

Oncelikle hücrelerden veya dokulardan numune elde ediliyor.

DNA ayrıştırılması yapılan örnek daha sonra **sanger** metoduyla elde **florasan** boyama yöntemiyle bazların boyanmasıyla elde ediliyor.

Farklı dalga boyuna sahip farklı bazlar, dedektörden geçerek bize bu grafiği oluşturuyor.

Daha sonra bu diziler veritabanlarına (ncbi, genbank) uygun formatta yükleniyor. (30 gün)



Zincirin tek tarafı



### Homo sapiens M-phase phosphoprotein 10 (MPHOSPH10), mRNA

NCBI Reference Sequence: NM\_005791.3

[GenBank](#) [Graphics](#)

```
>NM_005791.3 Homo sapiens M-phase phosphoprotein 10 (MPHOSPH10), mRNA
ATGCTGCATTGTGTCGGGAGTTGCTGACAGCCATGGCGCCGAGGTCTGGCGTGCACGGACCCTGGAGCG
GTGCTGACGGAAGTCGGCAAAGCCACGGGTGGCCCGAGTGCTTCTCACGATTCAAGAGGGATTGGCA
TCAAAGTTCACCTCTTTAACAAAAGTGCTTTATGACTTTAATAAAATATTAGAGAATGGTAGGATCCATG
GAAGCCCTTGCAAAAACCTTGATAGAAAATTTGATGATGAGCAGATTTGGCAACAACCTGGAATTGCA
AAATGAACCAATTTTACAATACTTTAGAAATGCAGTTAGTGAACAATTAATGATGAAGATATCAGTCTT
CTCCAGAGAGTGAAGAACAGGAACGTGAAGAGGATGGTTCCAGAGATAGAGGCTGATGACAAGGAGGACC
TAGAAGATTTAGAGGAGGAGGAAGTGTCCGACATGGGTAATGATGATCCTGAAATGGGTGAGAGAGCTGA
AAACTCAAGCAAATCTGATCTGAGGAAAAGCCCCGTTTTTCAGTGATGAGGATTCTGACCTTGACTTTGAT
ATCAGCAAATTTGAAACAGCAGAGCAAGGTGCAAAAACAAAGGACAGGGAAAACCAAGAGAAAAGTCCATAG
```

# FASTA FORMAT

- Fasta '>' ile başlayan açıklama satırı dizinin adını veya tanımlayıcısını, ayrıca ek bilgiler içerebilir. Ardından gelen satırda sıralama yer alır. Ardından bir başka sıralama yer alabilir.
- Örnek olarak >NC\_002023.1 Influenza A virus (A/Puerto Rico/8/1934(H1N1)) segment 1, complete sequence  
AGACGACCTA...

# UGENE NEDİR?

DNA ve protein dizileri, çoklu dizi hizalaması, 3B yapılar ve filogenetik ağaçlar için izleyiciler ve editörler.

- Dizi içerisinde arama yapılabilir.
- ClustalW, ClustalO, Muscle, K-Align, Mafft, T-Coffee gibi algoritmalarla multiple sequence alignment yapılabilir.
- Açık okuma pencereleri gözlemlenebilir.
- Filogenetik ağaç yapılabilir.
- Protein yapısı incelenebilir.
- FASTA (.fa) GenBank (.gb) Clustal (.aln) PDB (.pdb) gibi formatları destekler.

Uygulamayı açınca iki seçeneğimiz var. Ya dosyayı biz yükleyebiliriz.

Yada veritabanından indirebiliriz..

Öncelikle veritabanından örnek bir dosya indirelim.

Burada da iki seçenek var.

**NCBI Gebank arama** kısmında örneğin covid yazıp **ara** dediğimizde ilgili nükleotit dizilerini indirebiliyoruz. **Yada:**

**Uzaktan veritabanına eriş** modunda: doğrudan dizinin kimliğini yazabiliyoruz mesela makalelerde bu kimlik numarası yazıyor.

