Bioinformatics Assignment

Anand.M.P
B120361CS
S8, B.Tech Computer Science

Question 1:

Create an implementation for the Global Sequence Alignment algorithm.

Input:

It takes three inputs; two DNA sequences, and one scoring criteria, which would determine how the matches, mismatches, insertions, and deletions should be scored.

Output:

One of the plausible alignments and the corresponding score of the alignment

Solution:

The code for the above problem is attached with the mail.

Question 2:

Create an implementation for the Exon Chaining Problem.

Input:

A set of exons is provided as the input. Each input is given in the following format,

L R W

Where 'L', represents the left index of the interval, 'R' the right index of the interval, and 'W' the weight of the interval.

Output:

The maximum score that can be obtained by using non-overlapping exon sequences.

Solution:

The code for the above problem is attached with the mail.

Question 3:

Study of Clustal Omega :

Use the Clustal Omega tool to carry out Multiple Sequence Alignment, and generate a phylogram tree of 7 hemoglobin sequences, downloaded from UniPort, of 7 different organisms. Perform two different test cases and analyze the results obtained. Create a step by step documentation of the same.

Answer:

The two different test sets used for performing the activity have been listed below.

Test case 1:

1. Humans - *Homo sapiens*, Hemoglobin subunit beta
2. Rat - *Rattus norvegicus*, Hemoglobin subunit beta-1
3. Mouse - *Mus musculus*, Hemoglobin subunit beta-1
4. Rabbit - *Oryctolagus cuniculus*, Hemoglobin subunit beta-1/2
5. Sperm Whale - *Equus caballus*, Hemoglobin subunit alpha
6. Horse - *Physeter catodon*, Hemoglobin subunit beta-1/2
7. Chimpanzee - *Pan troglodytes*, Hemoglobin subunit alpha

Test case 2:

1. Cat - *Felis catus*, Hemoglobin subunit beta-A/B
2. Olive baboon - *Papio Anubis*, Hemoglobin subunit beta
3. Sheep - *Ovis aries*, Hemoglobin subunit beta
4. Atlantic cod - *Gadus morhua*, Hemoglobin subunit beta
5. Chicken - *Gallus gallus*, Hemoglobin subunit alpha-2
6. Golden hamster - *Mesocricetus auratus*, Hemoglobin subunit beta
7. Australian ghost bat- *Macroderma gigas*, Hemoglobin subunit alpha-1/2

Procedure:

1. Hemoglobin sequence for different organisms were collected from UniProt(www.uniport.org).

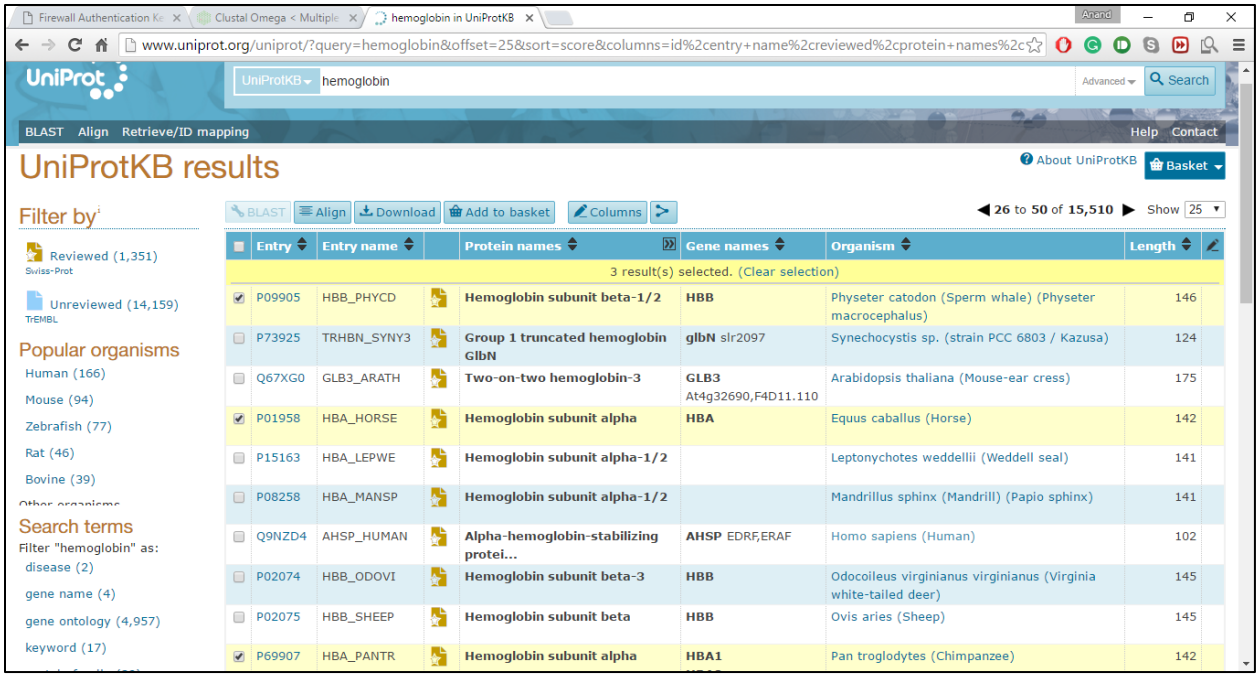


Figure 1 Hemoglobin sequences at UniProt website

2. Select the 7 hemoglobin sequences for the first test case and click download. This will download all the 7 sequences in FASTA format, which will be then used as input for Clustal Omega tool.

