleyebiliriz.

2.2 Filogenetik Ağaç Oluşturma

Filogenetik ağaç oluşturma, canlıların evrimsel ilişkilerini görsel olarak temsil eden ağaç yapılarının oluşturulması işlemidir. Bu ağaçlar, türlerin veya genlerin ortak atalardan nasıl evrildiğini ve birbirleriyle olan ilişkilerini gösterir. Filogenetik ağaçlar, genellikle genetik veriler (DNA, RNA veya protein dizileri) kullanılarak oluşturulur. Filogenetik ağaç oluşturma süreci genel olarak şu adımlardan oluşur:

- **Veri Toplama:** İlgili türlerin veya genlerin DNA, RNA veya protein dizileri toplanır.
- **Hizalama:** Bu diziler, homolog bölgelerinin karşılaştırılabilmesi için hizalanır. Hizalama, diziler arasındaki benzerlik ve farklılıkları belirler.
- **Model Seçimi:** Evrimsel model seçimi yapılır. Bu modeller, dizilerin evrimsel süreçte nasıl değiştiğini açıklar.
- **Ağaç Oluşturma:** Uygun algoritmalar kullanılarak (örneğin, Maksimum Parsimony, Maksimum Likelihood, Neighbor-Joining) filogenetik ağaç oluşturulur.
- **Ağacın Değerlendirilmesi:** Oluşturulan ağacın güvenilirliği çeşitli yöntemlerle değerlendirilir. Filogenetik ağaçlar, biyolojik araştırmalarda türlerin evrimsel geçmişini anlamak, türlerin sınıflandırılması, ve genetik çeşitlilik gibi konuların incelenmesinde önemli bir araçtır.

Subset tree 2



Şekil 16: Örnek filogenetik ağaç

4. Uygulama

4.1 Sekansları NCBI'den İndirme

I. NCBI GenBank'e Gidin: [NCBI GenBank](https://www.ncbi.nlm.nih.gov/genbank/)

II. **Sekans ID'lerini Arayın:** Arama kutusuna aşağıdaki sekans ID'lerini girin ve arama yapın:

- Neandertal 1 (Vindija 33.16): AM948965
- Bonobo(Pan paniscus): NC_001644
- Goril (gorilla gorilla): NC_001645
- Orangutan(Pongo pygmaeus): NC_001646
- Macaque (Macaca mulatta): NC_005943
- Human: FN677377
- Baboon: NC_001992
- Chimpanzee: NC_001643
- Denisovan: NC_013993
- Neandertal 2: FN673705
- Neandertal 3: FN677377
- Neandertal 4: KC879692
- Neandertal 5: KX663170

III. **Sonuçları Seçin ve İndirin:**

- Her bir sonucun detay sayfasına gidin.
- "Send to" menüsünden "File" ve ardından "FASTA" formatını seçin ve indirin.
- İndirilen dosyaları uygun isimlerle kaydedin (örneğin, neandertal1.fasta, neandertal2.fasta vb.)

4.2 Python ile Sekansları Birleştirme

İndirilen dosyaları tek bir FASTA dosyasında birleştirmek için aşağıdaki Python kodunu kullanabilirsiniz:

```
from Bio import SeqIO
from Bio.Align.Applications import ClustalwCommandline
from Bio import Phylo
import os

# Dosya isimleri listesi
fasta_files = ["neandertal1.fasta", "neandertal2.fasta", "neandertal3.fasta", "neandertal4.fasta",
               "neandertal5.fasta", "human.fasta", "denisovan.fasta", "chimpanzee.fasta",
               "orangutan.fasta", "macaque.fasta", "goril.fasta", "baboon.fasta", "bonobo.fasta"]

# Çıktı dosyasının adı
output_file = "combined.fasta"

# Birleştirilmiş sekansları yazmak için boş bir liste oluşturun
combined_sequences = []

# Her bir FASTA dosyasını okuyup sekansları listeye ekleyin
for fasta_file in fasta_files:
    if not os.path.exists(fasta_file):
        print(f"{fasta_file} bulunamadı.")
        continue
    print(f"{fasta_file} bulundu, işleniyor...")
    with open(fasta_file, "r") as handle:
        records = list(SeqIO.parse(handle, "fasta"))
        if not records:
            print(f"{fasta_file} dosyasında işlenecek sekans bulunamadı.")
        for record in records:
            combined_sequences.append(record)
            print(f"{record.id} sekansı {fasta_file} dosyasından eklendi.")

# Birleştirilmiş sekansları tek bir FASTA dosyasına yazın
with open(output_file, "w") as output_handle:
    SeqIO.write(combined_sequences, output_handle, "fasta")

print(f"{output_file} dosyasına {len(combined_sequences)} sekans birleştirildi.")
```

