Trouble Shooting GENI-ACT

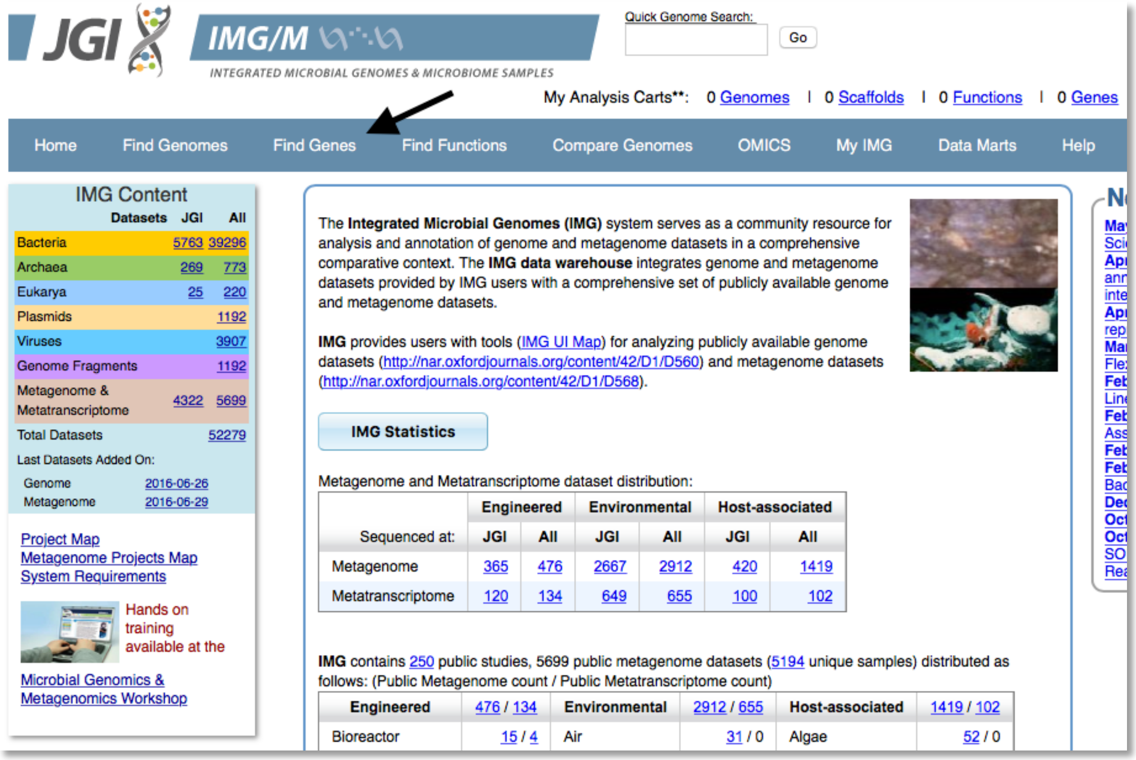
The GENI-ACT site is down. What can I do?

Unfortunately, the GENI-ACT site goes offline at times without warning. We know that you are often restricted in the timing you can work with your students and we have developed the following "work around" to help you salvage your work session with your students.

If you have not done so already, it is a good idea to have your students keep a copy of the FASTA formatted amino acid sequence of the gene on which you are working in a Word or text file saved on their computer or a flash drive. Teachers can access the basic information page of all their student's genes to keep a master list as well. At the very least the locus tags of genes students are working on should be available. The procedure below provides directions for you to retrieve an amino acid sequence on which to work. If you already have your FASTA formatted sequence available to you, skip to step 6 below.

Procedures

1. Log into IMG/edu using the following link:
   http://img.jgi.doe.gov/cgi-bin/edu/main.cgi.
2. Click on the Find Genes tab at the top of the page as indicated by the arrow in Figure 1 below and select the Gene Search option from the pull down menu.



IMG Content

Datasets	JGI	All
Bacteria	5763	39296
Archaea	269	773
Eukarya	25	220
Plasmids		1192
Viruses		3907
Genome Fragments		1192
Metagenome & Metatranscriptome	4322	5699
Total Datasets		52279

Last Datasets Added On:
Genome 2016-06-26
Metagenome 2016-06-29

The **Integrated Microbial Genomes (IMG)** system serves as a community resource for analysis and annotation of genome and metagenome datasets in a comprehensive comparative context. The **IMG data warehouse** integrates genome and metagenome datasets provided by IMG users with a comprehensive set of publicly available genome and metagenome datasets.

