**Proje 3**

**FASTA Algoritması :**

Query Seq: WATSONCRICK

Target Seq: BALLCRICKET

**Adım 1**: Query Seq’i Indexlerine ayır.



**Adım 2:** Alfabeye göre sırala ve mesela c harfinden 2 tane var onları tek bir yerde yaz.

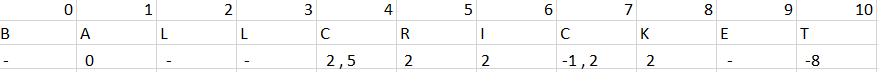


Not: C harfi hem 6. hem de 9. indexte olduğu için tek bir yerde yazdık.

**Adım 3:** Target tablosu yap.

Target tablosu yaparken eğer eşleşme varsa;

Değer = query – target



Örnek:

A değeri için target tablosunda 1. Indexte query tablosunda da 1. Indexte bu yüzden 1 – 1 = 0 değeri geldi

4. indexteki C değeri için query tablosuna bakıyoruyz 6 ve 9 . indexleri gösteriyor 6-4=2 ve 9-4=5 değerleri geliyor

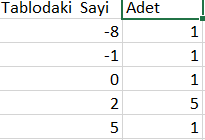
R değeri için query tablosunda 7 target tablosunda 5. Indexte 7-5=2 değeri geliyor.

Diğerleri de aynı bu şekilde hesaplanarak yazılıyor.

Ve query tablosunda olmayan B ve L gibi harfler için “- ” yapıp geçiyoruz.

**Adım 4:** Offset tablosu oluştur

Bu tabloyu oluştururken target tablosunda hangi değerden kaç tane olduğuna bakıyoruz.

Maximum tekrar eden sayı 2dir 5 kere tekrar etmiş. Pozitif olduğu için sağa doğru kaydırıcaz negative olsaydı sola doğru kaydırıcaktık.

**Adım 5:** Kaydırmaları yaparak final tablosunu oluştur.



Sonuç:



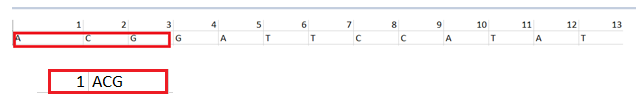
**BLAST Algoritması:**

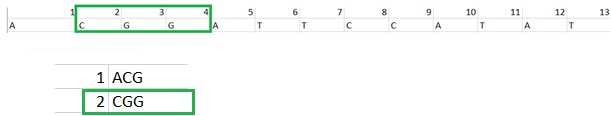
**Query Sequence:** GACAGC

**Database Sequence**: ACGGATTCCATAT

**Word Lengths = 3**

İlk önce database sequence’ı k-mers lara göre index tablosu hazırlayacağız. (k burda kelime uzunluğunu temsil etmektedir). Bunu yaparken database seq’ta üçer üçer ilerleyip olan kelimeleri tabloya ekleyeceğiz.





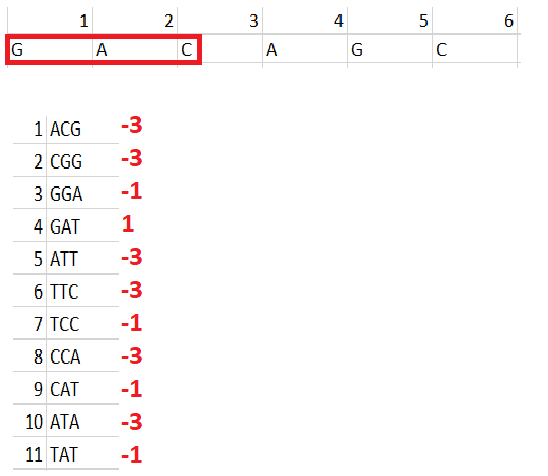
Bu şekilde ekleyerek devam ediyoruz ve en sonunda aşağıdaki tabloyu elde ediyoruz.



Şimdi HSSP(High Scoring Segments Pairs) kısmına geçiyoruz. Burada öncelikle bir score tablomuz olucak.

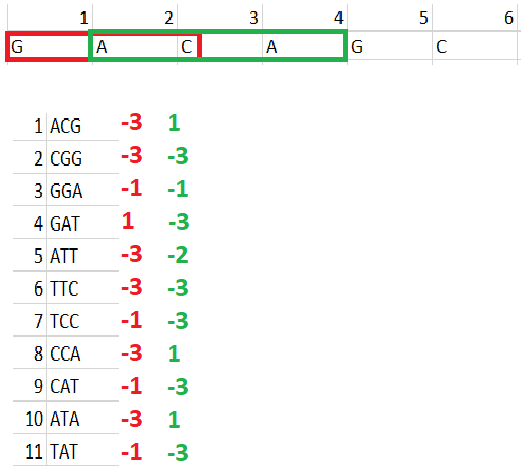
Bizim bu örnekte kullanacağımız skor tablomuz bu şekilde olucak.

