

BIFS 617: Advanced Bioinformatics

Dr. Alkharouf

Assignment # 4

Instructions:

Submit your solutions to the questions below in ONE Python script file, name it using the following format please: YourLastName_A4.py. Add comments to indicate which question each code snippet belongs to.

Your program must reflect your own work. For the discussion forums you are encouraged to work with your colleagues, but for the assignments please work on your own.

- 1) (4 pts) Write a program that tests to see whether a sequence has an AT repeat. For this problem, we'll define a repeat as 3 or more occurrences of AT, i.e. ATATAT...etc. Read in the sequence from a file (in FASTA format) to test your code. You can use the attached "Test.txt" sequence.

Return a message to the user whether or not an AT repeat exists in the sequence. Use at least one function.

- 2) (6 pts) EMBL (European Molecular Biology Laboratory) and EBI (European Bioinformatics Institute) are basically the equivalent of NCBI here in the US. They hold the same data but store the data in a different format. Your task is to parse out an EMBL record (see file attached) just like we did for GenBank records in the discussions.

EMBL's records are actually easier to parse out! Just parse out the sequence ID (line starts with ID), description (DE) and sequence (SQ). Use at least one function.