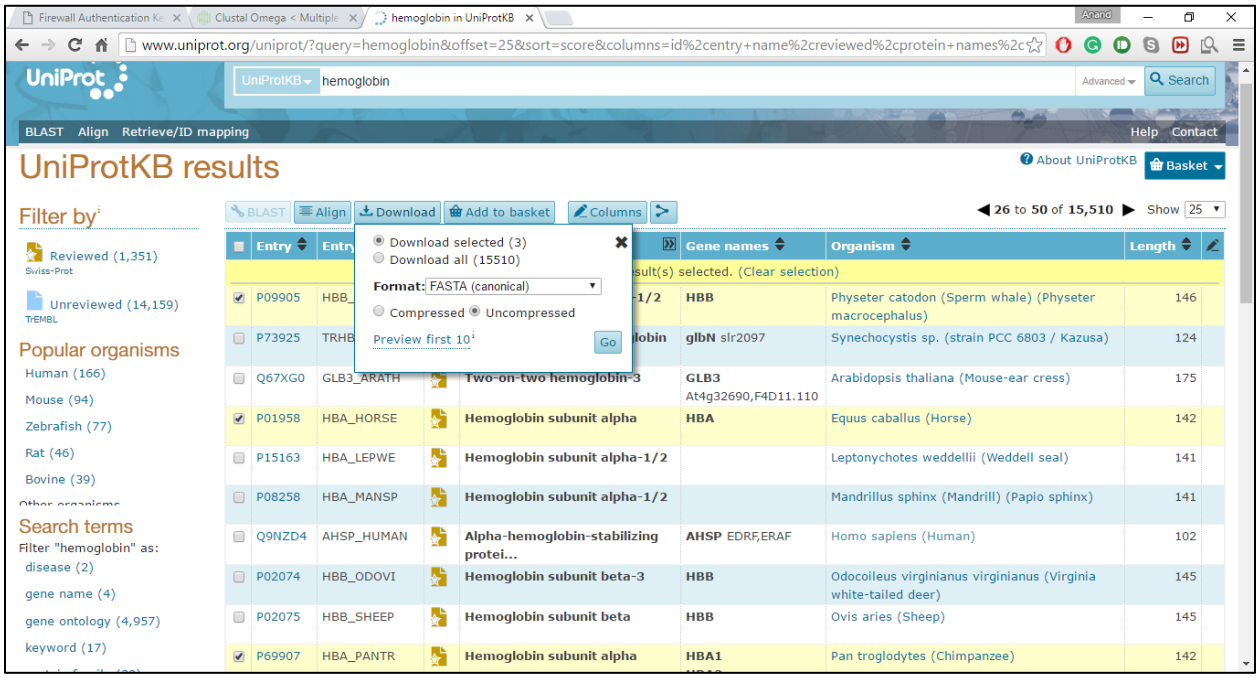


Figure 2 Download Hemoglobin Sequence

3. At the Clustal Omega website (http://www.ebi.ac.uk/Tools/msa/clustalo/), the sequence obtained from UniProt is added as the input. The default settings of the tool were left unchanged. Click the submit button to start processing the input.

