

**T.C. Fırat Üniversitesi**

**Bilgisayar Mühendisliği Bölümü**

**BMÜ450 - Biyoinformatik Dersi**

**BLAST İle Dizi Hizalama Ödevi Raporu**

**Ders Sorumlusu:** Prof. Dr. Mehmet KAYA

**Öğrenci İsim:** Mert İNCİDELEN

**Öğrenci No.:** 170260101

**ÖDEV TANIMI**

Ödevde, BLAST Algoritmasının araştırılması ve çalışma prensibinin anlaşılması; BLAST kullanan bir araç üzerinde örnek dizilerin hizalanıp, nasıl yapıldığının gösterilmesi beklenmektedir.

**1. BLAST ALGORİTMASI**

BLAST (Basic Local Alignment Search Tool), biyoinformatik çalışmalarında yaygın kullanılan veri tabanı arama ve hizalama aracıdır. Dizi veri tabanlarını aramak için birçok algoritma mevcuttur, ancak BLAST Algoritmaları hızı itibariyle diğerlerine kıyasla en popüler olanlarıdır. BLAST Algoritmasının temel stratejisi amino asitlerden oluşan sorgu (BLASTP) üzerinden örneklenmiştir.

“… **M E R T I N C I D E L E N** …” sorgusu, boyutu 3 olan (BLASTP için varsayılan boyut 3 amino asittir) kelimelere ayrılır. Böylece MER, ERT, RTI, TIN, INC, NCI, CID, IDE, DEL, ELE, LEN kelimeleri elde edilir. Bir sorgu için elde edilen kelime n-2 adet olmaktadır. Proteinler 20 farklı amino asit içerdiğinden, bir kelimenin her 203 tripeptitte bir kez (8000 tripeptit, 24000 amino asit içinde) rastgele oluşması beklenir, bu da herhangi bir proteinden daha uzundur ve dolayısıyla idealdir.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Sorgu:** | … | M | E | R | T | I | N | C | I | D | E | L | E | N | … |
| **Kelimeler:** | … | M | E | R |  |  |  |  |  |  |  |  |  |  |  |
|  |  |  | E | R | T |  |  |  |  |  |  |  |  |  |  |
|  |  |  |  | R | T | I |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  | T | I | N |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  | I | N | C |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  | N | C | I |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  | C | I | D |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  | I | D | E |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  | D | E | L |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  | E | L | E |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  | L | E | N | … |

Elde edilen boyutu 3 olan bu kelimelerin komşu kelimeleri bulunur ve bir eşik değeri belirlenir. Bir skor tablosu kullanarak eşik değerinin üzerinde skora sahip komşu kelimeler de sorgulamaya dahil edilir. Sonraki aşamada “MER” kelimesi için komşu kelimeler, eşik değeri 11 belirlenerek BLOSUM 62 tablosuna göre bulunmuştur. Eşik değeri düşürülerek arama kapsamı genişletilebilir.

|  |  |
| --- | --- |
| **KELİME** | **SKOR (BLOSUM62)** |
| **MER** | **15** |
| MEK | 12 |
| LER | 12 |
| MDR | 12 |
| MQR | 12 |
| MEQ | 11 |
| IER | 11 |
| VER | 11 |
| MKR | 11 |
| ~~MEN~~ | ~~10~~ |
| ~~QER~~ | ~~10~~ |
| **.**  **.**  **.** | |
| **Eşik değeri = 11** | |

Daha sonra hedef dizilerde aranan kelimeler bulunarak, isabetin olduğu bölgede her iki yönde eşleştirmeler genişletilir.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | **…** | L | R | **M** | **E** | **R** | T | I | N | C | I | D | E | L | E | N | **…** |  |
| **🡸** |  |  |  | I | I | I |  |  | I |  |  | I |  | I |  |  |  | 🡺 |
|  | **…** | C | E | **M** | **E** | **R** | Q | R | N | L | Q | D | R | L | T | W | **…** |  |

Genişletilen bölgenin HSP (high-scoring segment pair) puanı, skor tablosu üzerinden hesaplanır ve en iyi eşleşmeler bu şekilde saptanabilir. Yukarıdaki bölge için HSP puanı BLOSUM62 matrisine göre hesaplanmıştır.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **EŞ.** | **SKOR** |  | **EŞ.** | **SKOR** |  | **EŞ.** | **SKOR** |  | **EŞ.** | **SKOR** |  | **EŞ.** | **SKOR** |
| L-C | -1 | E-E | 5 | I-R | -3 | I-Q | -3 | L-L | 4 |
| R-E | 0 | R-R | 5 | N-N | 6 | D-D | 6 | E-T | -1 |
| M-M | 5 | T-Q | -1 | C-L | -1 | E-R | 0 | N-W | -4 |

