Programming for Bioinformatics

BIOL 7200

November 28th, 2016

Concepts and such for this week:

- Class
- Object
- Inheritance
- Subclass
- Passing arguments as a hash
- `package` – tell the interpreter what package you are in
- `bless` – make a hash reference a full-fledged object

Exercises:

BioPerl is a huge set of Perl classes which allow you to, among many other things, manipulate sequences.

0) Install BioPerl on your system using CPAN. It's straight-forward (and google-able).

Script 1

1) Use Bio::SeqIO to load a GenBank file
2) Convert it to a FASTA file
3) Convert it to an EMBL file

Script 2

4) Use the Bio::Tools::Run:StandAloneBlast module to BLAST multiple sequences (testSpeciesA.fasta) against a sequence database (testSpeciesB.fasta). The documentation for StandAloneBlast can be found at http://search.cpan.org/~cjfields/BioPerl-1.6.924/Bio/Tools/Run/StandAloneBlast.pm
5) Use a parser to find the top two best hits for each sequence from the returned BLAST parsing object. The parser will be one of the following: Bio::Tools::BPbl2seq, Bio::Tools::BPpslite, Bio::SearchIO or Bio::Tools::BPlite object

Deliverables:

- `convertgb.pl`
- `blastTop2.pl`

Additional instructions:

- You can use base Perl modules (*i.e.* modules that comes with the Perl installation) plus BioPerl modules mentioned above. You should NOT use any additional external Perl modules for this exercises.
- For implementation, your code should take the options below:

```
convertgb.pl –i [input_file] –f [fasta|embl] –o [output_file]
```

```
blastTop2.pl -i [input_file] –d [sequence_db.fa] –m [blast_method] –o
[output_file]
```