Buradaki örnekte: **1363** adlı dosya örneğin bir **MSV** adlı virüsün genom dizilimini indirebiliriz.

## 1363 adlı dosyayı açtığımızda..

Bu dosya **genbank** formatında bir dosya olduğunu görüyoruz. Burada yapabileceğimiz çeşitli işlemler var.

**Uzak Veritabanından Veri Al**

Kaynak Kimliği:

Veritabanı:  NCBI GenBank (DNA sequence)  NCBI protein sequence database

Dizine kaydet:  ENSEMBL  PDB  SWISS-PROT

Projeye ekle  Uygun diziyi  UniProtKB/Swiss-Prot  UniProtKB/TrEMBL

**İpucu: Genbank Kimlikleri boşluk veya noktalı virgülle ayırarak birden çok öge indirebilirsiniz.**

**NCBI Sıra Araması**

Term: Tüm alanlar covid

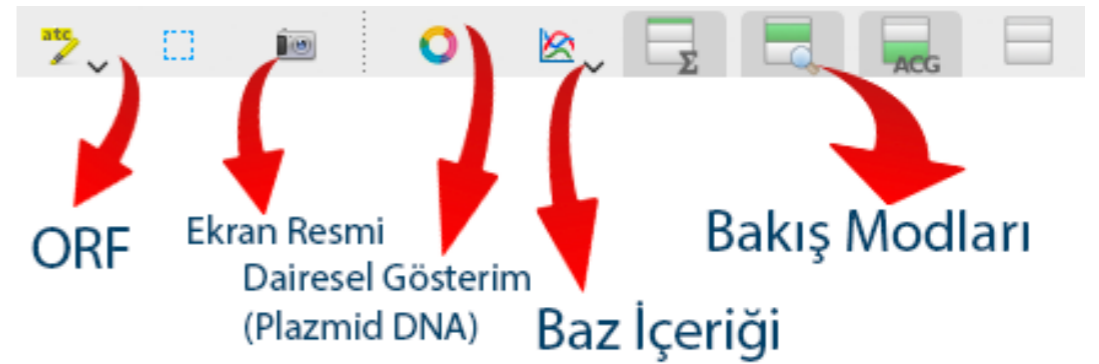
Veritabanı: nucleotide

Arama sorgusu: covid

Sonuçlar:

KİMLİK	Azalan	Boyut
OR759453	Human coronavirus NL63 isolate E202011025 RNA-dependent RNA polymerase gene, partial cds	393
OR759452	Human coronavirus OC43 isolate E202011017 RNA-dependent RNA polymerase gene, partial cds	380
OR759451	Human coronavirus 229E isolate 202103333 RNA-dependent RNA polymerase gene, partial cds	359
OR759450	Human coronavirus HKU1 isolate 202101291 RNA-dependent RNA polymerase gene, partial cds	384
OR759449	Human coronavirus HKU1 isolate 202101280 RNA-dependent RNA polymerase gene, partial cds	372
OR759448	Human coronavirus 229E isolate 202101279 RNA-dependent RNA polymerase gene, partial cds	334
OR759447	Human coronavirus NL63 isolate 202012228 RNA-dependent RNA polymerase gene, partial cds	366
OR759446	Human coronavirus 229E isolate 202011176 RNA-dependent RNA polymerase gene, partial cds	271

Sonuç sınırı: 20



1. Burada arayüzümüzde Açık okuma pencerelerini gösterebiliyoruz.
2. Eğer dairesel bir DNA ise dairesel görünüm
3. AT GC oranları (Örneğin **GC** seviyelerinin yüksek olduğu yerlerde 3lü hidrojen bağı olduğu stabilitenin yüksek olduğunu, dolayısıyla da erime sıcaklığının arttığını söylemek mümkün.) + DNA **esnekliği azalıyor** GC çok olan yerde grafikte görmek mümkün. Yaklaşırsak G veya C sayılarının arttığını görebiliyoruz.
4. Modlar arasında geçiş yapabiliyoruz daha ayrıntılı görmek için.