IMG provides users with tools (IMG UI Map) for analyzing publicly available genome datasets (http://nar.oxfordjournals.org/content/42/D1/D560) and metagenome datasets (http://nar.oxfordjournals.org/content/42/D1/D568).

Metagenome and Metatranscriptome dataset distribution:

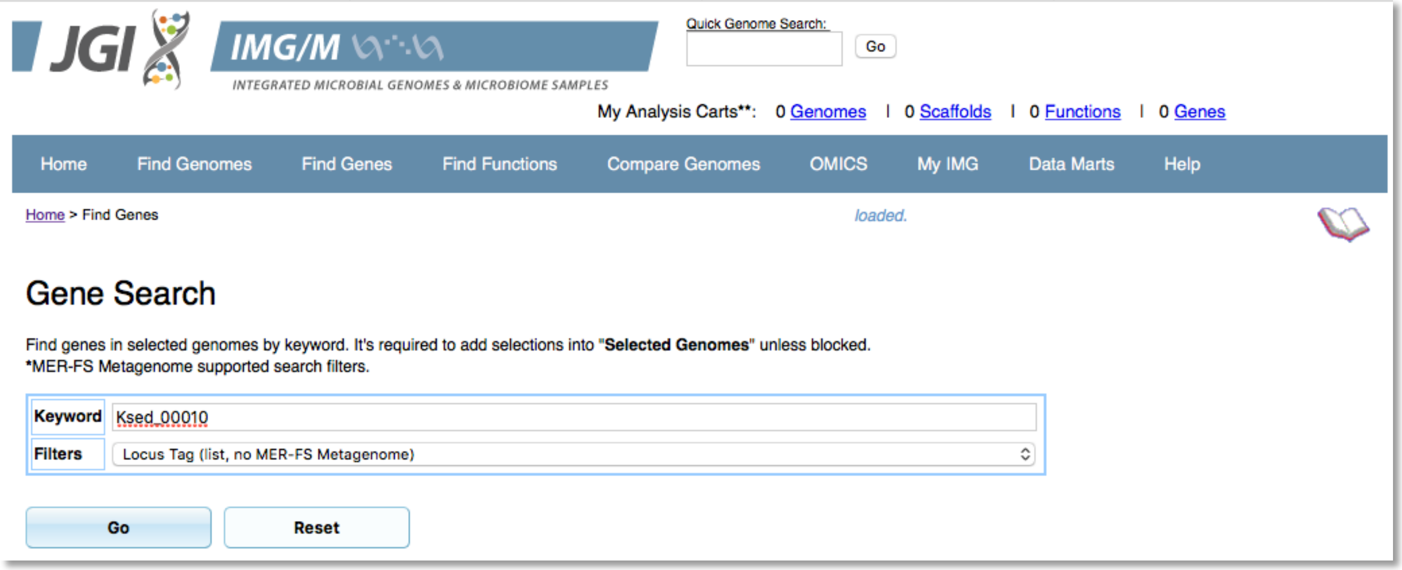
	Engineered		Environmental		Host-associated	
Sequenced at:	JGI	All	JGI	All	JGI	All
Metagenome	365	476	2667	2912	420	1419
Metatranscriptome	120	134	649	655	100	102

IMG contains 250 public studies, 5699 public metagenome datasets (5194 unique samples) distributed as follows: (Public Metagenome count / Public Metatranscriptome count)

Engineered	476 / 134	Environmental	2912 / 655	Host-associated	1419 / 102
Bioreactor	15 / 4	Air	31 / 0	Algae	52 / 0

Figure 1. The IMG/M Home page. The Find Genes tab is indicated by the arrow.

3. In the Gene Search window paste the locus tag for your gene in the keyword box, select Locus Tag (list) from the filters pull down menu (Figure 2) and click Go.
   - 3.1. The locus tags for all of our student's genes are on the set-up document sent to you by way of email when your gene assignments were set up.
   - 3.2. It is a good idea to keep the document at hand when working with your students.
   - 3.3. For those of you with group assignments, you should also keep a master list of which gene locus tag was assigned by you to each student in your group.
   - 3.4. If you have misplaced the document please notify Dr. Koury and he will resend it to you.