4.3 Sekansları Hizalama ve Filogenetik Ağaç Oluşturma

Birleştirilmiş FASTA dosyasını kullanarak filogenetik ağaç oluşturabilirsiniz

1) ClustalW ile Sekansları Hizalama:

```
# ClustalW kullanarak sekansları hizalama
clustalw_exe = "C:\\Program Files (x86)\\ClustalW2\\clustalw2.exe" # ClustalW'nin tam yolu
in_file = "combined.fasta"
clustalw_cline = ClustalwCommandline(clustalw_exe, infile=in_file)
stdout, stderr = clustalw_cline()
print("Sekanslar hizalandı.")
```

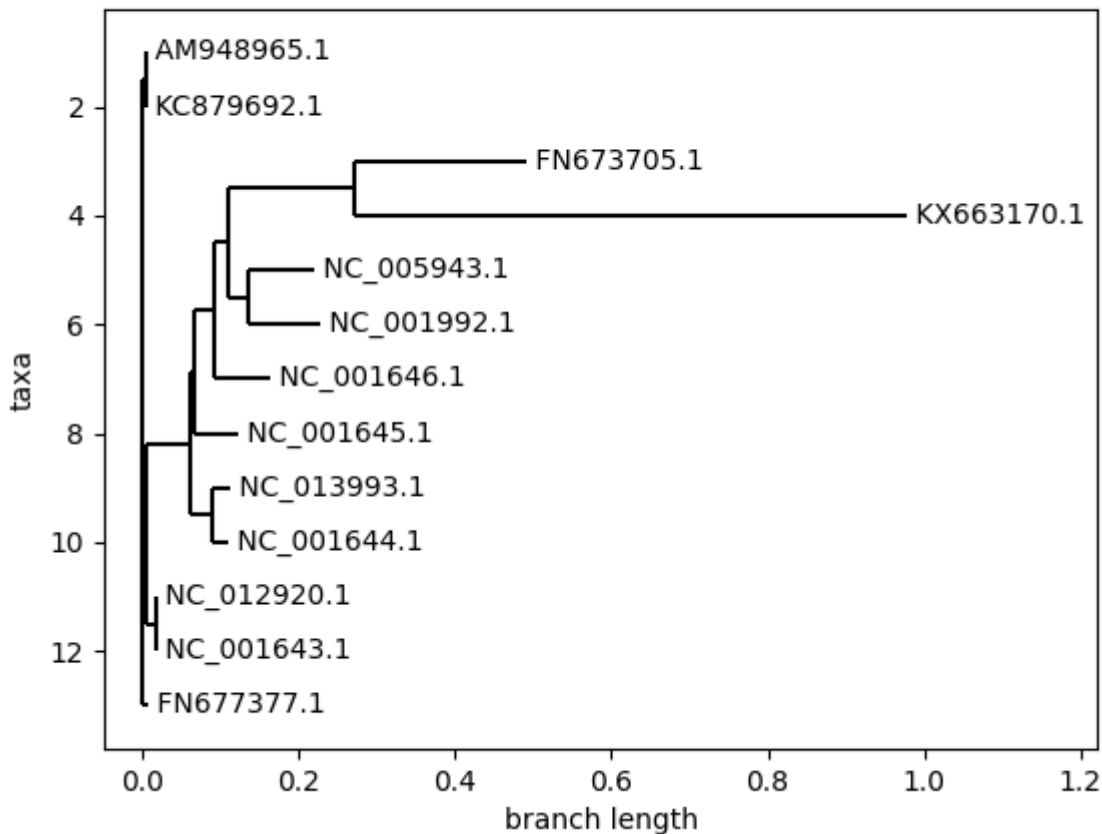
2) Filogenetik Ağaç Oluşturma ve Görüntüleme:

```
# Filogenetik ağaç oluşturma ve görüntüleme
tree = Phylo.read("combined.dnd", "newick")

# Filogenetik ağacı çizme
Phylo.draw(tree)
import matplotlib.pyplot as plt

# Filogenetik ağacı çizme ve görüntüleme
fig = plt.figure(figsize=(10, 5), dpi=100) # Görüntü boyutunu ve çözünürlüğünü ayarlayın
axes = fig.add_subplot(1, 1, 1)
Phylo.draw(tree, do_show=False, axes=axes) # Filogenetik ağacı çizme
plt.show() # Grafiği gösterme
```

4.4 Ağacın Görünümü



SONUÇ

Bu rapor, biyoinformatik alanında gerçekleştirilen çalışmaların kapsamlı bir incelemesini sunmaktadır. Biyoinformatik, moleküler biyoloji ve bilgisayar bilimlerinin kesişim noktasında doğmuş ve genetik ile biyolojik verilerin toplanması, analizi ve yorumlanmasını sağlayan önemli bir disiplinler arası alandır. Tarihsel gelişimi, teknolojik ilerlemeleri ve çeşitli uygulama alanları, biyoinformatiğin biyolojik bilimler için vazgeçilmez bir araç haline gelmesine katkıda bulunmuştur.