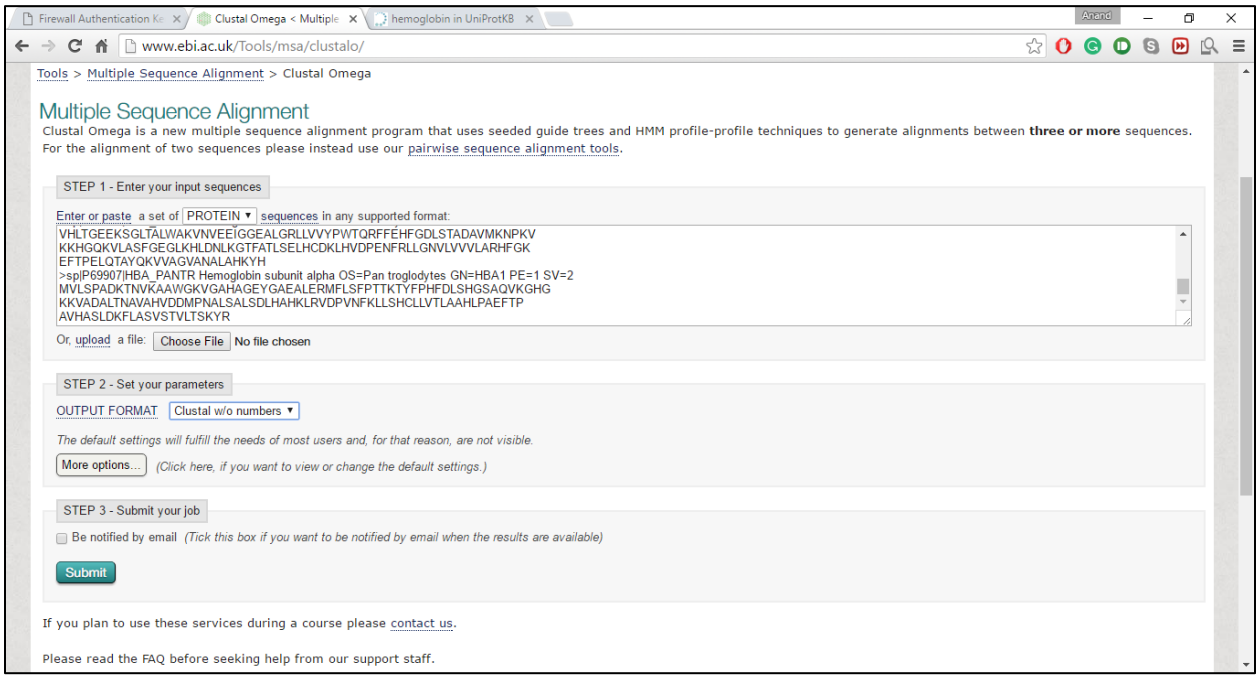
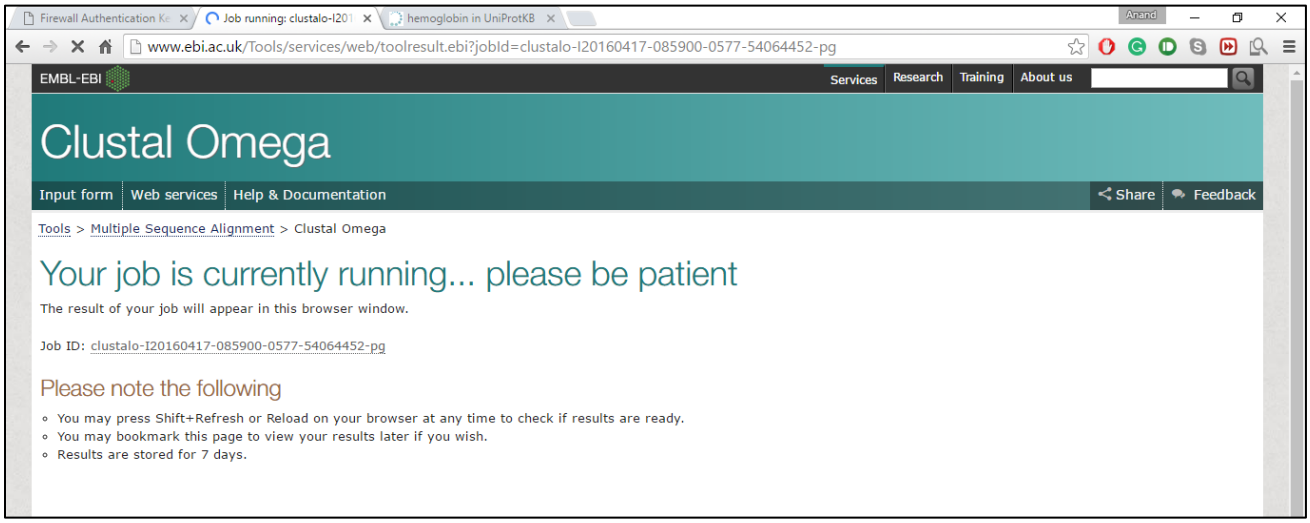