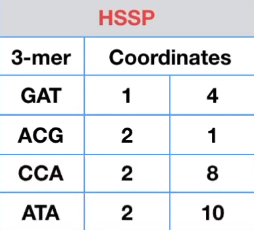
Query Seq’te üçer üçer sırayla ilerleyerek database sequence için daha önce elde etmiş olduğumuz tablo üzerinden ilerleyerek eşleşme durumlarını control edeceğiz ve HSSP threshold’umuzu geçen değerlerimizi HSSP tablomuza ekleyeceğiz. Örneğin;

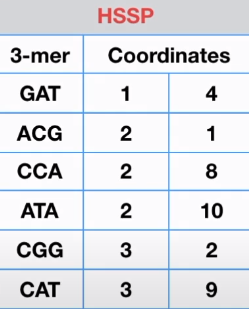
 1 olan threshold değermizi sadece 4. Pozisyondaki GAT değeri geçmiş bunu HSSP tablomuza ekliyoruz.



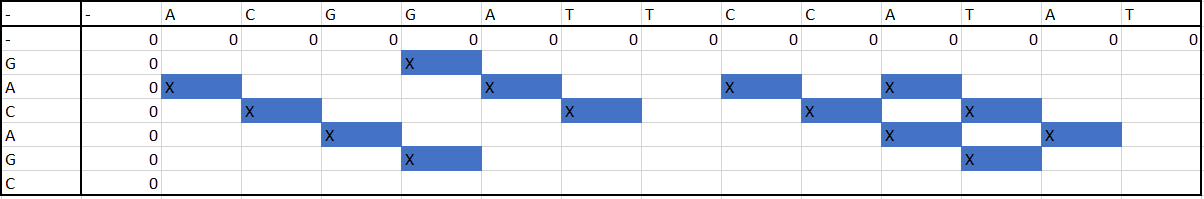
İkinci üçlü için;



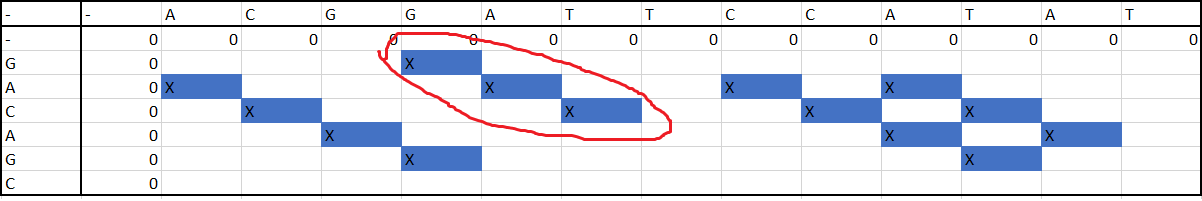
bu şekilde devam ettirdiğimizde en son HSSP tablomuz şu şekilde oluyor;

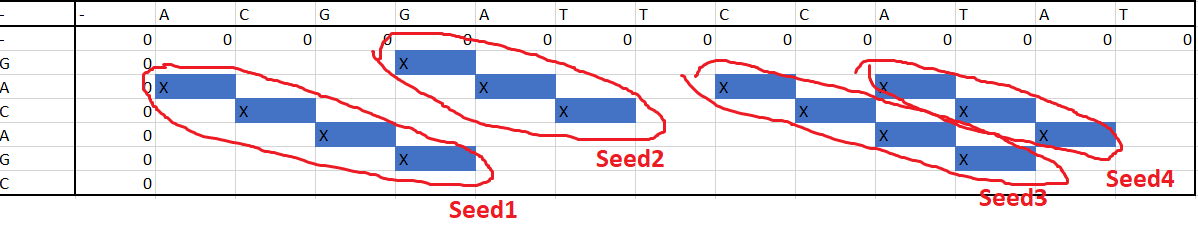


Şimdi bunu daha geniş bir tablodan görücek olursak



Yukarıda X işaretli mavi kutucuklar HSSP tablosundaki değerlerimizin tablo üzerinde işaretlenmiş halidir. Bunu yaparken coordinate kısmında GAT ,1, 4 ü ele alallım buradaki 1 bize query seq’teki ilk 3 harfi temsil ediyordu yani GAC.GAC ve GAT’nin kesiştiği yerleri çarpı haline getirdik.4 tane çarpı çapraz çapraza olan kısımlarda 2 tane kesişimin sonuna denk geldikleri için o şekilde oldu.





