

Python 3 Problem Set

1. Write a script that calculates the AT content of the DNA string found below. AT content is the proportion of bases that are either A or T. You will need to use some mathematical functions as well as a python function. It is ALWAYS a good idea to test your code with test data. For example, the below sequence is long and you don't know for sure how many As and Ts are present. Test your code with a DNA string that you KNOW the correct answer. In AATTGGCCA you know you have 3 As and 2 Ts.

```
1 GATGGGATTGGGGTTTTCCCCTCCCATGTGCTCAAGACTGGCGCTAAAAGTTTTGAGCTTCTCAAAAGTCTAGAGCCACCGTCC
```

2. Now add in lines of code to calculate GC content of the above DNA string. Run with test data first!!
3. Write a new script that prints out the reverse complement of the above DNA string.

Original Sequence 5'ATGCAGGGGAAACATGATTCAGGAC 3' Complement 3'TACGTCC-CCTTTGTACTAAGTCCTG 5' Reverse Complement 5'GTCCTGAATCATGTTTCCCCTG-CAT 3'

4. Make sure to commit your changes along the way. You can wait until the end to push them to your remote repo, if you like, or you can do it now. It is probably smart to commit after each problem set question.
5. Write a script to find the start position of an (EcoRI)[<https://www.neb.com/products/r0101-ecori#Product%20Information>] in the above sequence. Remember DNA sequences start with a 1 and a python string starts with a 0. Run with test data first.
6. What is the location in the DNA string of the EcoR1 cut site? Calculate the first and last nucleotides of each restriction fragment.
7. Extract the the restricion fragments. Print each fragment along with its postion in the whole DNA sequence and its length.

```
1 Fragment\tPostion\tLength
```

10. ADD/COMMIT/PUSH