BIOL 8803B
Exercises Week 10
November 2nd, 2015

Terms and operators for this week:

`=~` - Perform a regular expression operation

`s///` - Substitute a regular expression

`m//` - Match something to a regular expression

`.` – Match any character

`*` - Match 0 or more times

`+` - Match 1 or more times

`{}` – Match a certain number of times

`[]` – Match anything in the brackets

`[^x]` – Any character but x, note this is different from the normal `^` which means the start of the line

`?` – Make something in a regular expression optional and non-greedy

`()` – Group things in a regular expression

`|` - Allow choices in a regular expression

`~~` - The smart match operator

`qr` – Make a regular expression a scalar

`s` modifier – Makes the regex go past newline characters

`m` modifier – Make the ^ and $ special characters usable with each newline character

`g` modifier – Make a regular expression keep going after the first match

`say` – Like `print` but adds a newline to everything

Exercise:

1.) All 2 FASTA!

FASTA is generally acknowledged to be the greatest format ever. Therefore, this week you will be writing a script to convert any file that is given to it into a FASTA file with the same name, but with a .fna or .faa extension, depending on whether or not it contains nucleic acid or amino acid sequence. It is up to you and your script to determine the file format and the sequence type.

It should accept the following file types:

EMBL	http://www.bioperl.org/wiki/EMBL_sequence_format
FASTQ	http://www.bioperl.org/wiki/FASTQ_sequence_format
GenBank	http://www.bioperl.org/wiki/GenBank_sequence_format
MEGA	http://www.bioperl.org/wiki/MEGA_multiple_alignment_format
PIR	http://www.bioperl.org/wiki/PIR_sequence_format

- Code:
    - all2fasta.pl

- Your code should an input such as `a_file` and produce an output file called `a_file.fna`, or `a_file.faa` if `a_file` had amino acid sequence.
- Do not make your script work on extension detection. The format should be determined based on the file contents.