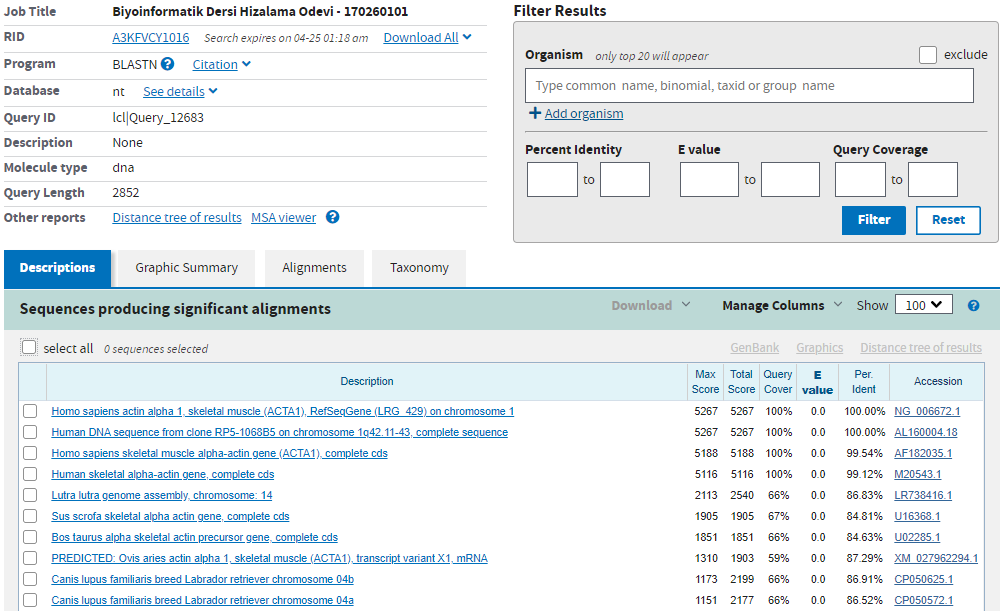
Pozitif skorlu eşleşmelerin değerlerini toplarsak (5+5+5+6+6+4 = 31), HSP puanı 31 olarak bulunur. Negatif skorlu eşleşmeler de dahil edilerek toplam puan 17 olarak bulunabilir.

**2. BLAST İle Dizi Hizalama Uygulamaları**

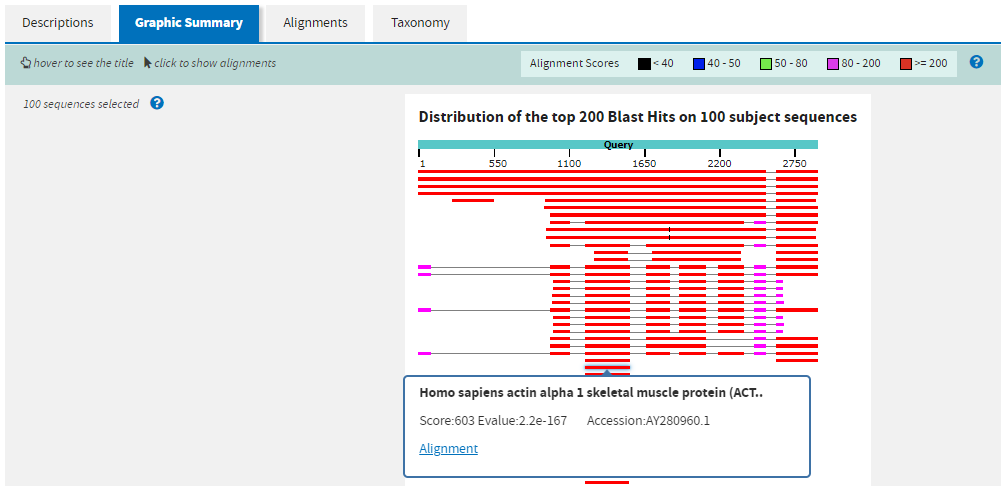
**2.1. BLASTn İle Uygulama**

Dizi hizalama işlemi için **blast.ncbi.nlm.nih.gov/Blast.cgi** adresinde bulunan araçtan faydalanılmıştır. Eşleştirme için ikinci örnek uygulamada dizi olarak **ACTA1** kimliğine sahip (ID No. 58) 2852 nükleotit uzunluğunda bir dizi kullanılmıştır. BLASTN (nükleotit-nükleotit karşılaştırma ve eşleştirme) ile ACTA1 dizisi standart veri tabanı üzerinde benzerlik derecesi yüksek olacak şekilde sorgulanmış ve aşağıdaki veriler elde edilmiştir.

