BCH 571 Bioinformatics for Life Scientists

Lab 1

First install Python and PyCharm.

Program 1- (call this lab1.1.py)

Using Pycharm, program a "Hello World" program that prints "Hello Word" to the screen.

Program 2- (call this lab1.2.py)

Write a program that takes a FASTA file (input.FASTA) that contains to aligned sequences, and output two separate sequences in FASTA format (output1.FASTA and output2.FASTA). The input file has gap characters ("-") that need to be removed. Use only the functions and methods that I described in the first lecture: print, string concatenation, len, str, upper, lower, replace, getting part of a string, find, open, write, and close.

input.FASTA looks like:

 >Seq1
AGG-UGC-GA-CCAUACCGUGUUGAAAAU-UC-UGCAUCCCGUCC-GAUCUGCAA-AGACAAG-
CAACACAGGGCCC-AGUCAGUA--GU-GCGGUG-GGUGAC-CACGU-GCGAAU--ACU-GUGGUG-UU-GCACUUU-
-
>Seq2
GCC-UACGG--CCAUACUAGUCUGAAAAC-GC-CCGAUCUCGUCU-GAUCUCGGA-AGCUAAG-
CAGAUUCAGGCCU-GGUUAGUA--CUUGGAU-G-GGAGAC-C-GCCUGGGAAU--ACC-A-GGUG--
CAGUAGGCUUU

output1.FASTA should contain:
>Seq1
AGGUGCGACCAUACCGUGUUGAAAAUUCUGCAUCCCGUCCGAUCUGCAAAGACAAGCAACACAGGGCCCAGU
CAGUAGUGCGGUGGGUGACCACGUGCGAAUACUGUGGUGUUGCACUUU

output2.FASTA should contain:
>Seq2
GCCUACGGCCAUACUAGUCUGAAAACGCCCGAUCUCGUCUGAUCUCGGAAGCUAAGCAGAUUCAGGCCUGG
UUAGUACUUGGAUGGGAGACCGCCUGGGAAUACCAGGUGCAGUAGGCUUU