Week 03 - Unlocking Packages
Assigned September 19th, 2016
Due September 26th, 2016

This week you will be repeating Week 2's assignment, with the exception that you get to use any available packages. So, yes you have to do it all again with the aid of a package!

1) Create a random sequence generator.

Requirements:

- Use 'argparse' to get command line arguments
- Biopython might be useful
- Produce multi-FASTA output, with a header as described below
    - Sequence count
    - Sequence GC content
    - Length of sequence
- *** Include optional argument to set a random seed ***

Example Usage:

```
./random-seq.py TOTAL GC READ_LENGTH --seed 12345
```

Example output:

```
>sequence_001 gc=0.30 length=20
ATGCCATTATATGCCATTAT
>sequence_002 gc=0.25 length=20
ATGCCATTATGAAAATTTCA
...
```

2) FASTQ Summary Stats

Files For This Exercise

- Staphylococcus aureus MiSeq FASTQ

Requirements

- Use 'argparse' package to get command line arguments
- Numpy and/or Biopython might be useful
- Output summary statistics of the FASTQ
    - Total Reads, Total Base Pairs
    - Min, Max and Mean read length
    - Read length distribution
    - Min, Max, and Mean quality per read

    - Mean quality per read
- *** Include optional file to output data to ***

Example usage:

```
./fastq-stats.py FASTQ --output stats.txt
```

Example Output (doesn't have to be exact!):

Total Reads: 120000
Total Base Pairs: 12234984bp
Mean Read (Min, Max): 107.33 (33, 210)
Read Lengths:
33:44
…
210: 20
Mean Quality (Min, Max) Per Read: 36.67 (13, 38)
Mean Quality Per Base:
1: 36
2: 37
…
210: 24

3) Extract Proteins From GenBank File

Files For This Exercise:

- Staphylococcus aureus N315 GenBank

Requirements:

- Use 'argparse' package to get command line arguments
- Biopython, use it!
- ***Output two multi-FASTA files of the CDS features
    - One is the nucleotide sequences
    - The other the amino acid sequences***
- ***Include optional argument to retain original order of CDS features.***

Example Usage:

```
./extract-cds.py GENBANK_FILE DNA_OUTPUT AA_OUTPUT --ordered
```

Example Output:

>SA_RS00145 product=chromosomal replication initiator protein DnaA
MGDAVLDQYVRTYIVLKLKSKPNKLHQMSKKYVSAKSQAQTLEYLMEQEWFTDEEM
DILKRGRNAKSHTKAKNTDVQTYRKSSAIEAVIGFLYLEKREERLEALLNKIITIVNER
>LOCUS_TAG product=...
MGD...TIV