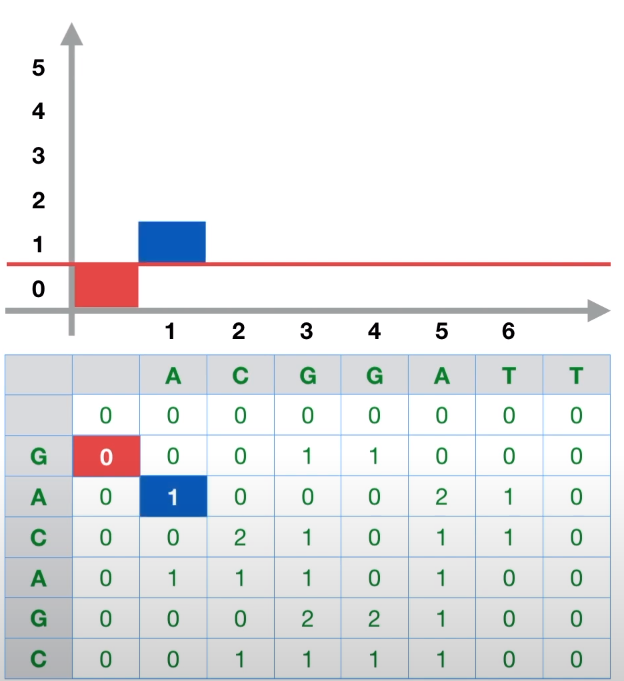
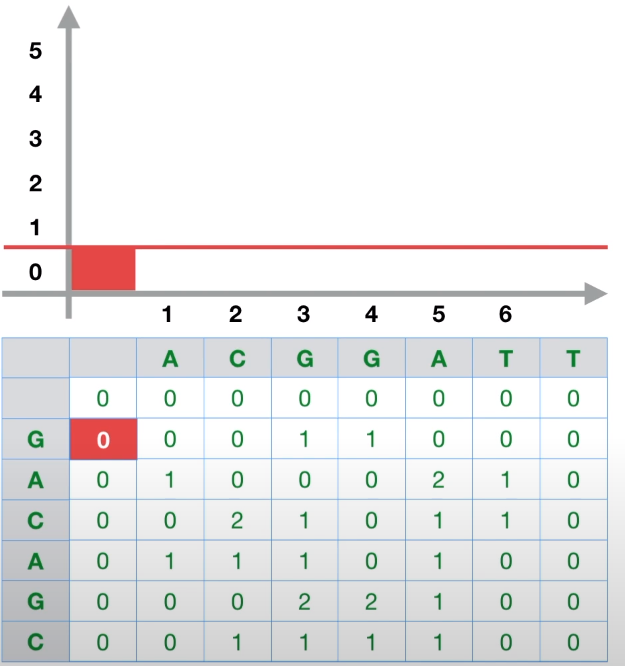
Bir sonraki aşama için yeni bir skor şemasına ihtiyacımız var.



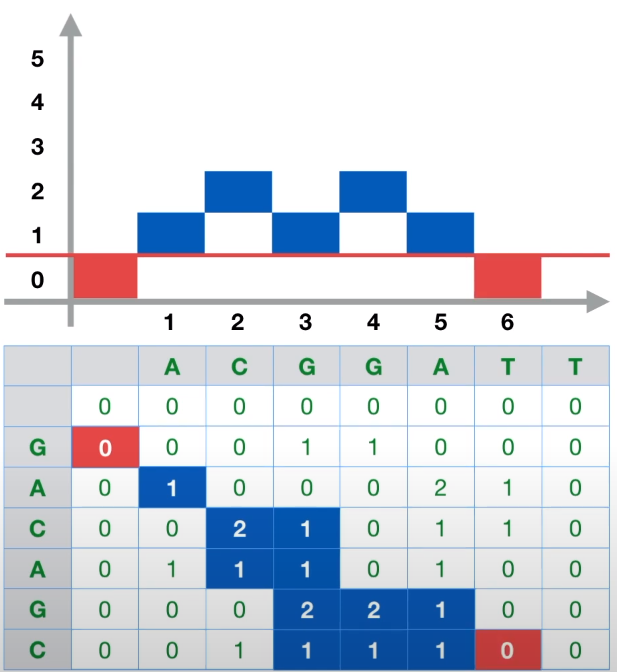
Şimdi Smith-Waterman algoritmasına göre tablomuzu dolduruyoruz