# Sağdaki Arayüz

1. Dizi içerisine arama yapma yapabiliyoruz. (ÖRN : **ATG**) Altta sonuç sayısı yazıyor.
2. Burada belirlediğimiz etiketler için için açık okuma pencereleri için veya farklı etiketler için farklı renkleri gösterebiliyoruz. (Ek açıklamaları göster açık olmalı.)
3. Bu kısımda da **seçtiğimiz (seç) kısmın uzunluğunu, içeriğini, moleküler ağırlığını, erime sıcaklığını** (PCR için önemli) **ve baz içeriklerini.** **CTRL + A** ve **scroll** istediğimiz aralığı seçebiliyoruz. (ssDNA = Tek Sarmallı), (dsDNA = Çift Sarmallı)



# Kromatogram Görüntüleyici

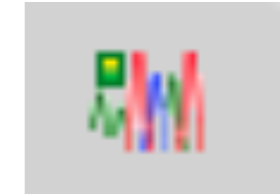
## Dizi nasıl elde ediliyor?

**(A01.abi) Örneğin burada;** DNA dizileme cihazlarından elde edilen bir dizi var fakat bu dizi .abi formatında olduğu için sanger sonucunu görebiliyoruz dizinin.

**OPT + Scroll** yaparak yakınlaştırıp görüntüleyebiliriz.

Az önce bahsettiğimiz dalga boylarına göre dizinin nasıl oluşturulduğunu görebiliyoruz.

Ayrıca açık okuma pencereleri yine dizide gözüküyor. Onlara da birazdan geçeceğim.



Ardından eğer bu bir Kromozom ise  
Görüntüleme için  
**(Kromozom.ugenedb)** kromozom  
görüntüleme aracını kullanacağız.

