

CSC / BIO 310

Bioinformatics

Instructor: Dr. Laurie J. Heyer

Assignment #5

Due Thursday, Feb 7

Instructions: Write comments in your file to indicate the solutions to each of the following problems. You should also include the number of each problem in the print statement.

Write a comment at the top of the program containing the five lines at the top of this assignment page, in addition to the names of people in your programming team.

You may not consult with anyone outside of your programming team, other than me.

In addition to the accuracy of your solutions, you will be graded on the readability of your code and your output, and the use of good programming practices as discussed in the text and in class.

1. Write a subroutine that returns the desired codon from a given DNA sequence that contains a coding sequence. The subroutine should take two arguments: the DNA sequence and the number of the codon that will be returned. You may assume that the first ATG in the sequence is the first codon. Test your subroutine with this DNA sequence:
AcGTtTAtGcgaTtaGAcGTaGctAtCGaTtAAgCaACC
In your main program, write a loop to call your subroutine for each codon beginning with the start codon and continuing until a stop codon is reached. The loop in your main program should also do some error checking based on the value returned from the subroutine. Print each codon on a separate line.
2. Modify the code in regex.pl to create a subroutine that returns the number of 7-mers in ecoli.txt that contain a particular regular expression. The subroutine should have one argument, the desired regex. The subroutine should not print each 7-mer that matches, but merely return the number.
3. Assuming that there is only one missing 7-mer in ecoli.txt, write a loop in your main program to automate the search procedure for the missing 7-mer. Your loop should call the subroutine you created in #2. Print the missing 7-mer.