Biopython: Introduction

Colby Witherup Wood

Northwestern IT Research Computing Services

Request a consultation: *bit.ly/rcs_consult*

```
This workshop will begin at
        1:02 pm Central
```

Biopython is used for working with sequence data in Python

"sequence data" can mean:

DNA

RNA

Protein

Multiple Sequence Alignments

"working with" can mean:

Renaming sequences

Filtering and removing certain sequences from a group of sequences

Combining sequences

Transforming sequences: reverse complement/transcription/translation

Plotting sequence lengths

Plotting GC content

Converting between file formats

BLASTing sequences and parsing the results

Most Biopython tools are available in GUIs

Why should you use Biopython instead?

- Have total customization
- Save a record of what you did in a script
- Automate your analyses – let the computer do the work
- Parallelization – run the same script on multiple genes or files
- Combine your scripts into pipelines

Sample pipeline to build a phylogenetic tree

Sort sequences into gene families (Orthofinder or BLAST with Biopython)

Create fasta file for each gene family (Biopython)

Rename sequences with sample id (Biopython)
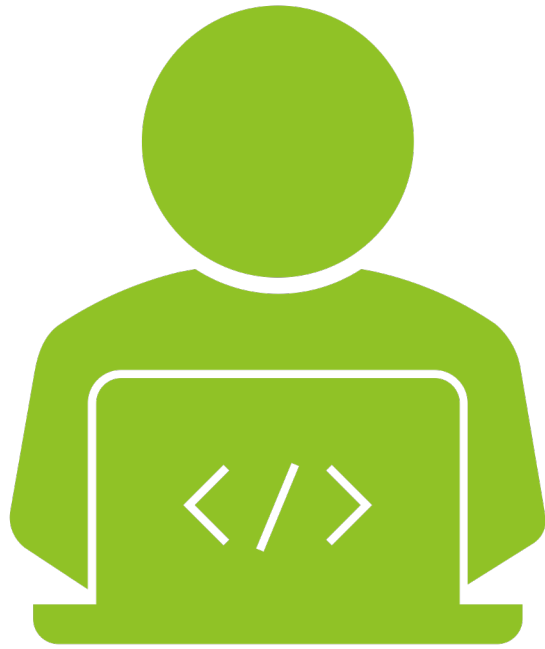
Remove short sequences (Biopython)

Align genes (Mafft)

Trim sequences (TrimAl)

Convert file to phylip format (Biopython)

Make gene trees (RaxML)

Build phylogenetic tree (Astral)

Biopython is suitable for beginning Python programmers

Workshop Materials

Two options for this workshop:

1. On your own computer (must have Biopython installed): Go to github.com/aGitHasNoName/Biopython. Click the green button that says Code. Select Download ZIP. Unzip the folder and move the folder someplace easy to find in your Documents or Desktop. Open up Jupyter Lab (or Jupyter Notebook) from Anaconda (or from the command line). Navigate to the Biopython folder. Open up Biopython.ipynb.

2. To access the notebook online: Go to colab.research.google.com, select GitHub, search for agithasnoname/Biopython. Select Biopython.ipynb.

If option 1 is giving you trouble, please use option 2.

This workshop was inspired by the Biopython Tutorial and Cookbook: biopython.org/DIST/docs/tutorial/Tutorial

Workshop plan

- Please ask questions in the Zoom chat and I will answer during exercises or breaks

- If my internet goes out at any point, that means everyone gets a 10-minute break

- Installation instructions and links to materials are in the email that went out

- If you have any trouble on your own computer, please switch to Google Colab