

MCB 5472 Assignment #2: Introduction to Perl part #2

January 29/14

The purpose of assignment #2 is to introduce some more complex Perl operations and data structures and to use them to perform simple bioinformatics tasks. In particular we will use: arrays; input arguments; for and foreach loops; simple regular expressions; and hashes. We will also work in groups to figure out the logic of the code that we are trying to write and draft it out as pseudocode for clarity. Implementation will be done individually.

Before you start: download any protein-coding gene from NCBI in fasta format. You will be using this as your input for the scripts that you write today. For each script, supply the input and output file names as command arguments; do not hard-code them as part of the script.

In class: Split into groups of 4-5 people and develop pseudocode for each of the assignment questions, checking with Peter and/or Jonathan before you move on to the next question. Once you have your pseudocode logic approved, then move on to actually writing the required code.

Pseudocode is a method of outlining programming logic without worrying about exact syntax. For example, see the assignment #2 lecture notes.

Before the start of the next class: Write each script and run it. Send the script, pseudocode, input file, and output file to Jonathan as a .zip folder (jonathan.klassen@uconn.edu; please remember to put MCB 5472 somewhere in the subject line).

Please use the following naming convention:

Script: [your name]_[question number].pl

Output: [your name]_[question number].out

Input fasta file: [your name].input.fasta

- (1) [5 marks] Write a script that takes reads your input file and prints it to an output file without the fasta header.
- (2) [5 marks] Write a script that takes your input sequence and outputs its reverse compliment sequence.
- (3) [5 marks] Write a script that takes your input sequence and outputs its translation.