

### **BF527: Applications in Bioinformatics**

Please prepare a typed report to submit through Blackboard. Please include printouts of all Python code that you write. Your code should follow the guidelines laid out in class, including commenting. Partial credit will be given for nonfunctional code that is logical and well commented. This assignment must be completed on your own.

## **HOMEWORK #2**

**See Blackboard for assignment and due dates**

### **PROBLEM 2.1 (50%):**

Thousands of genomes are being sequenced. To understand the motivation behind sequencing and the insights gained from sequencing, we would like you to read a report on a sequenced organism and answer some questions about it. Here are some compendia of sequenced organisms:

[http://en.wikipedia.org/wiki/List\\_of\\_sequenced\\_eukaryotic\\_genomes](http://en.wikipedia.org/wiki/List_of_sequenced_eukaryotic_genomes)

[http://en.wikipedia.org/wiki/List\\_of\\_sequenced\\_bacterial\\_genomes](http://en.wikipedia.org/wiki/List_of_sequenced_bacterial_genomes)

[http://en.wikipedia.org/wiki/List\\_of\\_sequenced\\_archaeal\\_genomes](http://en.wikipedia.org/wiki/List_of_sequenced_archaeal_genomes)

Select an organism with an available reference and then locate and skim the associated paper. Answer the following questions (1-2 sentences each).

- (1)** Why was this organism selected for study?
- (2)** What technology was used in the sequencing project?
- (3)** Give an example of how the specific biology or lifestyle of the organism is reflected in its genome.
- (4.5)** List two other facts of interest reported in the paper.

Please provide a citation to the article. You may replace 1-3 with other facts of interest if they prove difficult to answer based on your selected paper.

**THOUGHT QUESTION (NOT GRADED).** Is there a question raised in the paper that could be answered with additional bioinformatics work?

**PROBLEM 2.2 (25%):**

```
indices = [22, 0, 39, -6, 14, 8, 30, -22]
sequence = 'IF WE KNEW WHAT WE WERE DOING IT WOULD NOT BE CALLED
RESEARCH'
```

Given the above sequence and list of indices, use a ~~for~~ loop to cycle through the list of indices and print BOTH the index and the corresponding amino acid at that position in the sequence.

Your result should look like:

```
22  E
0   I
39  N
-6  S
14  T
8   E
30  I
-22 N
```

**PROBLEM 2.3 (25%):**

```
sequence = 'DELGCFVGTA'
```

Given the above protein sequence, use a **for** loop and **if** statement to cycle through each amino acid and print the amino acid and whether it is “hydrophobic” or “not hydrophobic”. The following amino acids are hydrophobic: A, V, I, L, M, F, Y, and W.

Your result should look like:

```
D  not hydrophobic
E  not hydrophobic
L  hydrophobic
G  not hydrophobic
C  not hydrophobic
F  hydrophobic
V  hydrophobic
G  not hydrophobic
T  not hydrophobic
A  hydrophobic
```