

Introduction to Biopython

Alissa Williams

April 12, 2019

What is Biopython?

- Biopython is a set of Python tools written specifically for biological applications.
- The main language for biological applications used to be Perl (BioPerl)
 - Biopython has now caught up to BioPerl functionality
 - Python syntax is more intuitive than Perl syntax

What can we do with Biopython?

- Parse files of various types:
 - fasta/fastq, GenBank, Blast, Clustalw, PubMed, ExPASy, UniGene, SwissProt, etc
 - “Parsing” is breaking up data into useable chunks/objects
- Read in ^those files and print them to other files
- Go through all entries in a file and perform a command on each of them individually
 - Example: reverse complement, get the sequence name associated with a sequence, etc.
- Remember: there are multiple ways to do the same thing!