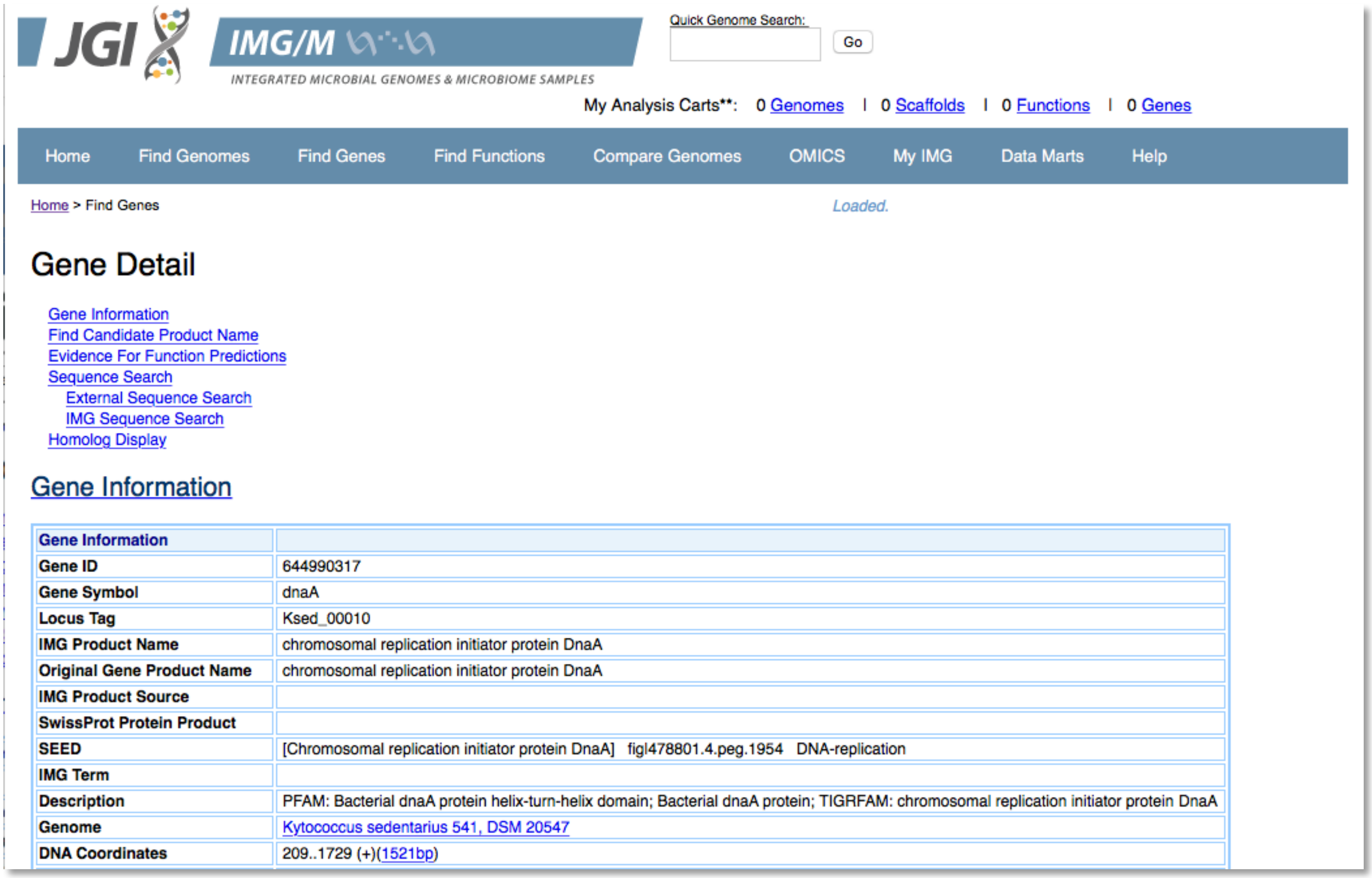
Gene Search

Find genes in selected genomes by keyword. It's required to add selections into "**Selected Genomes**" unless blocked.
*MER-FS Metagenome supported search filters.

Keyword	Ksed_00010
Filters	Locus Tag (list, no MER-FS Metagenome)

Figure 2. The Gene Search window at IMG/M.

4. The page that will appear is called the Gene Detail page in the IMG/M database (Figure 3). It contains a wealth of information known about the gene.



