BIFS 617: Advanced Bioinformatics
Dr. Alkharouf
Midterm Exam

Instructions:

Submit your solutions to the questions below in ONE Python script file, name it using the following format please: YourLastName_Midterm.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 and exams please work on your own.

1) Attached is a file called genomic_dna.txt. It contains a DNA sequence that is comprised of two exons and an intron. The first exon runs from the start of the sequence to the 63 bp, and the second exon runs from the 91 bp to the end of the sequence. Write a program that will print out to files the coding and non-coding regions of the sequence (i.e. The exons in one file called coding.txt, and the intron into another file called non_coding.txt).

2) Modify your code for Q1 above so it calculates what percentage of the sequence is coding and displays the result to the screen.

3) Attached is a file called sequences.txt, it contains 3 sequences (one sequence per line). Also attached is a file called AccessionNumbers.txt. Write a program that reads in those files and produces 3 separate FATSA files. Each accession number in the AccessionNumbers.txt file corresponds to a sequence in the sequences.txt file. Remember a FASTA formatted sequence looks like this:

   ```
   >ABCD1234
   ATGCTTTACGTCTACTGTCGTATGCTTTACGTCTACTGACTGTCGTATGCTTACGTCTACTGTCG
   ```

   The file name should match the accession numbers, so for 1st one it should be called ABCD1234.txt.

   Note: Print out the sequences in upper case AND remove any special characters in the sequence!

4) Write a program that checks to see if two DNA sequences given as input by the user are reverse compliments of one another.

5) Write a program to read a file, and then print its lines in reverse order, the last line first. You can use the sequence.txt file (attached) to test your program with.

6) Write a program that will predict the size of a population of organisms. The program should ask for the starting number of organisms, their average daily population increase (as a percentage), and the number of days they will multiply. For example, a population might begin with two organisms, have an average daily increase of 50 percent, and will be allowed to multiply for seven days. The program should use a for loop to display the size of the population for each day. So for the previous example, the output should look like:

```
Day              Organisms
----------------------------
1                 2.0
2                 3.0
3                 4.5
4                 6.75
5                 10.125
6                 15.1875
7                 22.78125
```

Input validation: Do not accept a number less than 2 for the staring size of the population. Do not accept a negative number for average percent daily population increase. Do not accept a number less than 1 for the number of days they will multiply.

7) Write a program that takes user entered lines from the keyboard and stores them in an array. When the user enters "quit", the program prints out all the lines sorted (i.e. a line starting with an “Ab…” would print out before one starting with “Ac…”).

8) Now modify the program so it tells you how many lines have been entered, and then prints out only lines 2, 3, and 4.

9) Ask the user for a list of sequence lengths, separated by whitespace (Example: 100 123 45 …etc.). Store the sequence lengths as one string in a variable, then:

a) Split that string and create an array from it (each number being an element of the array)
b) Use a for loop to get the sum of all sequence lengths.
c) Print the average.

10) Write a program that asks the user to enter the number of calories and fat grams in a food item. The program should display the percentage of the calories that come from fat. One gram of fat has 9 calories, therefore:

Calories from fat = fat grams * 9

The percentage of calories from fat can be calculated as follows:

Calories from fat / total calories

If the calories from fat are less than 30 percent of the total calories of the food, it should also display a message indicating the food is low in fat.

Note: The number of calories from fat cannot be greater than the total number of calories in the food item. If the program determines that the number of calories from fat is greater than the number of calories in the food item, it should display an error message indicating the input is invalid.