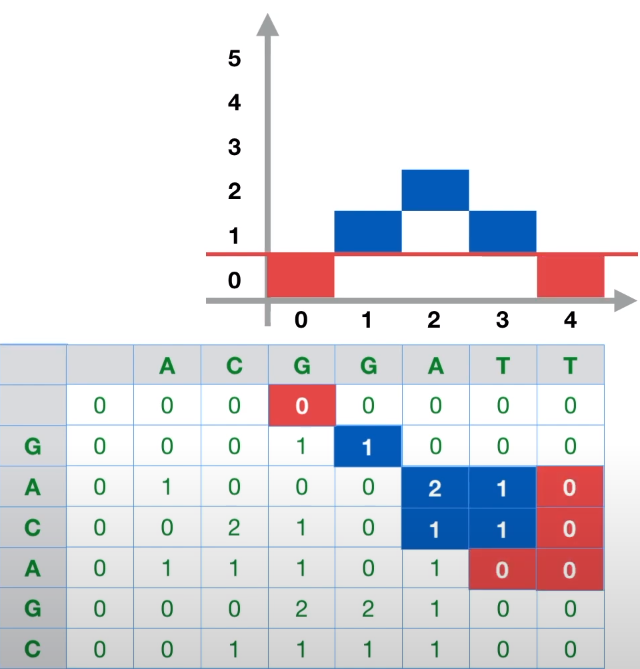
Şimdi bakıcağımız Seed’lerdeki yerlere göre bir score tablosu çizeceğiz diagonal olarak Seed’lerimize bakıyoruz; (Yana doğru olan kırmızı çizgi skor tablomuzdaki score thresholdunu gösteriyor yani 1 ve bu değerin altındakiler kırmızı üstündekiler mavi olmuş oluyor.)



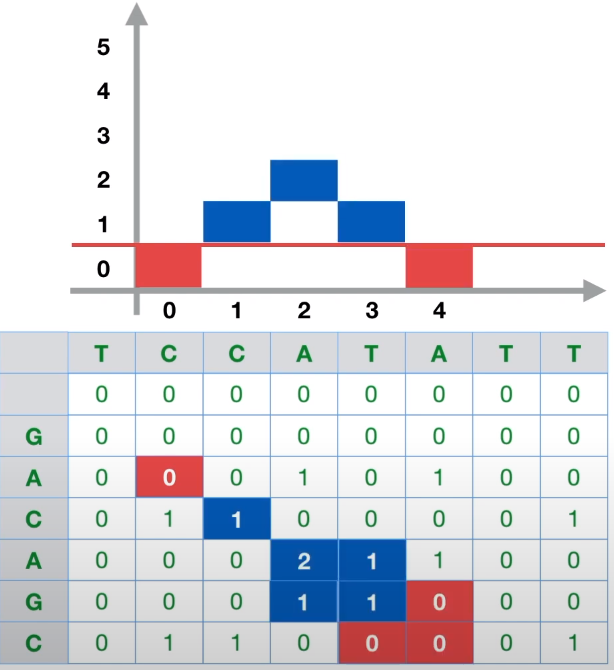
Bu şekilde ilerlettiğimizde Seed 1 için;



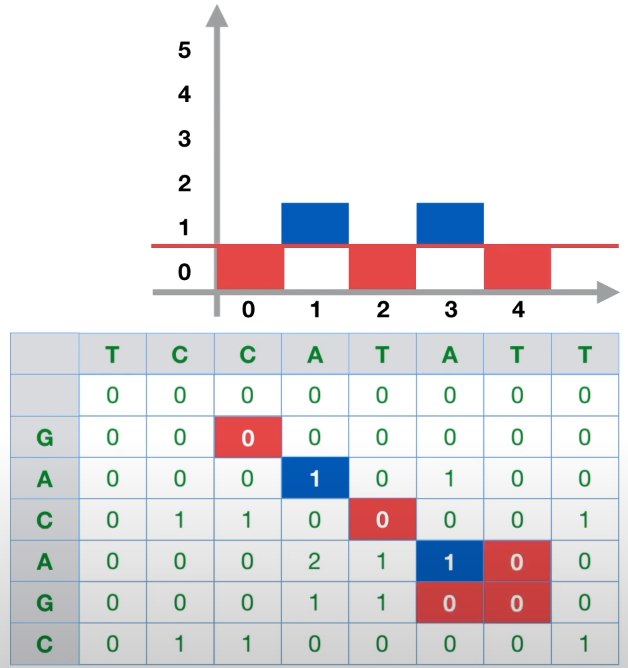
Seed 2 için ;



Seed3 için;