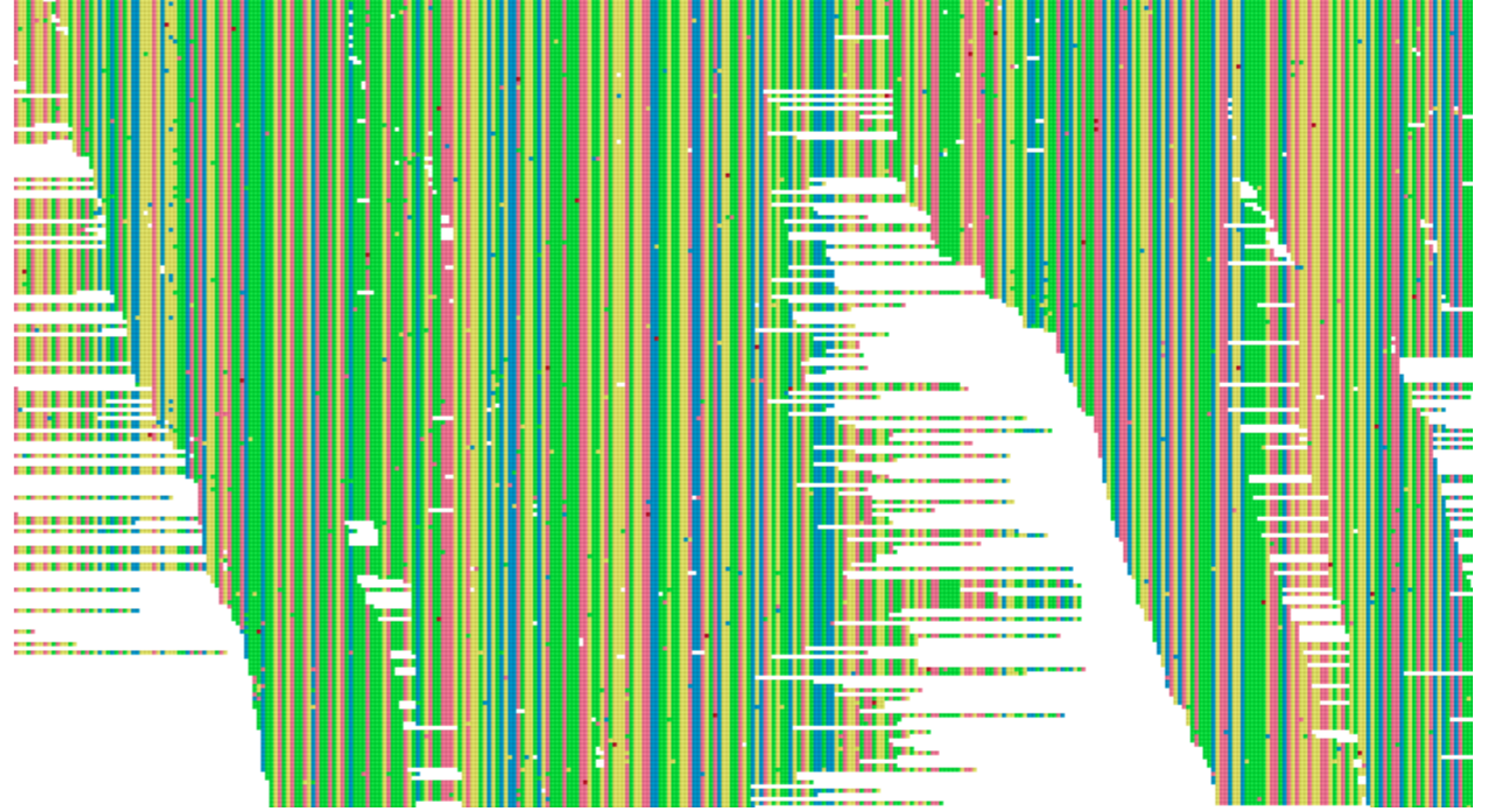
**Boşluklar** ekson

**Dolu bölgeler** intron

olarak kabul ediliyor.

Kayma ise ışığın dalga boyundan  
kaynaklanıyor. Bazı farklar mevcut buda  
ışığın kaymasından kaynaklı farklılıklar.  
(Örnekte 321 defa döngü yapılmış.  
Doğruluğu arttırmak için.)

Aslında yukarıdan aşağı farklılar snp'leri  
ifade ediyor gibi gözüksede aslında  
bunlar ufak okuma hataları.





Önceki **haftalarda** [https://www.sciencedirect.com/science/article/abs/pii/S147692711930088X?](https://www.sciencedirect.com/science/article/abs/pii/S147692711930088X?via%3Dihub) adresindeki makalede CSF3 genindeki snp'lerden bahsetmiştik.