Raporda ele alınan temel konular arasında, biyoinformatik veri tabanları ve kullanılan teknolojiler öne çıkmaktadır. GenBank, ENSEMBL, UniProt ve PDB gibi veri tabanları, genetik ve protein verilerinin saklanması ve analizi için kritik öneme sahiptir. Bu veri tabanları, araştırmacıların biyolojik verileri daha etkin bir şekilde kullanmalarına olanak tanıyan güçlü araçlar sunar. Özellikle veri madenciliği araçları (BioMart), dizi hizalama yazılımları (CLUSTAL) ve dizi karşılaştırma araçları (BLAST), biyoinformatik analizlerin temel taşlarını oluşturmaktadır.

BioPython kütüphanesi, biyoinformatik analizlerde sağladığı kolaylıklar ve esneklik ile raporda önemli bir yer tutmaktadır. DNA, RNA ve protein dizilerinin işlenmesi, genetik veri tabanlarından veri çekme, sekans hizalama ve filogenetik analizler gibi çeşitli biyolojik veri işleme görevlerini destekleyen BioPython, biyologlar ve bilim insanları için pratik ve verimli bir araçtır.

Filogenetik ağaç oluşturma süreci, türlerin evrimsel ilişkilerini anlamak ve genetik çeşitliliği incelemek için kullanılan önemli bir yöntemdir. Bu süreç, ilgili türlerin veya genlerin DNA, RNA veya protein dizilerinin toplanması, hizalanması, evrimsel model seçimi, ağaç oluşturma ve ağacın güvenilirliğinin değerlendirilmesi aşamalarını içermektedir.

5. KAYNAKÇA

- [1] - <https://tr.wikipedia.org/wiki/Biyoenformatik>
- [2] - <https://dergipark.org.tr/tr/download/article-file/196783#:~:text=Biyoinformati%C4%9Fin%20ba%C5%9Flang%C4%B1c%C4%B1n%C4%B1%20kesin%20olarak%20belirlemek,biyoinformatik%20i%C3%A7in%20ba%C5%9Flang%C4%B1%C3%A7%20kabul%20edilebilir.>
- [3] - <https://tip.baskent.edu.tr/kw/upload/464/dosyalar/cg/sempozyum/ogrsmpzsnm16/16.P12.pdf>
- [4] - <https://biyoinformatikdunyasi.blogspot.com/2018/08/biyoinformatige-giris.html>
- [5] – Uygulamalı Biyoinformatik, Paul M.selzer, Richard J.marhofer, Oliver Koch-2019
- [6] - https://www.researchgate.net/figure/Growth-of-GenBank-sequences-and-NCBI-web-users-through-2019-Figure-borrowed-from-the_fig2_343364994
- [7] - <https://tr.wikipedia.org/wiki/DNA>
- [8]- <https://tr.wikipedia.org/wiki/RNA>
- [9]- <https://tr.wikipedia.org/wiki/Protein>
- [10]- <https://tr.wikipedia.org/wiki/Gen>
- [11]- [https://tr.wikipedia.org/wiki/Santral_dogma_\(molek%C3%BCler_biyoloji\)](https://tr.wikipedia.org/wiki/Santral_dogma_(molek%C3%BCler_biyoloji))
- [12]- <https://biyoinformatikdunyasi.blogspot.com/2018/11/biyoinformatikte-veri-tabanlar.html>
- [13]- <https://www.ncbi.nlm.nih.gov/genbank/>
- [14]- <https://www.ensembl.org/index.html>
- [15]- <https://ab.org.tr/ab14/bildiri/90-uzun.pdf>
- [16]/[https://tr.wikipedia.org/wiki/Protein_Data_Bank#:~:text=Protein%20Data%20Bank%20\(PDB\)%2C,bir%20veri%20taban%C4%B1%20olarak%20%C3%A7al%C4%B1%C5%9F%C4%B1r.](https://tr.wikipedia.org/wiki/Protein_Data_Bank#:~:text=Protein%20Data%20Bank%20(PDB)%2C,bir%20veri%20taban%C4%B1%20olarak%20%C3%A7al%C4%B1%C5%9F%C4%B1r.)
- [17]- <https://www.bioconductor.org/packages/release/bioc/html/biomaRt.html>
- [18]- <https://www.ebi.ac.uk/jdispatcher/msa/clustalo>
- [19]- <https://blast.ncbi.nlm.nih.gov/Blast.cgi>
- [20]- https://www.tutorialspoint.com/biopython/biopython_installation.htm