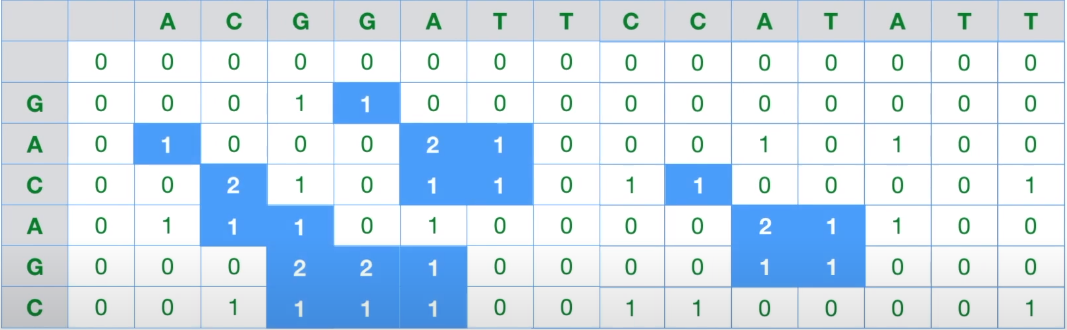


Seed4 için;



Seed 4 için herhangi bir extension bulunmadığını söyleyebiliriz.

Bizim sonuç grafiğimiz aşağıdaki şekilde olmuş oldu

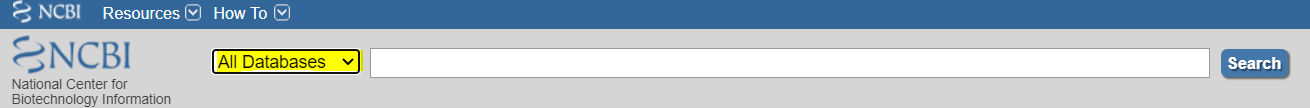


Ve bu değerlere göre yapılabilcek local hizalamalar ;

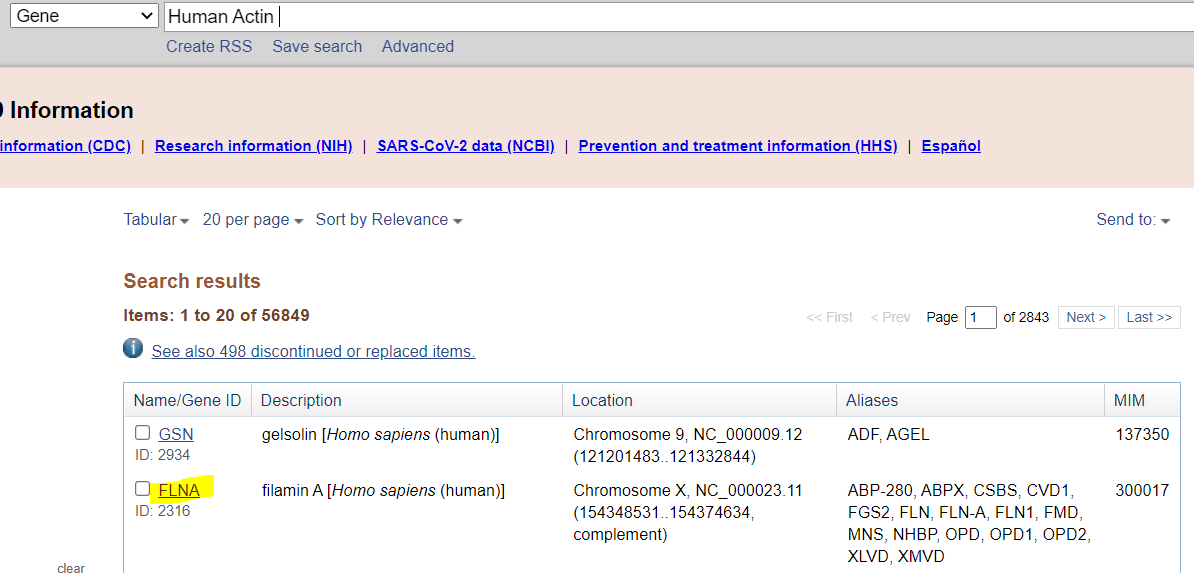


Bu şekildedir.