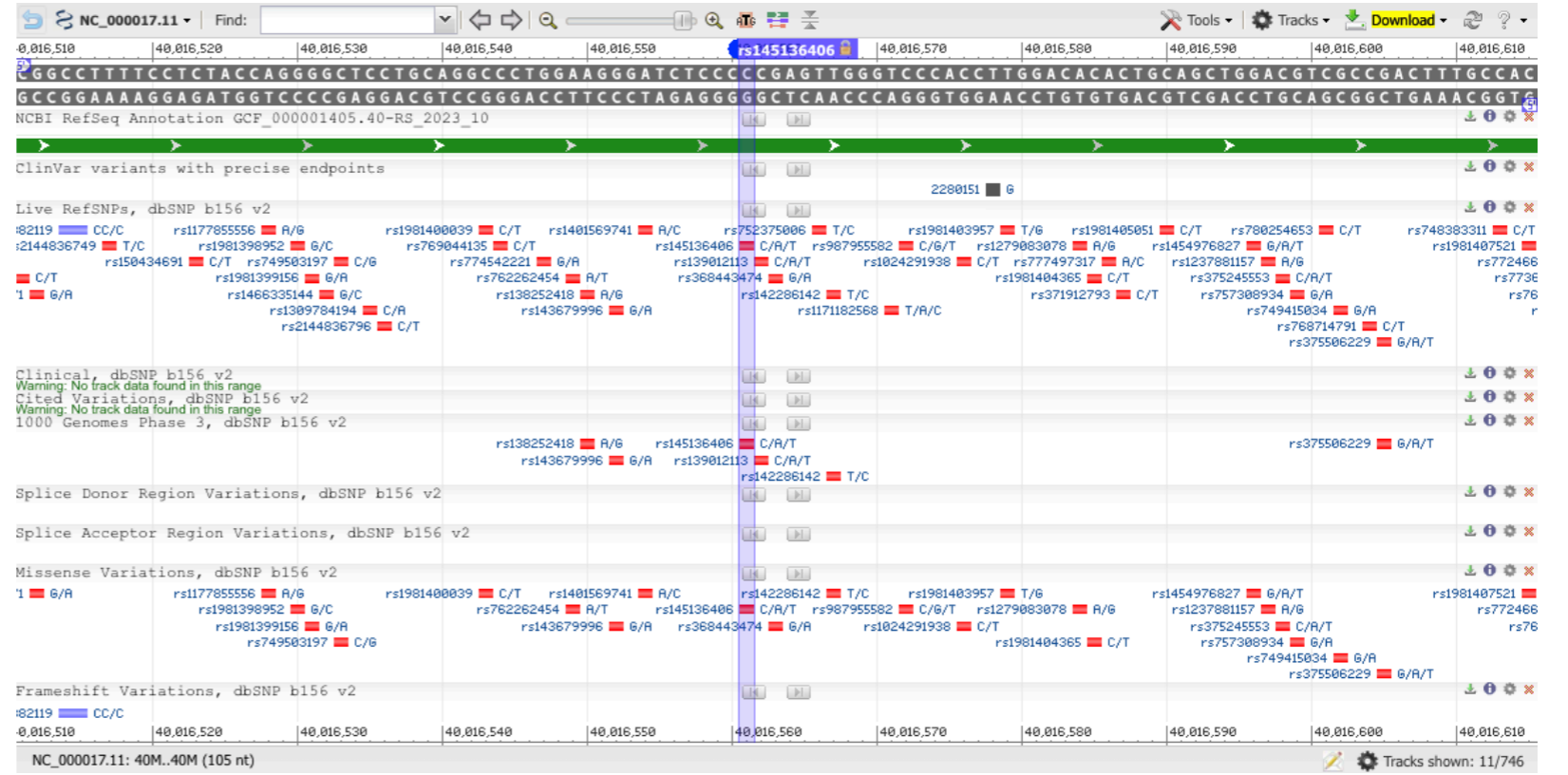
<https://www.ncbi.nlm.nih.gov/snp/rs145136406> (**CSF3 ile ilgili makalede**) adresinde bu snp'ye ulaşmak mümkün. 17. kromozomun p yani kısa kolundaki belirli bir aralığın dizisi yüklenmiş.

Zaten **yeşil** ile gösterilen bölgede bu değişikliğin **CSF3 geninde** olduğu gösterilmiş.

Biz burada en uzak kısmı seçip görünen aralığı indirebiliriz. Tabi böyle yaparak bütün kromozomu indirmiş olduk.

**Dosyayı açtığımızda** açık okuma pencerelerini görüntüleyip bu alanların **ATG veya AUG başlat kodonuyla başladığını görebiliriz. Bu bölgeler esas proteini kodlayan bölgelerdir.** Diğer bölgeler işlevseldir.

