Python 8 Problem Set

Don't forget to use a small test data set when you are testing your code. Make sure you know what the correct answer should be

1. Take a mulit-FASTA Python_08.fasta file from user input and calculate the nucleotide composition for each sequence. Use a datastructure to keep count. Print out each sequence name and its compostion in this format `seqName\tA_count\tT_count\tG_count\C_count`

Here is a structure of a handy datastructure to store this information

```
1  seqs[geneName][nucleotide]=count
2
3  seqs['geneA']['A'] = 2
4  seqs['geneA']['T'] = 3
5  seqs['geneA']['G'] = 3
6  seqs['geneA']['C'] = 1
7
8
9  seqs['geneB']['A'] = 1
10 seqs['geneB']['T'] = 5
11 seqs['geneB']['G'] = 2
12 seqs['geneB']['C'] = 2
```

2. Write a script that takes a multi-FASTA file Python_08.fasta from user input and breaks each sequence into codons (every three nucleotides is a codon) in just the first frame. Print out a multi-FASTA of codons for each sequence called 'Python_08.codons-frame-1.nt'

3. Add in exception handling. Throw and handle (try/except) the exception

- if no input is provided

- if the file cannot be opened
- if the file does not end in '.fasta' or '.fa' or '.nt'
- if a non ATGCN charcter is found in the sequence

4. Now produce codons in the first three reading frames for each sequence and print out a FASTA sequence record for each frame and print to a file called 'Python_08.codons-3frames.nt'

For example

```
1  >seq1-frame-1-codons
2  ATG TTG
```

5. Now reverse complement each sequence and print out all six reading frames to a file called 'Python_08.codons-6frames.nt'

6. Translate each of the six reading frames into amino acids. Create one file for which you print the six reading frames (Python_08.codons-6frames.nt) and one file for which you print the translation of the six reading frames (Python_08.translated.aa).

7. Find the longest peptide sequence (M–>Stop) of all the six translated reading frames for a single sequence. Do this for all the sequence records. For each sequence, print out in FASTA format the six frames of codons to one file (Python_08.codons-6frames.nt), the translations to

a second file (Python_08.translated.aa), and the single longest translated peptide to a third file (Python_08.translated-longest.aa).

8. Finally determine which subset of codons produced the longest peptide for each sequence record. Print this to a fourth file in FASTA format (Python_08.orf-longest.nt).