Assignment 3 – Due on October 5

1) Write a Parser in Python that reads in Balati_COI.fst and writes a fasta file to the standard output (stdout) in which every sequence was cropped by 10 nucleotides from the 5' end (beginning of sequence) and 5 nucleotides from the 3' end (end of sequence).

   You may use BioPython or a dictionary (see the example scripts we worked through).

   In addition to cropping the 5' and 3' ends of each sequence, rename each sequence. Every sequence should be labeled in the following manner: 'Genus_Species_AccessionNumber' (e.g., >Holothuria_mammata_GQ214743.1). Recall that strings can be converted to lists with the *split* method/function. Lists, in turn, can be converted into strings using the *join* function.

   http://www.pythonforbeginners.com/data-types/python-join-examples - follow the examples in the tutorial and then apply what you learned to your script. Note that you can do better than the examples in the link: tabs are encoded by \t and newlines by \n.

   Hint: strings are indexed in the same way lists are.

2) Write a script in Python that translates the mitochondrial sequences contained in Balati_COI.fst in the 3 forward reading frames. Choose the appropriate genetic code for the translations! Questions to ask yourself when choosing the genetic code: What organism(s) am I dealing with? Which organelle in the cell do the sequences originate from?

   Your script should take a file name as an argument, so that the script can be used for any fasta file. Use the appropriate BioPython methods to translate each sequence using the applicable genetic code (see https://www.ncbi.nlm.nih.gov/Taxonomy/Utils/wprintgc.cgi for all available codes).

   The tutorial at http://biopython.org/wiki/Seq should have all the information you need.

   Additional exercise – not required:
   How could you make the script more flexible and allow the user to specify the genetic code to be used? There are several ways to achieve this in Python. One way would be to ask the user for input: http://www.pythonforbeginners.com/basics/getting-user-input-from-the-keyboard. Can you write a script that asks the user to specify the genetic code to be applied to the translations? Consider printing some instructions prior to asking the user for input. You may want to verify the user's input prior to executing the rest of your script's instructions! A conditional statement could help here; if the user specifies invalid output you may want to print an informative error message.

Comment your code thoroughly, so that others understand what it is doing (this also demonstrates that you understand what your code does).

Please submit your 2 scripts (3 if you do the additional exercise) to me as a tar archive by the due date.