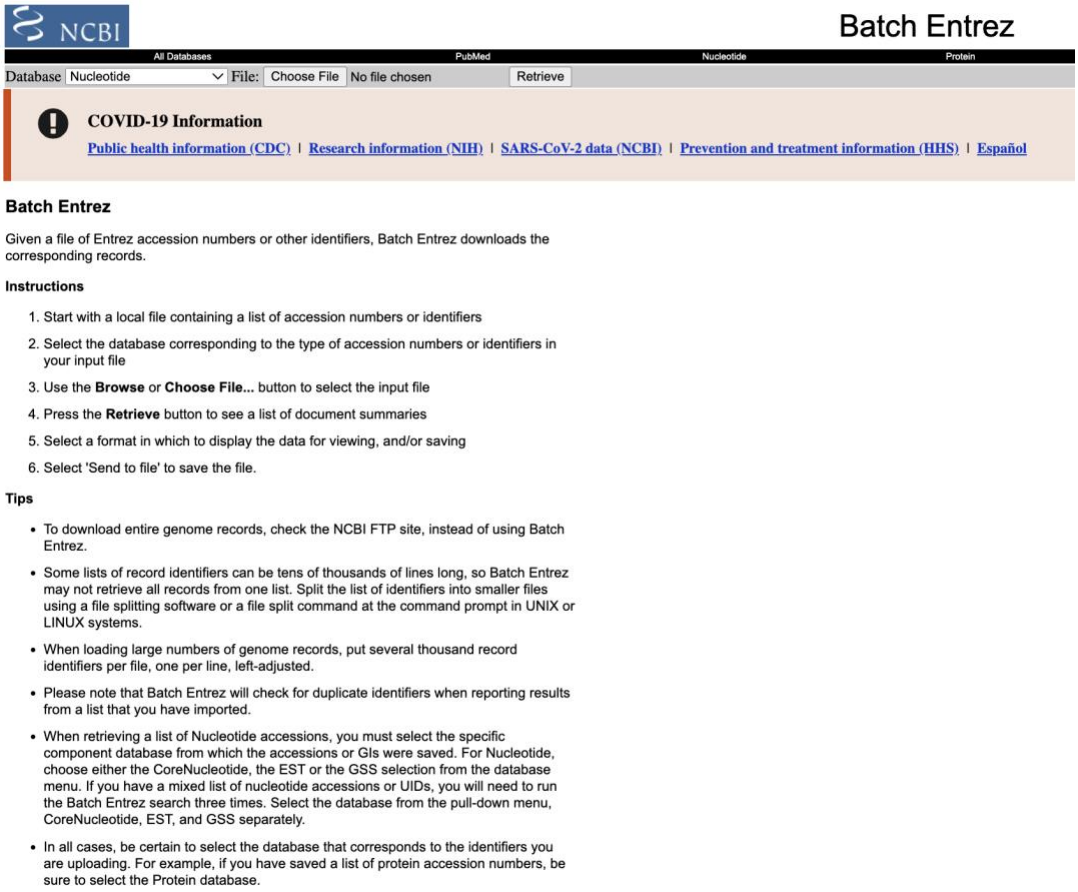
Retrieve transcript sequences from protein IDs on NCBI

1) Generate a plain text file that has one gene protein ID per line.
   Example plain text file:

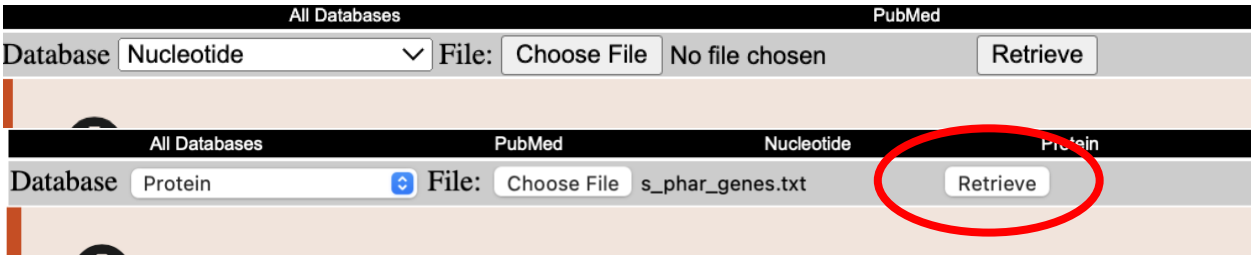
CAE1329857.1
CAE1289477.1
CAE1153723.1
CAE1315817.1
CAE1296720.1
CAE1158496.1
CAE1140664.1
CAE1252791.1
CAE1328585.1
CAE1167180.1
CAE1323142.1
CAE1323437.1

2) Navigate to Batch Entrez on a web browser. https://www.ncbi.nlm.nih.gov/sites/batchentrez

3) You should see a site that looks like the following:



4) In the upper left-hand corner of the site, you will need to change the database from Nucleotide to protein if you have protein IDs. You will also need to supply the plain text file you created in step 1. When you are all done click “Retrieve”.