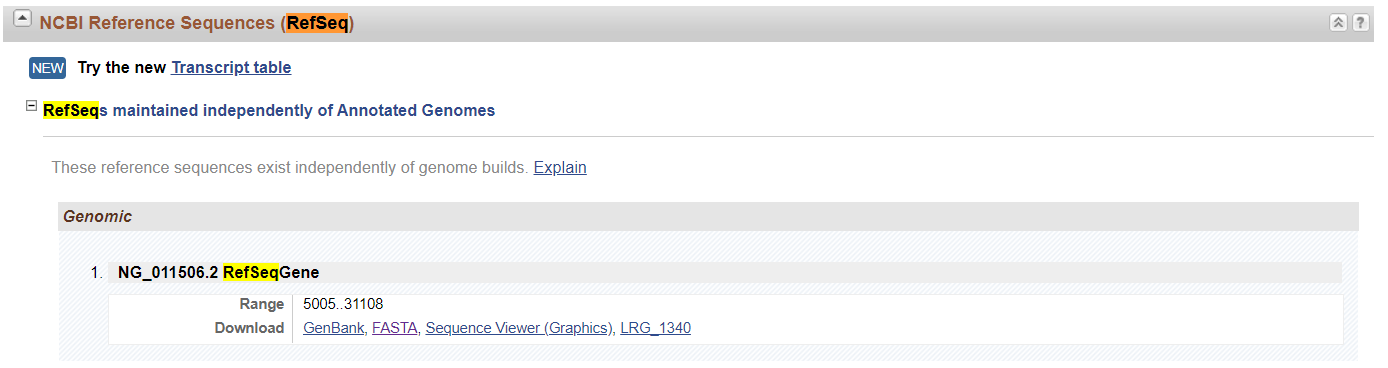
BLAST Uygulama:



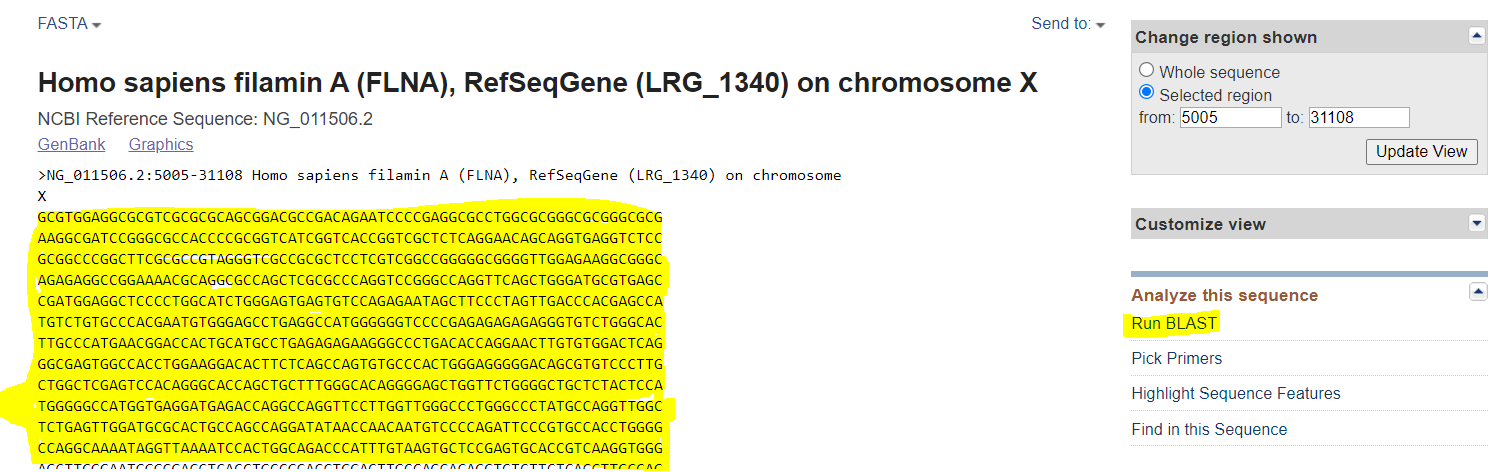
All databases kısmınden Gene seçeneğini seçiyoruz sonra orada Human Actin’I aratıyoruz.



