

Task

I. Multiple introns from genomic DNA: (15 points)

The file `genomic_dna.txt` contains a section of genomic DNA, and the file `introns.txt` contains a list of start/stop positions of introns. Each intron is on a separate line and the start and stop positions are separated by a comma.

Write a function that will exclude the intron segments and return the dna sequence without introns.

The function:

1. **Function name:** `ExcludeIntrons`
2. **Input parameters:**
 - i. `dnaFileName.txt`
 - ii. `intronsFileName.txt`
3. **Return value:**
 - i. Dna sequence without introns

Example:

```
def ExcludeIntrons (inputFileName, intronsFileName):  
    #Implementation  
    return outputDna
```

Output from the given file should be:

```
returned=ExcludeIntrons("genomic_dna.txt","introns.txt")  
print(returned)
```

GAAT

Note that:

- Please be sure that the output from this function doesn't have any trailing spaces or newlines at the end of the text.
- Start counting the nucleotides from zero

II. Fragment lengths (5 points)

Implement a function should calculate the size of the **two fragments** that will be produced when the DNA sequence is digested with restriction enzyme

ACTGATCGATTACGTATAGTAGAATTCTATCATACATATATATCGATGCGTTTCAT

Hint: The sequence contains a recognition site for the restriction enzyme, which cuts at the G*AATTC (the position of the cut is indicated by an asterisk).

The function:

4. Function name: getFragmentSize

5. Input parameters:

- i. the sequence in string format

6. Return value:

- i. Length of the first fragment
- ii. Length of the second fragment

Example:

```
def getFragmentSize(seq):  
    return len(fragment1), len(fragment2)
```

```
Fragment1, Fragment2 = getFragmentSize("ACTGATCGATTACGTATAGTAGAATTCTATCATACATATATCGATGCGTTTCAT")  
print(Fragment1)  
print(Fragment2)
```

22

33

III. Gene Expression (10 points)

- In the task folder, you'll find a text file called data.csv, containing some made-up data for a number of genes.
- Each line contains the following fields for a single gene in this order:
 - Species name
 - Sequence
 - Gene name
 - Expression level

Using data.csv to:

1. Print out the gene names for all genes belonging to *Drosophila melanogaster* or *Drosophila simulans*.
2. Print out the gene names for all genes between 90 and 110 bases long.
3. Print out the gene names for all genes whose AT content is less than 0.5 and whose expression level is greater than 200.
4. Print out the gene names for all genes whose name begins with "k" or "h" except those belonging to *Drosophila melanogaster*.
5. For each gene, print out a message giving the gene name and saying whether its AT content is high (greater than 0.65), low (less than 0.45) or medium (between 0.45 and 0.65).

Important Notes:

- 1) This task is an individual task
- 2) Add all the functions in a .py file
- 3) Name the .py file <Task1_secNo_BN>.py (for example : if you are in section 1 with BN 23 then the project file name is Task1_1_23.py)

Submission Deadline: 11:59 PM on May 30, 2022