Navigating the NCBI database

- Here is an exercise in which we learn more about the NCBI database

- Goal 1: track down cytochrome oxidase 1 gene from Aspergillus nidulans; and it should be greater than 450 bp
- Goal 2: Download the 16S rRNA gene for Escherchia Coli and get sequences that are between 1500 and 2500 nucleotides long and save each sequence in a file with the accession number.

Obtaining DNA sequences: search queries

- To design a more sophisticated search query:

www.ncbi.nlm.nih.gov/bookshelf/br.fcgi?book=helpentrez&part=EntrezHelp

- Terms to consider:
  - Boolean operators (*AND*, *OR*, *NOT*)
  - Field specifiers: place these in brackets
    - *ORGN* or *organism*
    - *SLEN* (e.g. “16S AND 200:1650[SLEN]”)
    - *Gene* (e.g. “16S[gene]”)
    - Range (of accession numbers)

Install Biopython AND Install Linux/Bash

http://biopython.org/DIST/docs/install/Installation.html

https://docs.microsoft.com/en-us/windows/wsl/install-win10

http://osxdaily.com/2014/02/12/install-command-line-tools-mac-os-x/

Entrez Direct Documentation

- http://www.ncbi.nlm.nih.gov/books/NBK179288/