

CSC / BIO 310

Bioinformatics

Instructor: Dr. Laurie J. Heyer

Assignment #4

Due Tuesday, Feb 5

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. Generate a single DNA nucleotide using the probability model in which the probability that a base is an A is 0.18, and the remaining probabilities are determined as we discussed in class, assuming balanced representation of complementary bases (A's and T's, C's and G's) on each strand. You will need to use an If, elsif, ... structure to decide which letter to produce, based on the value of a single random number.
2. Generate a random string of 100 DNA bases according to the model described in #1. To do this problem, you should embed the code you wrote for #1 inside of a while loop. The code on page 109 is a good template on which you can model your while loop. In this problem, you need to count down from 100 (or count from 1 up to 100) to progressively produce your nascent string of DNA.
3. Determine the numbers of A, C, G and T in the random DNA sequence you created in #1, and print the expected and observed numbers, and the percent "error" in the observations. Print your results in a nicely formatted table using printf.
4. Compute and report the GC-skew of the sequence you made in #1. Protect your code from a division by 0 error. You can read about GC-skew (and check your answer!) by using this online tool:
http://www.genomicsplace.com/gc_skew/gc_skew.html