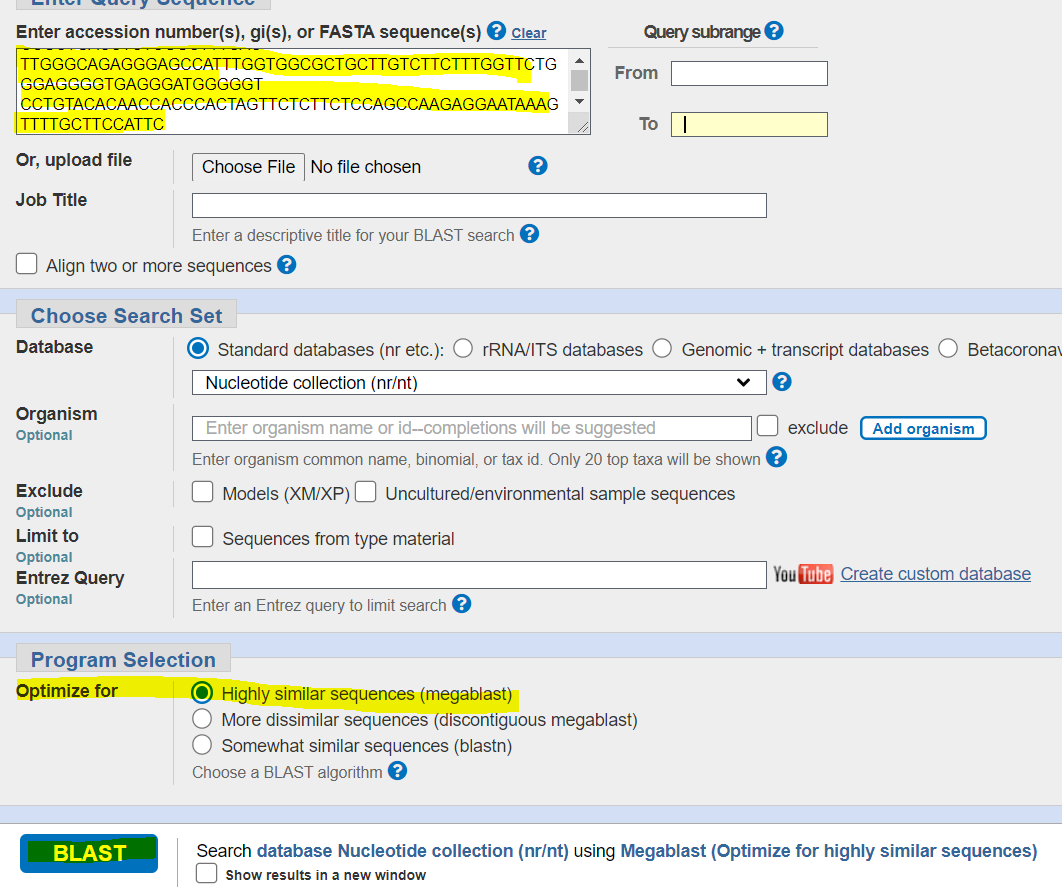
FLNA kısmına tıklıyoruz.



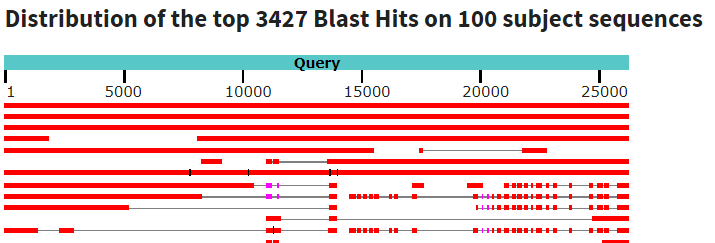
Burada RefSeq kısmında BLAST kullanacağımız için FASTA seçeneğini seçiyoruz.



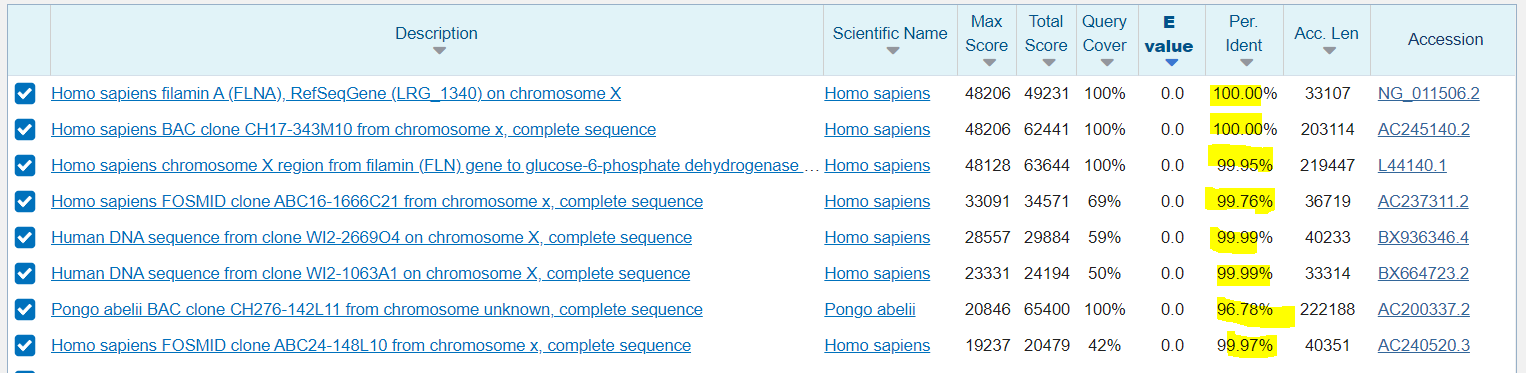
Gen kısmının tamamını kopyalayıp sağ taraftaki menüden Run BLAST seçeneğine tıklıyoruz.