Gene Detail

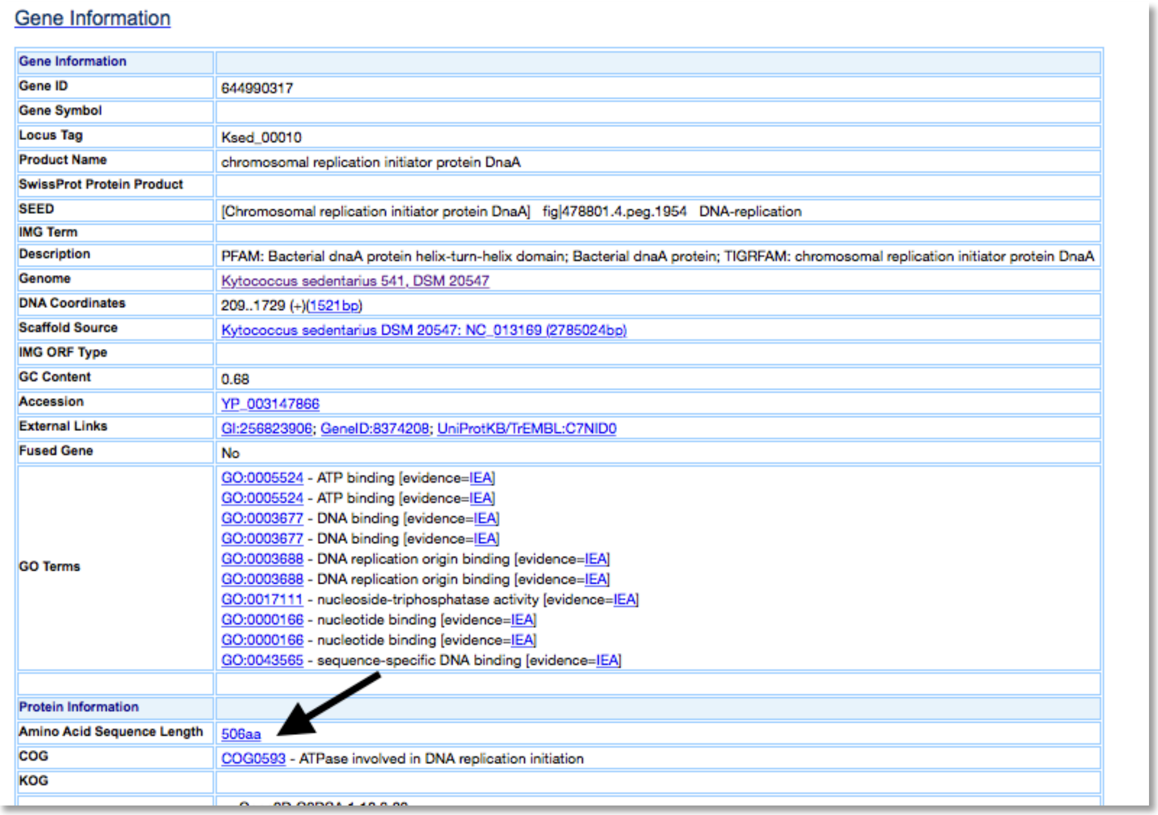
- Gene Information
- Find Candidate Product Name
- Evidence For Function Predictions
- Sequence Search
  - External Sequence Search
  - IMG Sequence Search
- Homolog Display

Gene Information

Gene Information	
Gene ID	644990317
Gene Symbol	dnaA
Locus Tag	Ksed_00010
IMG Product Name	chromosomal replication initiator protein DnaA
Original Gene Product Name	chromosomal replication initiator protein DnaA
IMG Product Source	
SwissProt Protein Product	
SEED	[Chromosomal replication initiator protein DnaA] figl478801.4.peg.1954 DNA-replication
IMG Term	
Description	PFAM: Bacterial dnaA protein helix-turn-helix domain; Bacterial dnaA protein; TIGRFAM: chromosomal replication initiator protein DnaA
Genome	Kytococcus sedentarius 541, DSM 20547
DNA Coordinates	209..1729 (+)(1521bp)

Figure 3. The IMG Gene Detail Page. Only the uppermost portion of the page is shown.

5. Scrolling further down the page you will lead you to a section entitled "amino acid sequence length" with a hyperlink in the form of the number of amino acids in the sequence (arrow, Figure 4).