Figure 3 Clustal Omega welcome screen

4. Wait for some time while the clustal omega tool processes the input.



5. The Multiple Sequence Alignment (MSA) for the input sequences is generated as the output along with the phylogenetic tree.

```
CLUSTAL O(1.2.1) multiple sequence alignment


sp|P01958|HBA_HORSE       -MVLSAADKTNVKAAWSKVGGHAGEYGAEALERMFLGFPTTKTYFPHFDLS------HGS
sp|P69907|HBA_PANTR       -MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLS------HGS
sp|P02091|HBB1_RAT        MVHLTDAEKAAVNGLWGKVNP--DDVGGEALGRLLVVYPWTQRYFDSFGDLSSASAIMGN
sp|P02088|HBB1_MOUSE      MVHLTDAEKAAVSCLWGKVNS--DEVGGEALGRLLVVYPWTQRYFDSFGDLSSASAIMGN
sp|P09905|HBB_PHYCD       -VHLTGEEKSGLTALWAKVNV--EEIGGEALGRLLVVYPWTQRFFEHFGDLSTADAVMKN
sp|P68871|HBB_HUMAN       MVHLTPEEKSAVTALWGKVNV--DEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGN
sp|P02057|HBB_RABIT       MVHLSSEEKSAVTALWGKVNV--EEVGGEALGRLLVVYPWTQRFFESFGDLSSANAVMNN
                           : *:  :*: :.  *.**     : *.*** *::: :* *: :*  *          .

sp|P01958|HBA_HORSE       AQVKAHGKKVGDALTLAVGHLDDLPGALSNLSDLHAHKLRVDPVNFKLLSHCLLSTLAVH
sp|P69907|HBA_PANTR       AQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAH
sp|P02091|HBB1_RAT        PKVKAHGKKVINAFNDGLKHLDNLKGTFAHLSELHCDKLHVDPENFRLLGNMIVIVLGHH
sp|P02088|HBB1_MOUSE      AKVKAHGKKVITAFNDGLNHLDSLKGTFASLSELHCDKLHVDPENFRLLGNMIVIVLGHH
sp|P09905|HBB_PHYCD       PKVKKHGQKVLASFGEGLKHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVVVLARH
sp|P68871|HBB_HUMAN       PKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHH
sp|P02057|HBB_RABIT       PKVKAHGKKVLAAFSEGLSHLDNLKGTFAKLSELHCDKLHVDPENFRLLGNVLVIVLSHH
                           :** **:**  ::  .: *:*.:  ::: **:**..**:*** **:**.. :: .*. *

sp|P01958|HBA_HORSE       LPNDFTPAVHASLDKFLSSVSTVLTSKYR
sp|P69907|HBA_PANTR       LPAEFTPAVHASLDKFLASVSTVLTSKYR
sp|P02091|HBB1_RAT        LGKEFTPCAQAAFQKVVAGVASALAHKYH
sp|P02088|HBB1_MOUSE      LGKDFTPAAQAAFQKVVAGVATALAHKYH
sp|P09905|HBB_PHYCD       FGKEFTPELQTAYQKVVAGVANALAHKYH
sp|P68871|HBB_HUMAN       FGKEFTPPVQAAYQKVVAGVANALAHKYH
sp|P02057|HBB_RABIT       FGKEFTPQVQAAYQKVVAGVANALAHKYH
                          :  :***  ::: :*.::.*:..*: **:
```

Figure 4 Multiple Sequence Alignment for test case 1

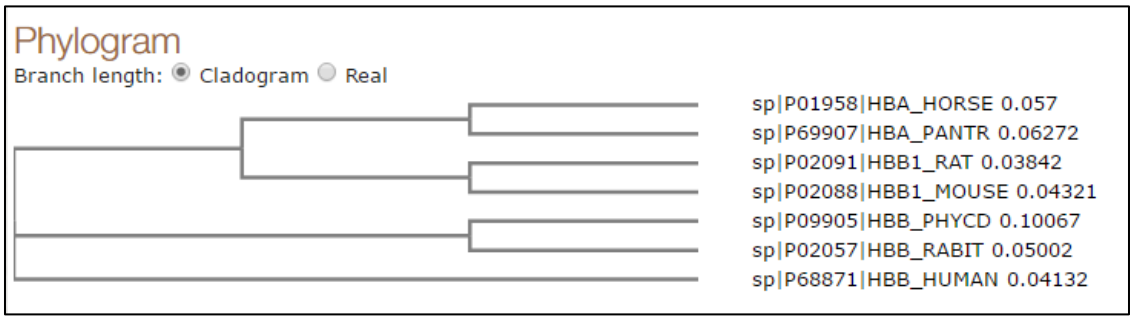
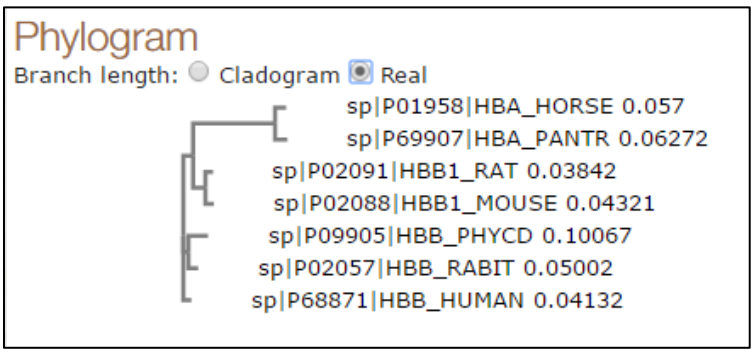