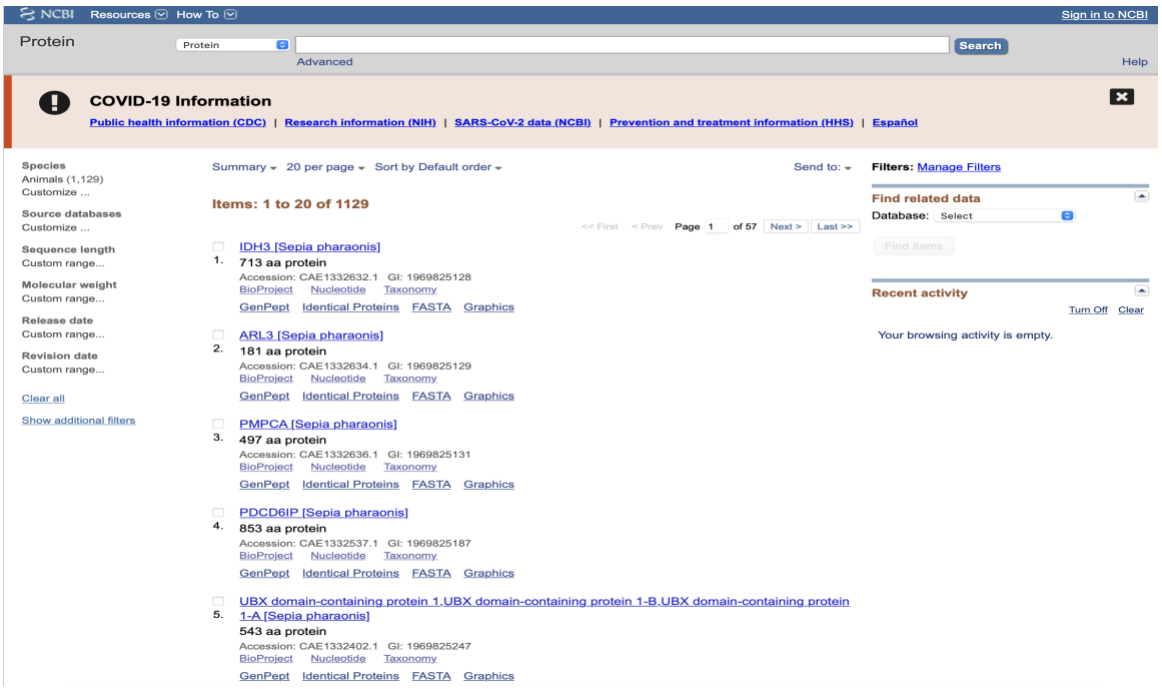
5) Batch Entrez should retrieve the records for each line in your plain text file and return your fetch results that should look like the following:

Received lines: 1129
Rejected lines: 0
Removed duplicates: 0
Passed to Entrez: 1129
Retrieve records for 1129 UID(s)

6) Click on Retrieve records for ## UID(s).

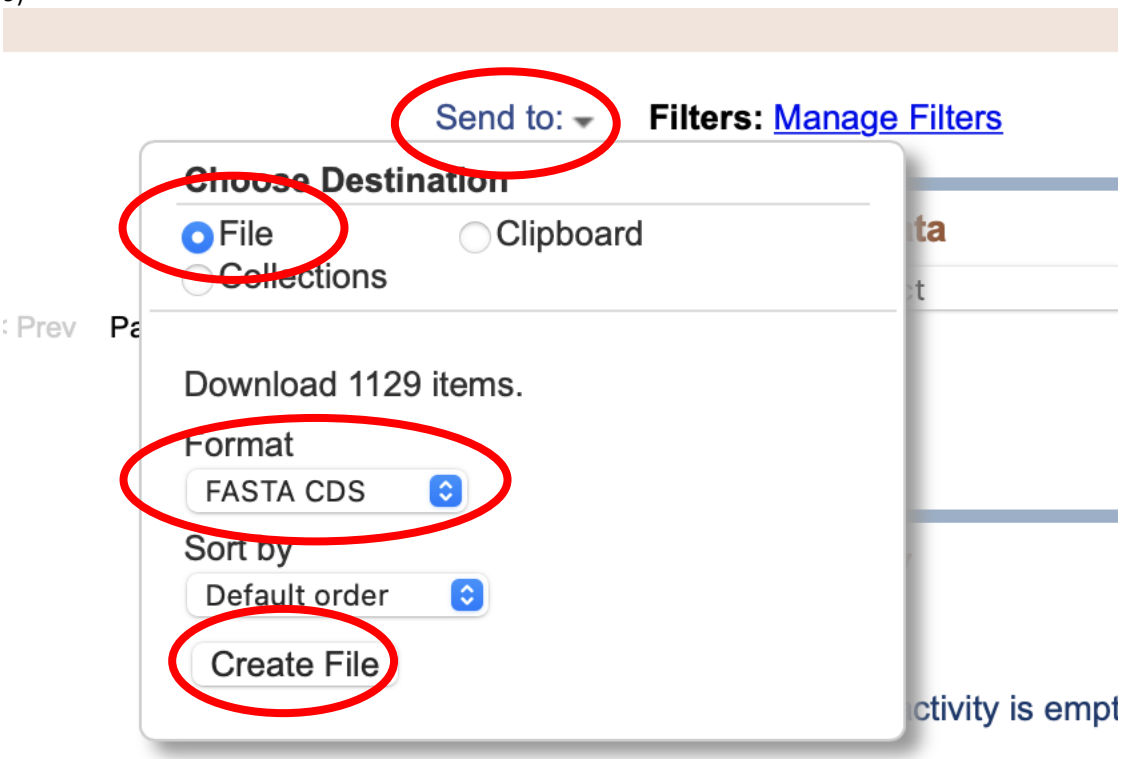
Received lines: 1129
Rejected lines: 0
Removed duplicates: 0
Passed to Entrez: 1129
Retrieve records for 1129 UID(s)

7) This should bring you to a site with your individual records that should look like the following:



8) In the upper middle of the page, click on the drop-down box that says “Send to:”. Within the drop-down box, select “File”. This will populate a new field. When the new field shows, click on the drop-down box under “Format” and select “FASTA CDS”. Finally, click “Create File”

9)



10) A download will initiate that has your transcript gene information.