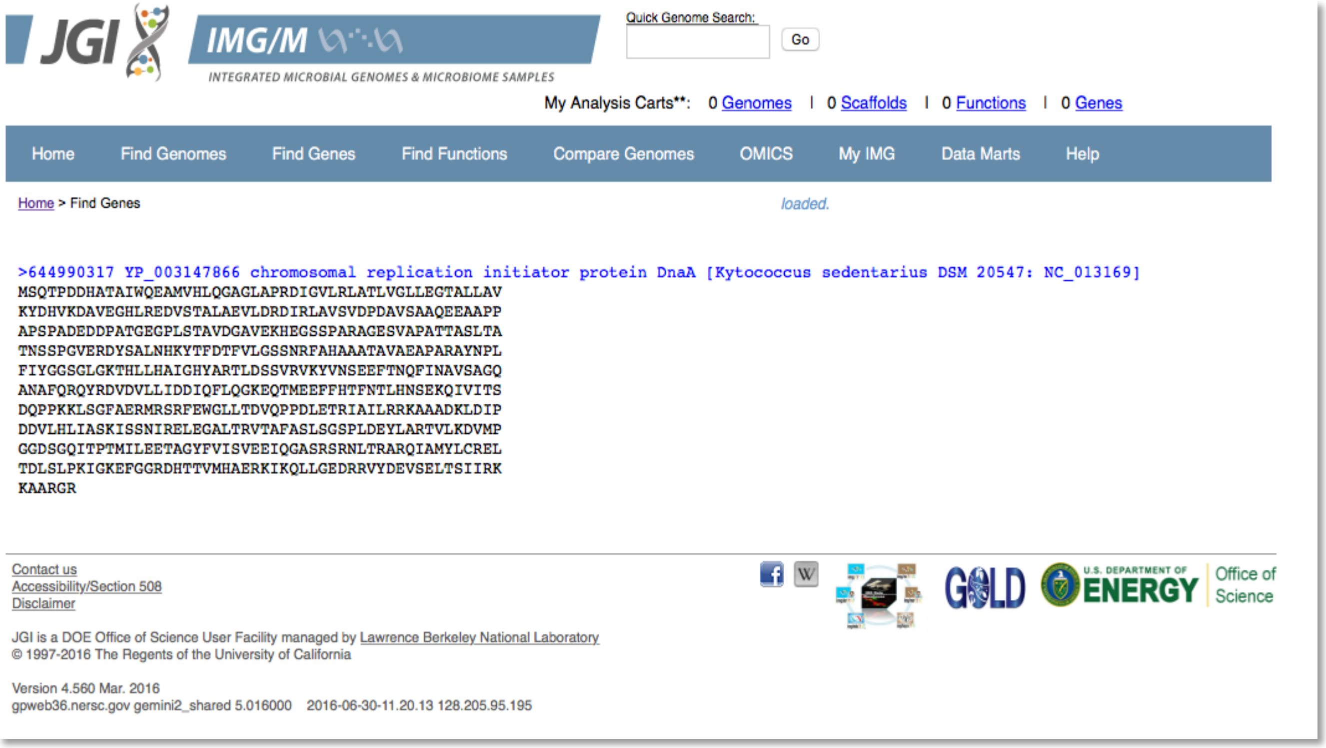


Gene Information

Gene Information	
Gene ID	644990317
Gene Symbol	
Locus Tag	Ksed_00010
Product Name	chromosomal replication initiator protein DnaA
SwissProt Protein Product	
SEED	[Chromosomal replication initiator protein DnaA] fig|478801.4.peg.1954 DNA-replication
IMG Term	
Description	PFAM: Bacterial dnaA protein helix-turn-helix domain; Bacterial dnaA protein; TIGRFAM: chromosomal replication initiator protein DnaA
Genome	Kytococcus sedentarius 541, DSM 20547
DNA Coordinates	209..1729 (+)(1521bp)
Scaffold Source	Kytococcus sedentarius DSM 20547: NC_013169 (2785024bp)
IMG ORF Type	
GC Content	0.68
Accession	YP_003147866
External Links	GI:256823906; GeneID:8374208; UniProtKB/TrEMBL:C7NID0
Fused Gene	No
GO Terms	GO:0005524 - ATP binding [evidence=IEA] GO:0005524 - ATP binding [evidence=IEA] GO:0003677 - DNA binding [evidence=IEA] GO:0003677 - DNA binding [evidence=IEA] GO:0003688 - DNA replication origin binding [evidence=IEA] GO:0003688 - DNA replication origin binding [evidence=IEA] GO:0017111 - nucleoside-triphosphatase activity [evidence=IEA] GO:0000166 - nucleotide binding [evidence=IEA] GO:0000166 - nucleotide binding [evidence=IEA] GO:0043565 - sequence-specific DNA binding [evidence=IEA]
Protein Information	
Amino Acid Sequence Length	506aa
COG	COG0593 - ATPase involved in DNA replication initiation
KOG	

