 Due: 16^{th} Sept

CMPSC 300
Bioinformatics
Fall 2019

Lab 2:
DNA and Python3 Basics

Summary

To strengthen the understanding of DNA structure and DNA replication. To learn and enhance your Python3 programming skills, including how to create variables, write assignment statements, manipulate lists, create repetitive and conditional statements. To utilize basic Python programming skills to write a program that processes and manipulates the DNA sequence.

GitHub starter link

https://classroom.github.com/a/ntDDrfDl

To use this link, please follow the steps below.

- Click on the link and accept the assignment.
- Once the importing task has completed, click on the created assignment link which will take you to your newly created GitHub repository for this lab.
- Clone this repository (bearing your name) and work on the practical locally.
- As you are working on your practical, you are to commit and push regularly. You can use the following commands to add a single file, you must be in the directory where the file is located (or add the path to the file in the command):
  - `git add -A`
  - `git commit -m ‘‘Your notes about commit here’’`
  - `git push`

Handed out: 9^{th} Sept. 2019

Alternatively, you can use the following commands to add multiple files from your repository:

- `git commit <`*nameOfFile*`> -m ''Your notes about commit here''`
- `git push`

Be sure to read the `README.md` file in the GitHub Classroom repository for instructions on how to complete your first assignment.

DNA and Python Basics

Bioinformatics requires the use of computational techniques to solve biologically-related problems. Therefore, in this course it is essential for you to have some basic understanding of both biological and computer science concepts related to Bioinformatics before we approach Bioinformatics problems.

Seemingly all goals of Bioinformatics research depend on automation for completion. In many cases, bioinformaticians are required to create their own programs and tools in order to complete their cutting-edge work. In this lab, you will also create a basic tool to help in your future work in the field.

Part 1: DNA Structure Worksheet

Navigate to the activity your GitHub Classroom repository (`activity/activitySheet.md`). Please complete the assignment by writing in your responses in markdown formatting into the file.

Part 2: Getting to Know Python

Python Help

Before we can start writing Python programs for Bioinformatics solutions, you need to get comfortable with the structure and the syntax of this programming language. Just like with any natural or programming language, you need to learn various syntactical and semantic rules when writing in Python.

If you feel that your skill in Python3 needs some improvement, you are invited to read Chapters 1 and 2 in the "Think Python" textbook. To get started with Python, please complete Sections 3 and 4.1-4.5 in the Python Tutorial, which can be found in: `https://docs.python.org/3/tutorial/index.html`. NOTE: if you have a previous experience in using Python, please browse through the tutorial and then consider lending your expertise to your fellow classmate.

Now that you have been exposed to some basic rules of programming in Python, you are asked to practice those skills by writing programs to accomplish the following small tasks.

Python3 Coding: a DNA Base Counting Program

You are to write a Python program that:

Given: A DNA string of length at most 1000 nt.
Returns: Four integers (separated by spaces) counting the respective number of times that the bases,'A', 'C', 'G', and 'T' occur in the sequence. In addition, you are to add the pair-counts of "AT", "TA". "GC" and "CG". Your programming is to simply go though the string and count these combinations.

Sample Input:
AAAACCCGGT
Sample Output:

- A's: 4
- C's: 3
- G's: 2
- T's: 1

Note: your source file is to be called: `baseCounter.py`

Python3 Coding: Find the Complementary sequence of DNA

DNA's bases are paired as follows: A↔T and G↔C.

You are to write a Python program that:

Given: A DNA string of length at most 1000 nt.
Returns: The complementary base string of the entered DNA.

Sample Input:
AAAACCCGGT
Sample Output:

- TTTTGGGCCA

Note: your source file is to be called: `compSeq.py`

Ethical Portion

Locate the article in your repository by Michael Specter entitled, "Rewriting the code of life." (The New Yorker, 2017.

The reference to the article is `https://www.newyorker.com/magazine/2017/01/02/rewriting-the-code-of-life?reload=true`

Please complete the three questions found in the file, `ethics/reflections.md`

HANDED OUT: 9^{th} SEPT. 2019

Required Deliverables

- Your completed activity should be saved as `activity/activitySheet.md`.
- All of the deliverables (i.e., `baseCounter.py` and `comp.py`) of Part 2 are to be placed into your `src` folder in your GitHub Classroom repository. Please ensure that you have sent your files correctly to the GitHub correctly by visiting the web site of your repository and by checking the your source files. This will show you whether you have correctly pushed your files. Please ask questions, if necessary.
- The ethical reflections writing piece: `ethics/reflections.md` .

You should see the instructor if you have questions about assignment submission.