

Project Report: Script Programming

By

Faisal Fiaz 19920420-T454

In this project there were different tasks to calculate if the attribute contains valid bases, comparisons of the DNA sequences, finding the pair of non-matching bases, reading data from the .dat files and printing total number of bases, finding the separators in the files and plotting the length of genes as a bar charts and making a scatter plot of the number of swap mutations per gene against gene length.

Task 1 :

In first task we have to check if seq contains valid bases. For this we have to write code in `isdna` method which returns whether seq contains valid bases and then after checking that returns True otherwise False. So we use a loop which checks the length of seq, if seq not in self bases then it passes False other wise the function will return passes. So for testing purpose if we uncomment the line `myseq=dna.dnasequence(seq="U")` in the `project.py` file and comment the `myseq=dna.dnasequence(seq=mystr)` in the same file then it will show our desired result. So if we do this it will return output "seq contains invalid bases. Exiting program."

Task 2 :

In this task we have to make a method which will returns the complement of dna sequences. For this purpose I have made a method name `complement` which will return the complement of dna sequence and it will also return the comparison of two strings by using `isequal` method which is already defined. So it will compare the two strings name `crick_str` and `watson_str`, if they are complements then it will return True and if they are not complement it will return False.

Task 3 :

In this task we have created a method for finding the first pair of non-matching bases in 2 sequences. Therefore, I have created a method in `dnasequence` file named ***compare*** and in this method, I am checking if the bases are of equal length.

If they are of equal length the method should return the index of mismatching bases. If they are not of equal length then the method is terminating the program and if method

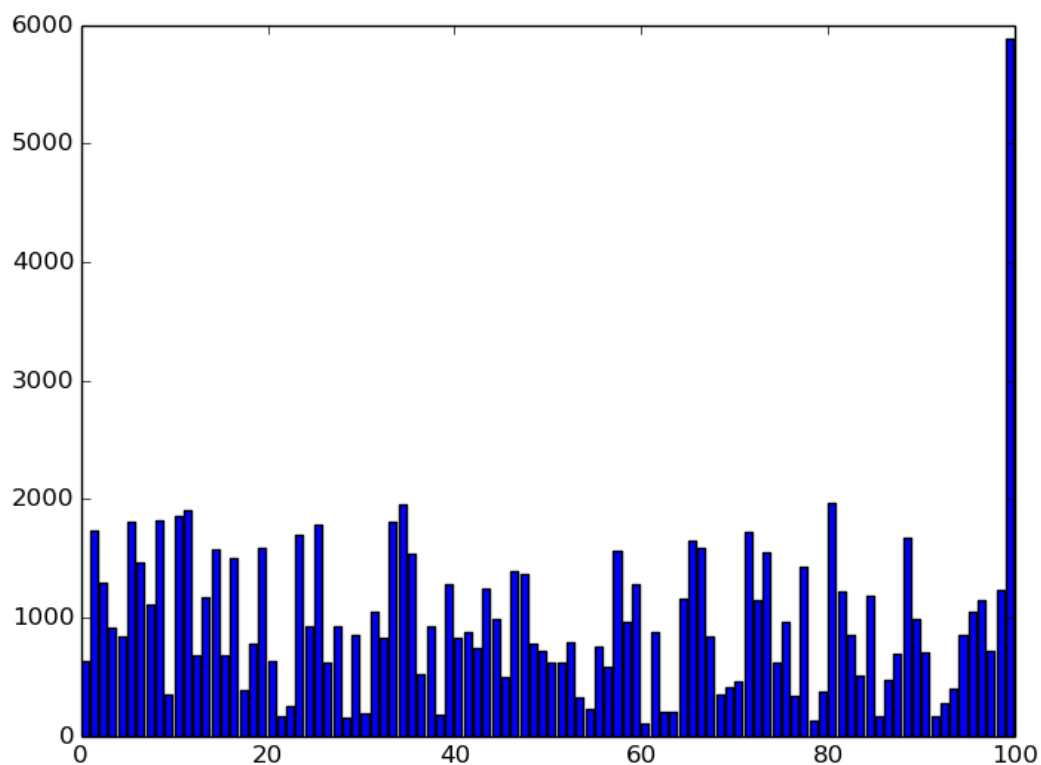
did not find any mismatch then it is returning -1. So for this task I have used if else statements and the output of first non matching bases in crick_str and crick2_str is 27.

Task 4 :

In this task we have to have to calculate the total number of bases in the genome_01 file. So for this I have added a new function in the project.py file named ***def read()***. This function reads the file of genome_01 and then prints the length of total number of bases this file contains.

Task 5 :

In this task we have to first read the file genome_01 and then find the separator “AAAAAAAAAATTTTTTTTTT”. After this it prints the length of first gene and plot the lengths as bar charts for the rest of genes. To do this, I have made a method named ***gene()*** in the dnasequences.py file which first converts the string into tostring() and then find the separator in the genome_01 file and then returns the length of first gene. After this it creates the list of dna sequences in the genome_01 file and plot the length of the genes as a bar charts.



Task 6 :

In this task we have to read two files name genome_01 and genome_02. After reading both files we have to count the number of non-equal bases in each gene. For this I have made a function name ***find_nonmatch()***. It will count all the non equal bases in each gene and also make a scatter plot of the number of swap mutations per gene against gene length.