Figure 4. The Gene Detail Page at IMG/edu showing where the amino acid sequence of the protein can be found. The arrow indicates a hyperlink that will lead you to the amino acid sequence.

6. Clicking on the Amino Acid Sequence Length hyperlink indicated by the arrow in figure 4 will open a FASTA formatted sequence, as shown in Figure 5. The sequence plus the FASTA header should be copied and pasted into a Word document or text editor. The FASTA header has a gene identifier number from the IMG database that is different from the locus tag we normally use. For consistency, replace the FASTA header from IMG/edu with the locus tag of your gene as described in the Basic Information Module Instructions and in the caption of figure 5.



```
>644990317 YP_003147866 chromosomal replication initiator protein DnaA [Kytococcus sedentarius DSM 20547: NC_013169]
MSQTPDDHATAIWQEAMVHLQGAGLAPRDIGVLRLATLVGLLEGTALLAV
KYDHVKDAVEGHLREDVSTALAEVLDRDIRLAVSVDPDAVSAAQEEAAPP
APSPADEDDPATGEGPLSTAVDGAVEKHEGSSPARAGESVAPATTASLTA
TNSSPGVERDYSALNHKYTFDTFVLGSSNRFAHAAATAVAEAPARAYNPL
FIYGGSGLGKTHLLHAIGHYARTLDSSVRVKYVNSEEFTNQFINAVSAGQ
ANAFQRQYRDVDVLLIDDIQFLQGKEQTMEEFFHTFNTLHNSEKQIVITS
DQPPKKLSGFAERMRSRFEWGLLTDVQPPDLETRIAILRRKAAADKLDIP
DDVLHLIASKISSNIRELEGALTRVTAFASLSGSPLDEYLARTVLKDVMP
GGDSGQITPTMILEETAGYFVISVEEIQGASRSRNLTRARQIAMYLCREL
TDLSLPKIGKEFGGRDHTTVMHAERKIKQLLGEDRRVYDEVSELTSIIRK
KAARGR
```

Contact us
Accessibility/Section 508
Disclaimer

JGI is a DOE Office of Science User Facility managed by Lawrence Berkeley National Laboratory
© 1997-2016 The Regents of the University of California

Version 4.560 Mar. 2016
gpweb36.nersc.gov gemini2_shared 5.016000 2016-06-30-11.20.13 128.205.95.195

Figure 5. The FASTA formatted sequence of Ksed_00010 as obtained from the IMG gene details page. The FASTA header in blue in the image above should be replace with the locus tag of your gene. In this case the edited header would read >Ksed_00010 amino acid sequence, as is described in the Basic Information Module instructions.

7. Once you have the FASTA formatted sequence of your protein, from either a saved file or from IMG/edu as described above, you can use the online module instructions to perform the annotation exercises.
8. You can access the online module instructions at:

   http://ubwp.buffalo.edu/wnygirahcp/educational-resources/

9. Select the instructions for whichever module on which you are planning to work to see what happens in that module. Your students will then be able to find the bioinformatics tools on the Web and know how to use them. The module instructions will tell your students what data to save.
10. Your students can save their work in a Word or PowerPoint document and then upload it to the GENI-ACT notebook when it becomes available. Below is a link to a webpage that will have the Word document (blank_GENIACT_notebook_template.doc) that has all the GENI-ACT notebook sections in it, active links to the GENI-ACT tools and active links to instructions for each module.

