

Python Unit 5 programming problem

October 13, 2016

This is the exercise for this week.

Exercise 1. Modify your Unit4 script so that it:

1. Reads a FASTA file and extracts:
 - (a) the accession number. We expect the sequences to be refseq and contain an accession number composed by the prefix “NM_” or “XM_” followed by a number (several digits), a dot and a number. For example NM_230423.3 or XM_000123.1
 - (b) the DNA sequence.
 - (c) [Note: use the files Human_RHOA.fasta and Human_VHL.fasta]
2. The program must accept the filename as an argument (do not hard-code the input filename).
3. Reads the standard genetic code from files “GeneticCode_standard.csv” and extract the information (codons and amino acid letter code) using regular expressions instead of using the string method split(\t).
4. Translates the DNA sequence in all six coding frames according to the standard genetic code.
5. Saves all translations to a single text file whose name is the accession number you got in step 1a.

Happy coding!