Figure 5 Phylogram for test case 1



6. Similar steps were performed for test case 2 and the following results were obtained.

Figure 6 Multiple Sequence Alignment for test case 2

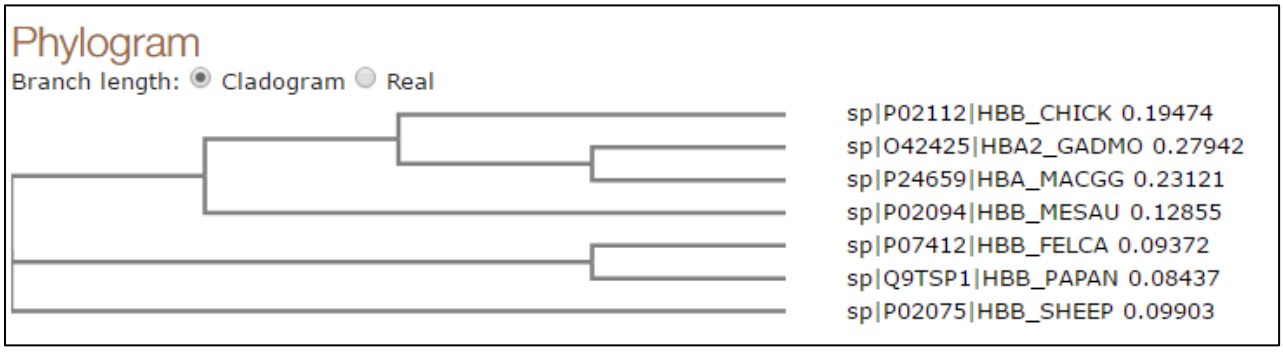
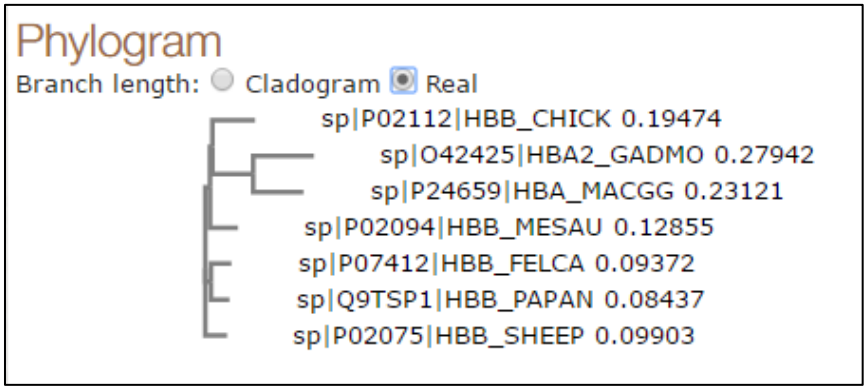


Figure 7 Phylogram for test case 2



7. Some of the Input Parameters used by the Clustal Omega for generating the output are as follows:

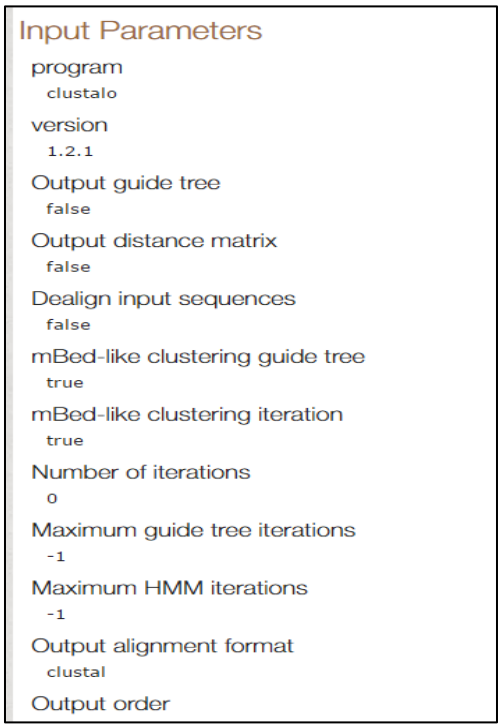


Figure 8 Input Parameters used by Clustal Omega

Observations:

Test Case 1:

It can be observed from the phylogram obtained from the test set 1 that humans are least similar to the rest of the organisms in that particular set. Whereas, rat and mouse are the most similar as it would have expected. This clearly indicates that they are closely related to each other.

Test Case 2:

It can be observed from the phylogram obtained from the test set 2 that sheep is least similar to the rest of the organisms in that particular set. Whereas, cat and olive baboon were found have great amount of similarities, indicating that they are closely related.

Question 4:

Collect the chromosome 4 sequence of humans from NCBI website and use this sequence to perform a Genscan and analyze the results obtained.

Procedure:

1. The chromosome 4 sequence was collected from the NCBI website. Taking into consideration the enormous size of the sequence, a small portion of it was later used as the input for Genscan.

