**MBG 6133 HW4**

The sequence, function, and structure knowledge of the protein world is deposited in Uniprot (<https://www.uniprot.org/>). Uniprot assigns a unique identifier to each protein, as in Q9BYF1 for human ACE2: [https://www.uniprot.org/uniprotkb/**Q9BYF1**/entry](https://www.uniprot.org/uniprotkb/Q9BYF1/entry)

This homework will guide you in downloading files directly from Uniport and working with them.

1. Create a folder, called Uniprot\_related

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

1. Pick five proteins of interest from any organism. List their Uniprot id’s below:

**Q6YGZ4, U5L0E8, B8Q4T4, G9IBF2, B4ZAC3**

Also enter these id’s in the Uniprot\_id\_list file that you will create and place under Uniprot\_related

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

1. Download the fasta sequences of these proteins in Uniprot\_related by updating the following command:

wget <https://rest.uniprot.org/uniprotkb/Q9BYF1.fasta>

OR

wget [https://rest.uniprot.org/uniprotkb/[Uniprot\_ID].fasta](https://rest.uniprot.org/uniprotkb/%5bUniprot_ID%5d.fasta)

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

1. By using a for loop in bash, write a script that counts the number of lines present in each sequence file.

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

**I made the script executable with**

**chmod 777 line-count.sh**

**command**

1. Update your script so that your script also calculates the length of each sequence.

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

1. Update your script so that the Uniprot id will be printed next to the sequence length calculated.
2. Write the final output in Uniprot\_id\_stat file.

metin içeren bir resim

Açıklama otomatik olarak oluşturuldu

Please provide all the necessary inputs, outputs, and scripts, together with this word file that I can use to track down what you did as a single zip file. Name you zip file as HW3\_NAME\_SURNAME