LAB #7
Bioinformatics and Python Programming

1. **Pre-Lab**: Make sure you finish implementing the two programs in the prior lab. In addition, look over the current lab, and begin designing your programs for the lab.

2. There have been many questions regarding the difference between an integer, int, and a floating point number, float.

   **Explain the difference between an int and float to receive 1 point for the lab!!!**

3. Since our most recent speakers have been talking about supercomputing and bioinformatics, let’s combine these topics in our lab. The San Diego Supercomputing Center, SDSC, hosts The Biology Workbench, which is a web application used by researchers for DNA sequencing. [http://workbench.sdsc.edu/](http://workbench.sdsc.edu/).

   There is a Biology Student Workbench, [http://medsocnet.ncsa.illinois.edu/MSSW/](http://medsocnet.ncsa.illinois.edu/MSSW/), that was developed by the National Center for Supercomputing Applications, NCSA. You need to register for an account with The Biology Workbench, [http://seqtool.sdsc.edu/register.cgi](http://seqtool.sdsc.edu/register.cgi), to use the Student Workbench. After you register for an account, then you can log into the Biology Student Workbench, [http://bighorn.animal.uiuc.edu/cgi-bin/sib.py](http://bighorn.animal.uiuc.edu/cgi-bin/sib.py) (notice that it is written in Python!).

   Log into the Student Interface to the Biology Workbench, and press the “New” button to create a new session called, “CS 160 Lab”. After this, you will be automatically be taken into this session, i.e. notice the current session change at the top. You can change the color of this session to visually distinguish it from others. Now, click the “Protein Tools” button, at the top of the page.

   Now, search for “Human DNA chromosome” from the “GenBank Primate Sequences” database, and select a DNA sequence to import (I used Human chromosome 20). Then, go to the bottom of the list and select “Import Sequence”. Next, search for “Mouse DNA” from the “GenBank Rodent Sequences” (I used the Mouse chromosome X), and import this into your proteins. Pick more proteins from other animals/plants, and import those sequences.

   Now, go to the bottom of the page and select the proteins to compare. Then, press the CLUSTALW button on the right to compare the proteins. Now, press the “Import Alignment” button at the top, and then use the different tools on the left to visualize the alignment.

   **Show a lab TA your analysis of protein sequences to receive 2 points on this lab.**
4. Before you write a program to compare two strings for similarities, you will take baby steps. First, write a program that takes a string from the user and prints the string horizontally, by supplying the string to the print function, prints the first and last character in the string, and then print the string vertically, one character at a time.

You will need to use the `len()` function that determines the length of a string, and you will have to use the brackets, `[ ]`, to access each character in the string. For example, here is a short program that takes the user's name as input, and it prints the first and last character in the user's name:

```python
name = input("Enter you name: ")
print("The first character in your name is: " + name[0])
print("The last character in your name is: " + name[len(name)-1])
```

Notice that the character in the string is accessed using 0 in the brackets.

1. Why do we use `len(name)-1` in the brackets to access the last character?
2. How does the program behave when you remove the `-1`?

Show a lab TA your short program and answer the two questions above to receive 2 points on this lab.

5. Now, adapt your program to take two strings as input, and you will compare the two strings, character by character. You need to count the number of occurrences where the characters in the two strings match, and you will output the percentage of matching characters between the two strings.

If the two strings are not the same length, then they are not matching for those extra characters. For example, Jennifer and Jen match on 37.5% of the characters.

In order to receive the rest of the 5 points for this lab, show your Python code for matching characters in two strings to a lab TA!!!