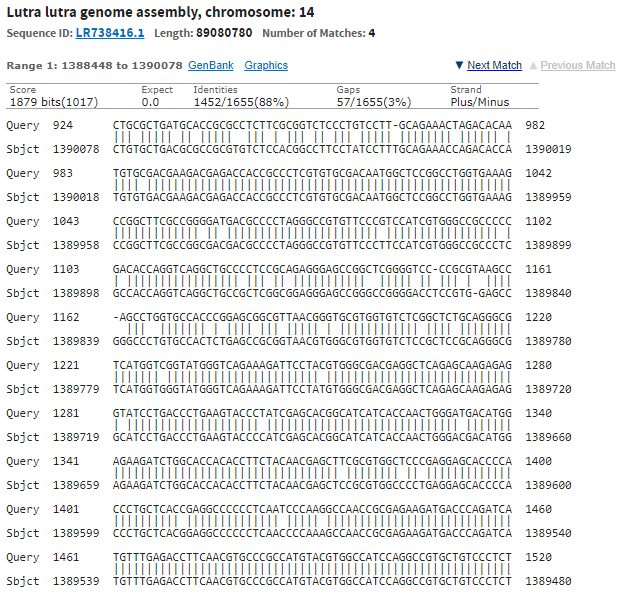
**Sorgulama sonrasında karşılaştırılan dizilerden benzerlik oranı yüksek olandan düşüğe doğru sonuç ekranında aşağıda görüldüğü şekilde sıralanmıştır.**



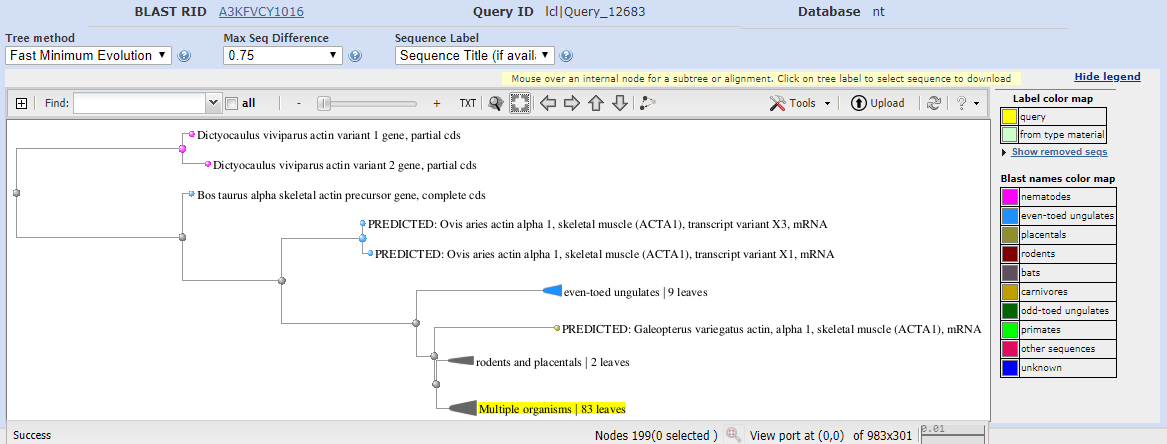
**Bu diziler bir grafik üzerinde hizalanma konumları ile kabaca gösterilmiştir. Kırmızı renk ile gösterilenler eşleşme skoru 200 ve üzerinde olan hizalamalardır.**