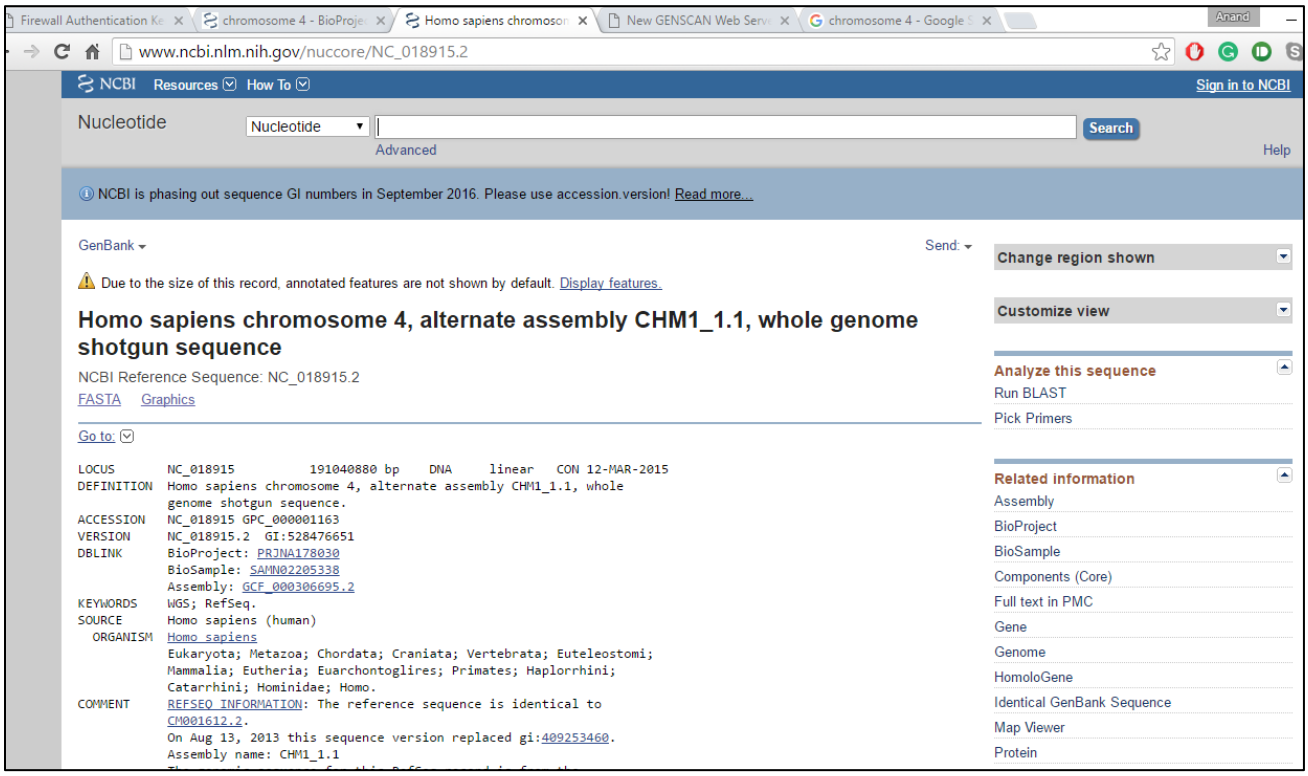


Figure 9 NCBI website

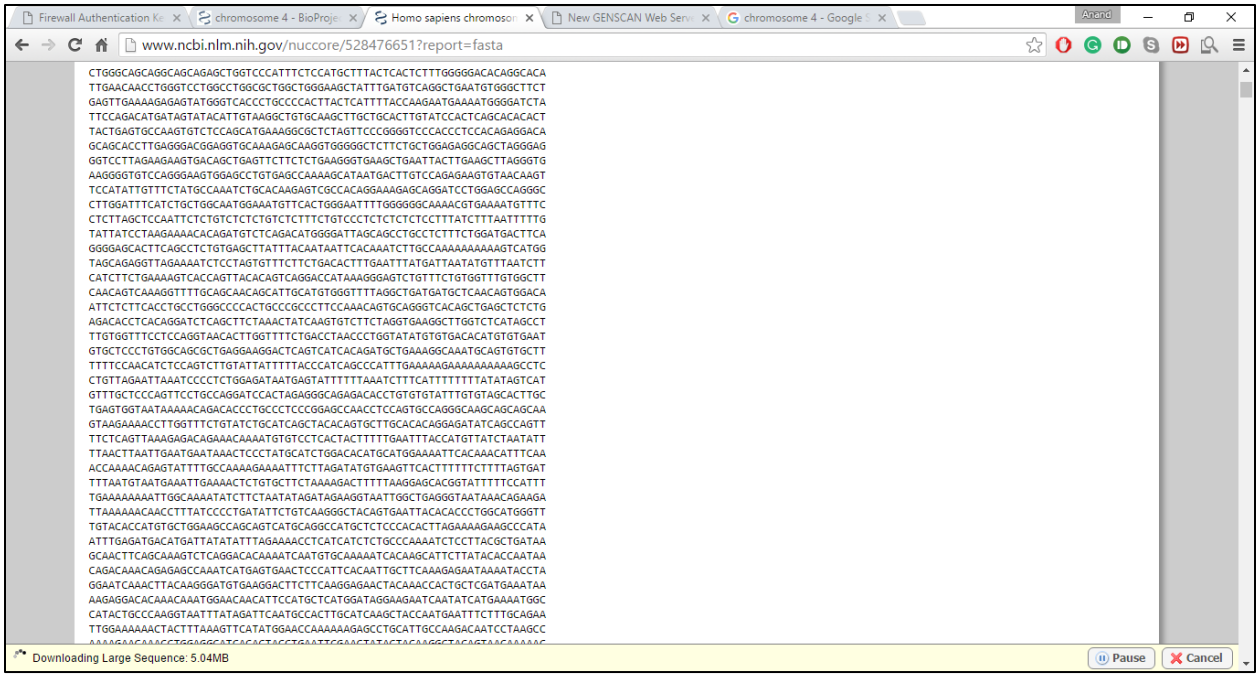


Figure 10 Downloaded chromosome 4 sequence

2. A small portion of the downloaded sequence was used to create a text file, which was later uploaded to Genscan website for processing. The default settings of the website were left unchanged. File