**ORF'leri göster diyip gösterebiliriz.**



*Sayfanın altında ilgili SNP seçili download seçeneceğini seçip indirebiliriz.*



Başka bir örnekte kovid-19 için 2 farklı varyantı indirdim. (**İkisini beraber yüklememiz gerekiyor.**)

1. Wuhan seafood market pneumonia virus isolate Wuhan-Hu-1, complete genome NCBI Reference Sequence: NC\_045512.1
2. Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome NCBI Reference Sequence: NC\_045512.2

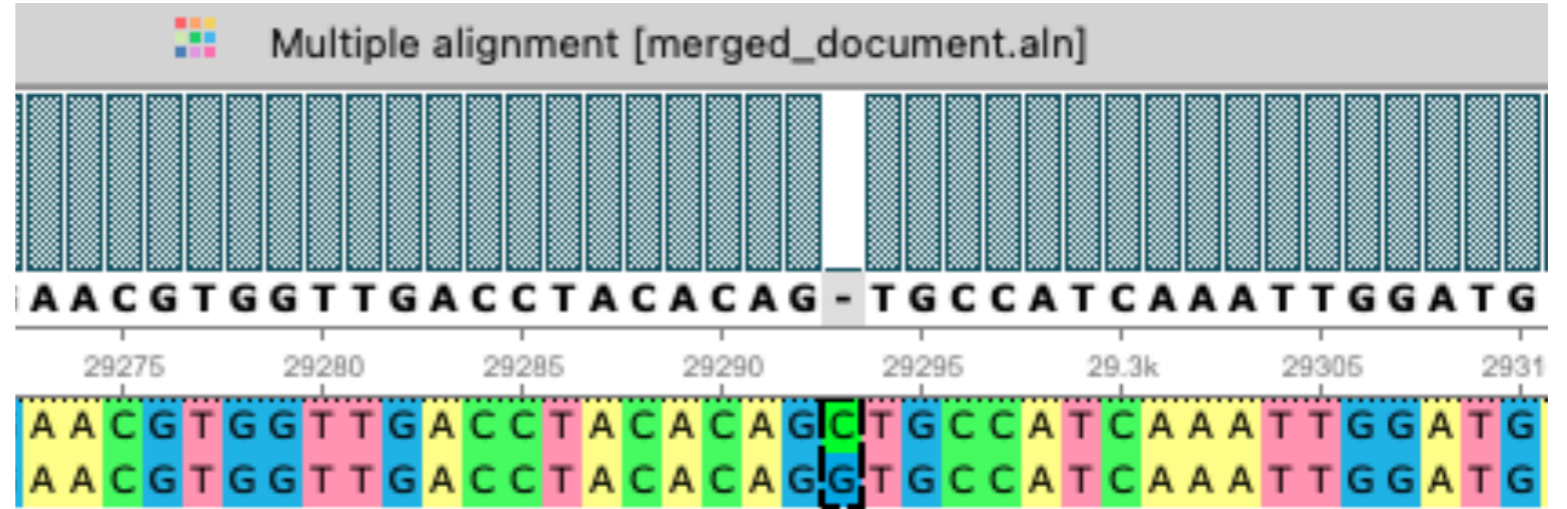
Bu iki varyantı sisteme yükleyip hizalama yapacağız. "**Dizileri Hizalamaya Birleştirin**". Fakat bu halde hizalı değiller.

Alınan iki dizi örneği kayık olduğu için bu iki varyantı kıyaslamamız mümkün olacak. İkisini seçip sağ tıkladığımızda farklı algoritmalar var. Güncel hizalama yapacağımız için **ClustalW** kullanacağım. (Biraz uzun sürebiliyor)

**Örneğin 29.293. yerde Tek Nükleotid Polimorfizmi (SNP) örnek olarak gösterilebilir. Sonlarda iyise varyant farkı iyice artıyor.**

Örnek Makale: <https://pubmed.ncbi.nlm.nih.gov/37948480/>

Nükleotit: [https://www.ncbi.nlm.nih.gov/nucore/NC\\_045512.1?report=fasta](https://www.ncbi.nlm.nih.gov/nucore/NC_045512.1?report=fasta)



## DotPlot (Nokta Grafiği)

Nokta grafiği oluşturarak yine iki dizi arasındaki benzerlik **görselleştirilebiliyor**.

