Homework 3

String manipulation

Write a python program to calculate some statistics from this DNA sequence. You will need to initialize a string variable in your script so your code can start something like this template code I started for you here.

1. Calculate and print the GC content of the DNA

2. Find the location of all the ATG codons in the sequence

3. Print the reverse complement of the sequence

Math calculations

Download the genome annotation for the Chr6 of Rice again (or reuse what you had from before) Rice Chr6 annotation.

1. Compute the number of gene and exon features in the file using python

2. Compute the numer of bases which are in genes and in exons

3. Report the % of the chromosome which is exonic (e.g. covered by exons). Assume the CDS exons in the GFF files are NON-overlapping for this problem.

- If you are a more advanced programmer, try to solve this problem by also correcting for the fact that alternative splicing isoforms will produce redundant instances of a CDS. You could also manipulate the input GFF3 file if that is easier - you do not have to read in the GFF3 - you could for example convert to BED format . . .

Use the following basic code template