was given the name chromosome_4_seq_1.txt. Run GENSCAN was clicked to start processing the input file.

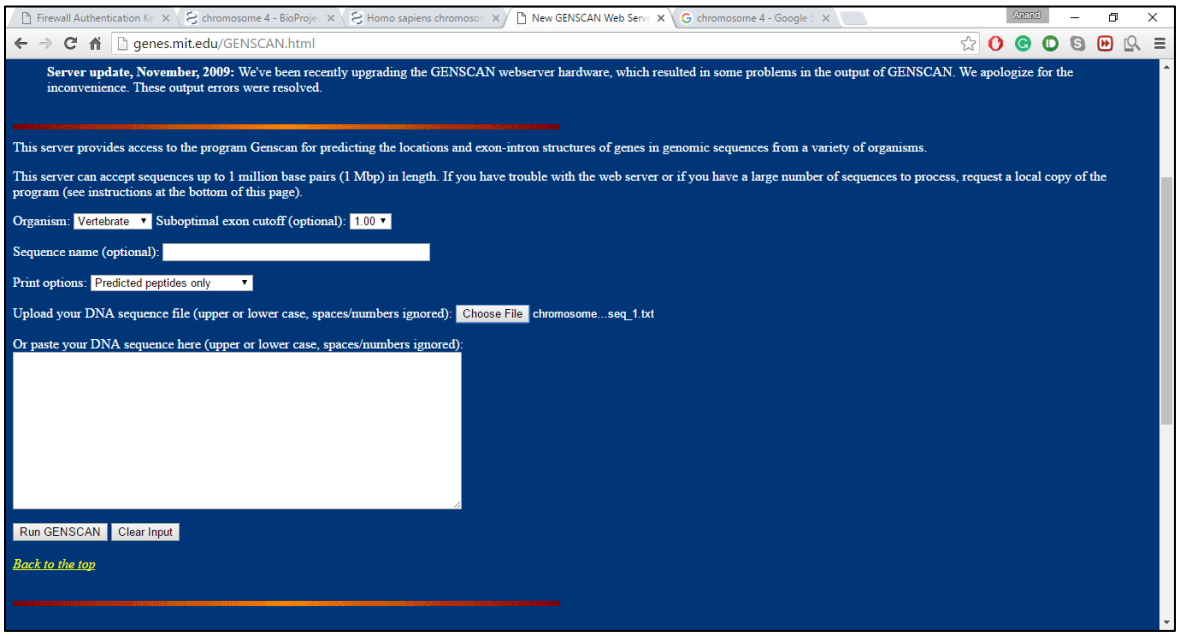


Figure 11 Genscan website

3. Following results were generated by by genscan after the processing was over. Predictions for a new set of genes or exons, and peptide sequences were provided as the output.