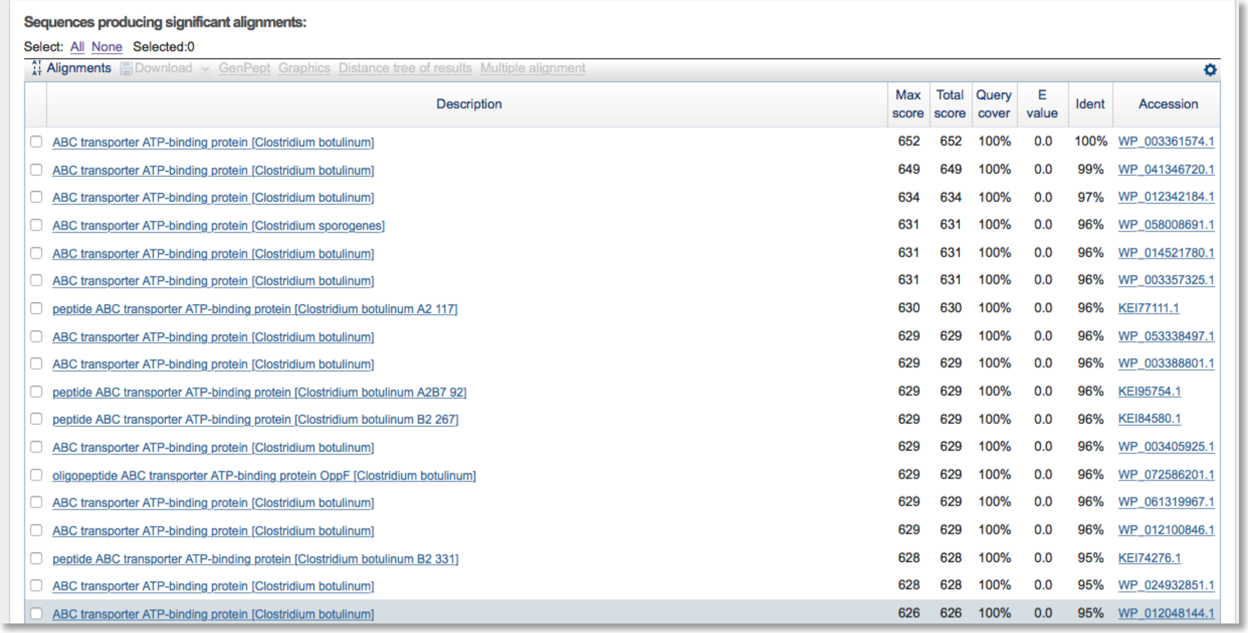
    The link is:

    https://sites.google.com/site/ubbclsgenomeannotation/educational-resources

What do I do if my blast searches seem to have all the top hits from the same genus or species?

If the bacterial species you are using to annotate is clinically significant or of great research interest, you may find that when you perform blast searches (particularly in nr) that you seemingly only get hits that are different strains or isolates of the same species. This obviously doesn't give you much information about how well conserved the protein on which you are working is compared to proteins in other genera. There is a method to modify blast to let you exclude such hits from your searches.

A gene from *Clostridium botulinum* will be used as an example to illustrate this process. Figure 1 shows the top nr blast hits for the protein sequence of the gene with the locus tag CLJ_B3418. You can easily see that all of the hits but one are from *Clostridium botulinum* with very high levels of coverage and identities. They are essentially all the same protein from different isolates of *Clostridium botulinum*.