Yine iki **covid** verisini yüklüyoruz ama bu sefer iki **ayrı sıra** modunda içeri aktarıyoruz.

Böylelikle birinci dizinin ikinci diziye minimum **4 baz çifti benzerliğini** görselleştiriyoruz.

Yine scroll ile yakınlaşıyoruz.

Buradaki fark **çizgiler benzer** bölgelerken, diğer kısımlar aynı pozisyondaki farklı diziler.

The image shows a software interface for creating a DotPlot. The window title is "Dotplot kullanarak dizileri karşılaştırın". The "Dotplot parametreleri" section includes:

- X eksen dizisi: NC\_045512.1 Wuhan seafood market pne
- Y eksen dizisi: NC\_045512.2 Severe acute respiratory s
- Yükleme Sırası: (empty)
- Doğrudan tekrarları ara (öntanımlı)
- Tersine çevrilmiş tekrarları ara (öntanımlı)
- Özel algoritma (Otomatik)
- Minimum tekrar uzunluğu: 4bp (1k)
- Kimliği tekrar eder: 100% (100)

Buttons at the bottom: Yardım, İptal Et, TAMAM.

The background shows a dot plot comparing two sequences. The x-axis is labeled "NC\_045512.1 Wuhan seafood market pneumonia virus isolate Wuhan-Hu-1, complete genome (minimum uzunluk 4, kimlik 100%)". The y-axis is labeled "NC\_045512.2 Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, c". The plot shows a dense grid of dots, with a zoomed-in view of a specific region.

Yine geçtiğimiz haftalarda bir makalede protein yapısının görselleştirilmesi geçmişti. Yazılım protein data bank formatını desteklediği için seçtiğimiz protein yapısını indirip inceleyebiliyoruz. <https://www.rcsb.org/structure/11BA>

Download Files

**Protein1** ve **Protein2** adlı dosyalarda birinde normal birinde **mutasyona** uğramış protein yapısı mevcut. Bunları 3 boyutlu görselleştirerek aralarındaki farkı gözlemek mümkün.

Görünüm Bağlantılar

Menüdeki görünüm menüsünden ayrıca görüntülemeyle ilgili farklı modlar seçilebiliyor. **Moleküler** yapılarıyla alakalı.

Görselin altındaki kamera işaretine basarak da görüntüyü çıktı olarak alabiliyoruz.



**Genom** dizisi üzerinde **seçim** yaptıkça bu dizi üzerindeki ilgili bölgelerin protein yapısının hangi bölgesinde **etkili** olduğunu görebiliyoruz.

**Birden fazla** aminoasit zinciri şeklindeki yapıyı oluşturuyor bu genom dizisinde.

**Evrimsel ağaç** yani filogenetik inceleme yapmak için bu seferde farklı organizmalardan elde edilen bir genin yapısını inceleyebiliriz.

.msf formatındaki ncbi'dan indirdiğimiz dosyayı yüklüyoruz.

Diziler farklı alanlardan alındığı için Öncelikle yine **hizalama** yapmamız gerekiyor.

Yaptığımız bu filogenetik ağaçlı bu farklı dizilerin evrimsel olarak birbirlerine yakınlıklarını gösteren bir şema elde ediyoruz

Örneğin burada **insan PA2M** adlı genin farklı organizmalardaki evrimsel yakınlıklarını görüyoruz. İnsan human olarak gösterilmiş. Mesela insan ile sürüngenler arasında yakınlık çıkmış bu grafikte. **Croat** (sürüngenlerden elde edilmiş).

**Birbirine** en yakın **sığır** ve **domuz** olarak sonuçlanıyor. (En alt ve en üst)