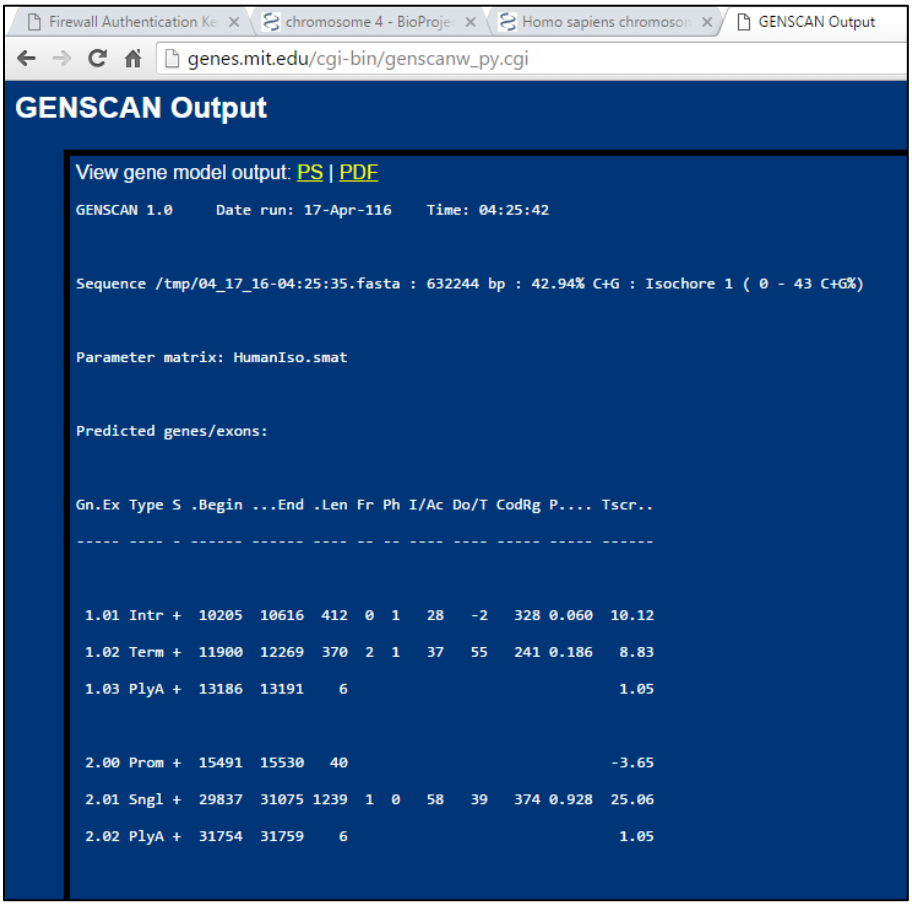


Figure 12 Predicted Genes or Exons

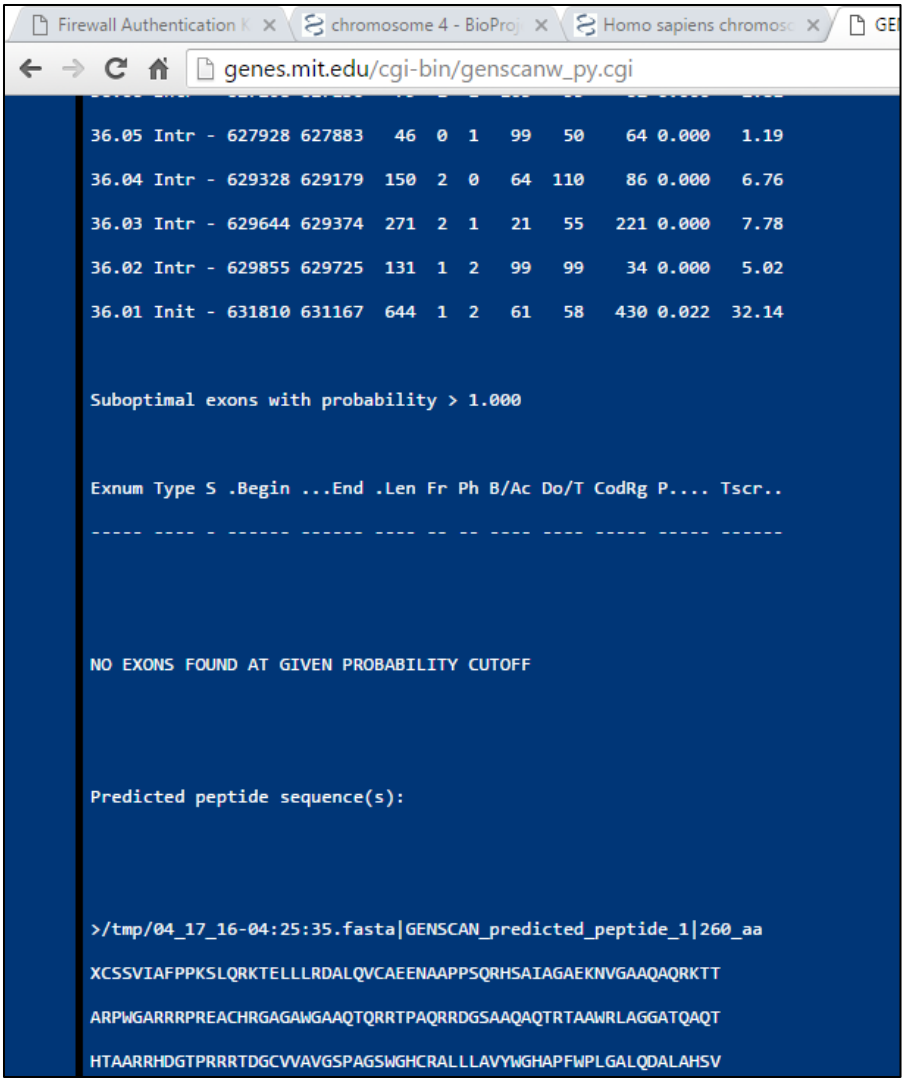


Figure 13 Predicted Peptide Sequence(s)