**Eşleşen dizilerden bir tanesi için eşleşen kısımların bir bölümünün sonuç dökümü ise aşağıdaki gibi gösterilmiştir.**



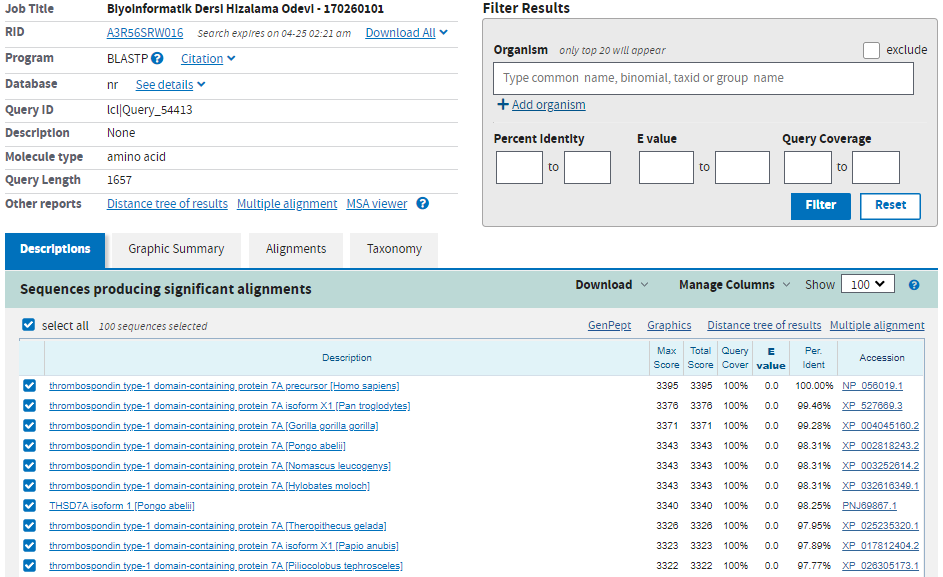
**Benzerlikleri esas alınarak sorgu sonucunda elde edilen verilere göre dizilerin filogenetik haritası ise aşağıdaki gibidir.**

****

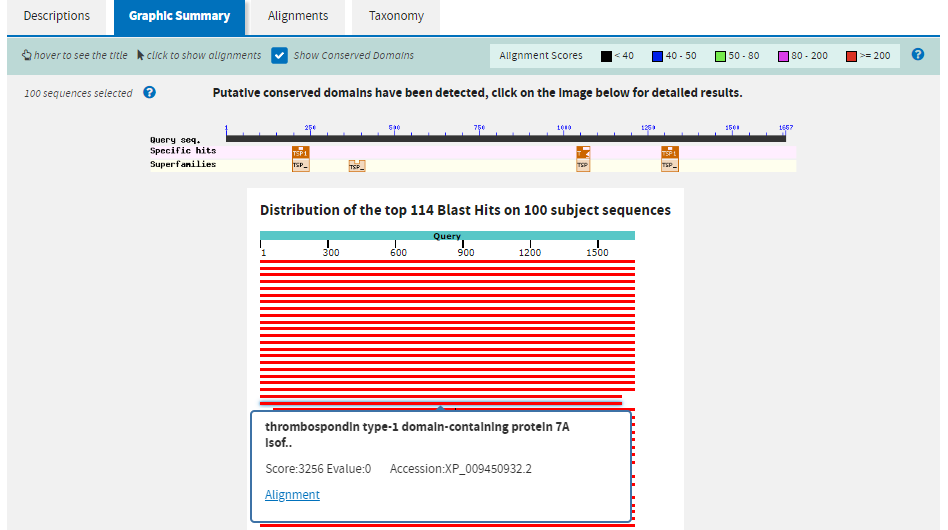
**2.2. BLASTp İle Uygulama**

Eşleştirme için ikinci örnek uygulamada dizi olarak **Thrombospondin Type-1** kimliğine sahip 1657 aminoasit içeren bir dizi kullanılmıştır. BLASTp (aminoasid-aminoasid karşılaştırma ve eşleştirme) ile Thrombospondin Type-1 dizisi veri tabanı üzerinde sorgulanmış ve aşağıdaki veriler elde edilmiştir.

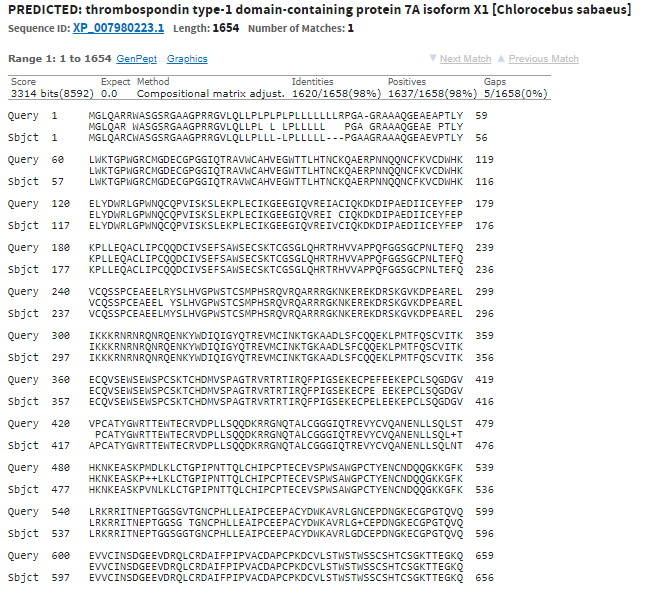
**Sorgulama sonrasında karşılaştırılan dizilerden benzerlik oranı yüksek olandan düşüğe doğru sonuç ekranında aşağıda görüldüğü şekilde sıralanmıştır.**