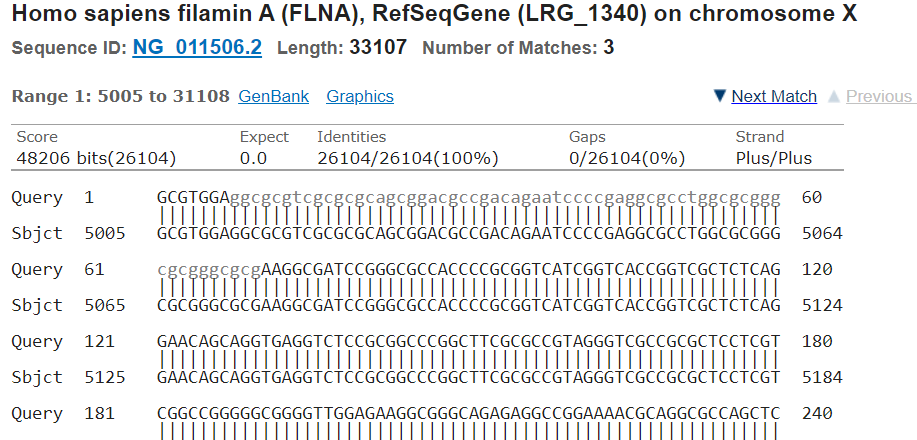
Yukarıdaki yere yapıştırıp, Optimize kısmında highly similar sequences kısmını seçiyoruz. Bu kısım bize yüksek benzirlik oranına sahip olduklarını getiricek ve BLAST’a tıklıyoruz.

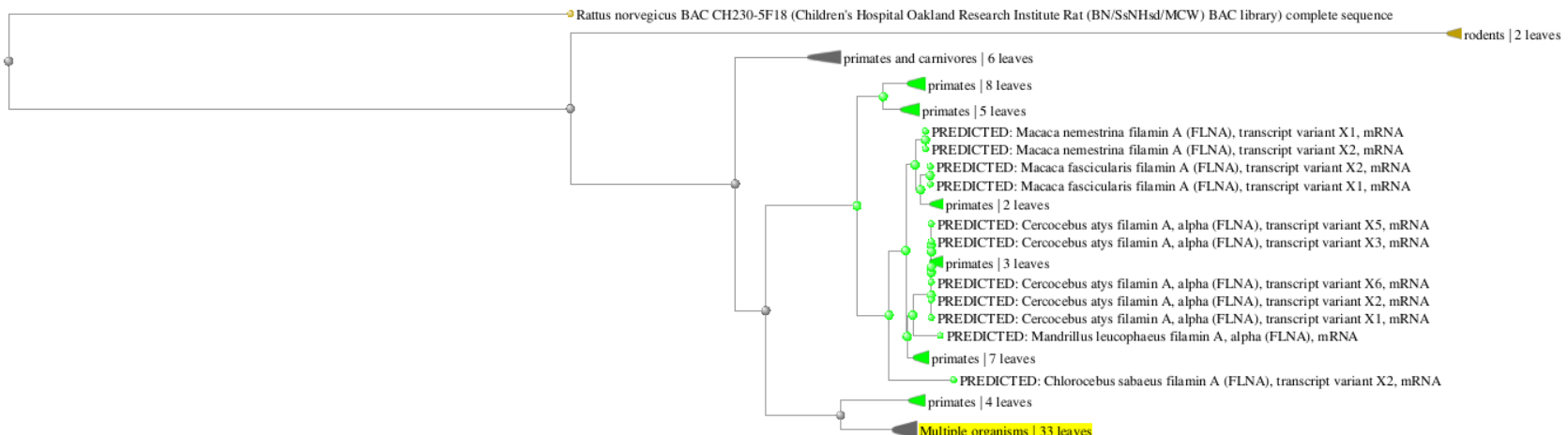


3427 adet yüksek uyumlu DNA dizisi bulundu.



Yüzdeler uyumluluğu gösterir ve %99 yüksek uyumlu olan genlerin yine Homo sapiens(insan) genine ait olduğunu görebiliriz.

Alignment kısmında yapılan hizalamaları görebiliriz. Fakat burası çok uzun olucak çünkü binlerce satır hizalama yapıldı.



Description kısmından Distance tree of result seçeneğini seçersek bu şekilde ağaç yapısını görebiliriz.