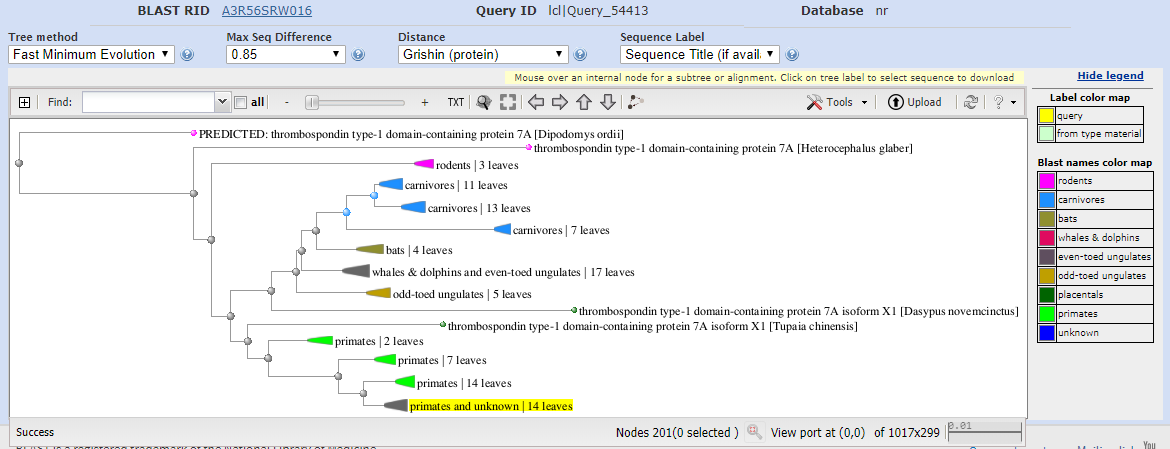
**Bu diziler için hizalama grafiği kabaca aşağıdaki gibi elde edilmiştir. Kırmızı renkli eşleşmeler, eşleşme puanı 200 ve üzerinde olanları ifade etmektedir.**



**Eşleşen dizilerden bir tanesi için eşleşen kısımların bir bölümünün sonuç dökümü ise aşağıdaki gibi elde edilmiştir.**

****

**Benzerliklerine göre sorgu sonucunda elde edilen veriler için dizilerin benzerlik haritası ise aşağıdaki gibidir.**

****

**3. YARARLANILAN KAYNAKLAR**

**[1]**<http://resources.qiagenbioinformatics.com/manuals/clcmainworkbench/800/index.php?manual=How_does_BLAST_work.html>

**[2]**<https://bio.libretexts.org/Bookshelves/Cell_and_Molecular_Biology/Book%3A_Investigations_in_Molecular_Cell_Biology_(O'Connor)/9%3A_Protein_Conservation/9.3%3A_BLAST_algorithms_are_used_to_search_databases>

**[3]**<https://en.wikipedia.org/wiki/BLAST_(biotechnology)>

**[4]**<https://biotechgo.org/tr/?view=article&id=247:lo2&catid=136>

**[5]**<https://www.ccg.unam.mx/~vinuesa/tlem/pdfs/Bioinformatics_explained_BLAST.pdf>

**[6]**<http://www.yasinhoca.com/2016/11/dna-dizilerini-karslastrma-blast.html>

**[7]**<https://www.comp.nus.edu.sg/~ksung/algo_in_bioinfo/slides/Ch5_database.pdf>

**[8]**<https://bioinf.comav.upv.es/courses/biotech3/theory/blast.html>

**[9]**<https://www.youtube.com/watch?v=4AcnZnZRss8>

**[10]**<https://www.youtube.com/watch?v=g0nSH17psDc>

**[11]**<https://www.youtube.com/watch?v=NnY2f5111FU>