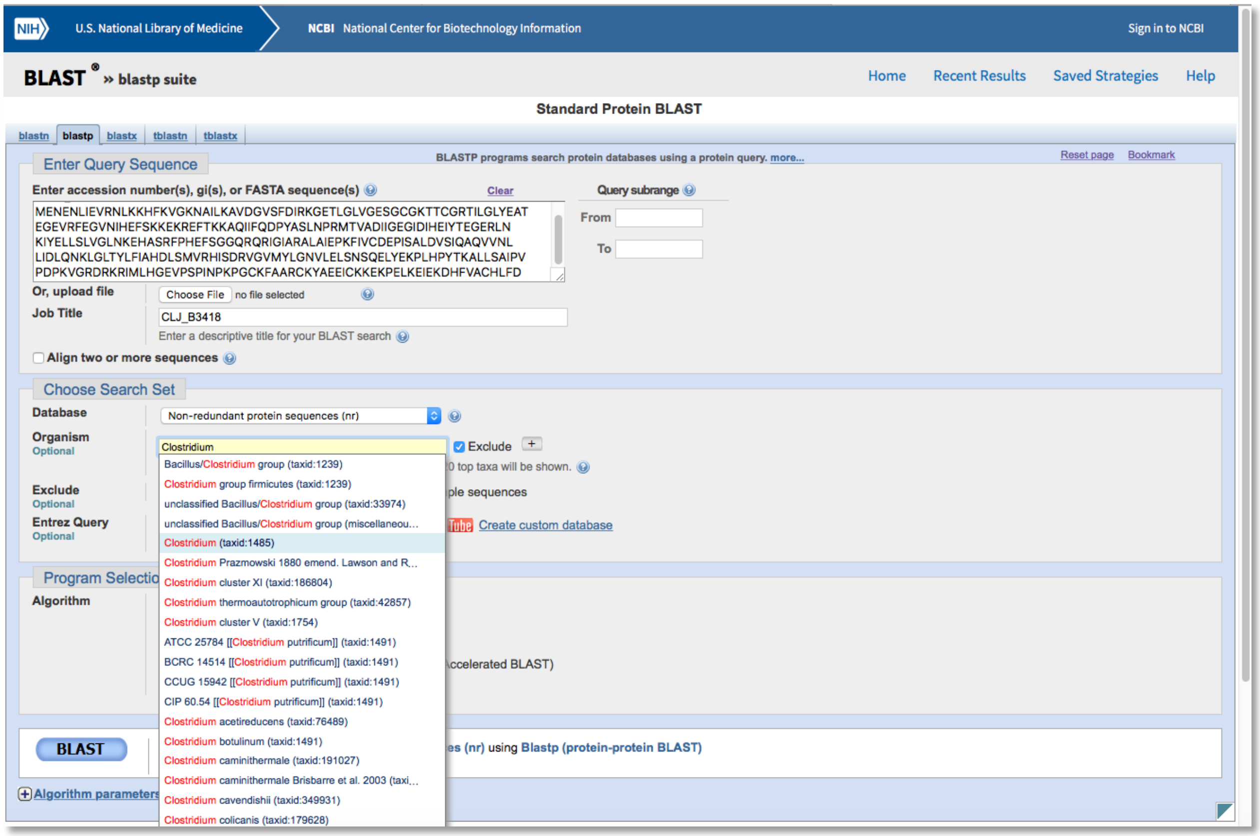
Sequences producing significant alignments:

Select: All None Selected:0

	Description	Max score	Total score	Query cover	E value	Ident	Accession
☐	ABC transporter ATP-binding protein [Clostridium botulinum]	652	652	100%	0.0	100%	WP_003361574.1
☐	ABC transporter ATP-binding protein [Clostridium botulinum]	649	649	100%	0.0	99%	WP_041346720.1
☐	ABC transporter ATP-binding protein [Clostridium botulinum]	634	634	100%	0.0	97%	WP_012342184.1
☐	ABC transporter ATP-binding protein [Clostridium sporogenes]	631	631	100%	0.0	96%	WP_058008691.1
☐	ABC transporter ATP-binding protein [Clostridium botulinum]	631	631	100%	0.0	96%	WP_014521780.1
☐	ABC transporter ATP-binding protein [Clostridium botulinum]	631	631	100%	0.0	96%	WP_003357325.1
☐	peptide ABC transporter ATP-binding protein [Clostridium botulinum A2 117]	630	630	100%	0.0	96%	KEI77111.1
☐	ABC transporter ATP-binding protein [Clostridium botulinum]	629	629	100%	0.0	96%	WP_053338497.1
☐	ABC transporter ATP-binding protein [Clostridium botulinum]	629	629	100%	0.0	96%	WP_003388801.1
☐	peptide ABC transporter ATP-binding protein [Clostridium botulinum A2B7 92]	629	629	100%	0.0	96%	KEI95754.1
☐	peptide ABC transporter ATP-binding protein [Clostridium botulinum B2 267]	629	629	100%	0.0	96%	KEI84580.1
☐	ABC transporter ATP-binding protein [Clostridium botulinum]	629	629	100%	0.0	96%	WP_003405925.1
☐	oligopeptide ABC transporter ATP-binding protein OppF [Clostridium botulinum]	629	629	100%	0.0	96%	WP_072586201.1
☐	ABC transporter ATP-binding protein [Clostridium botulinum]	629	629	100%	0.0	96%	WP_061319967.1
☐	ABC transporter ATP-binding protein [Clostridium botulinum]	629	629	100%	0.0	96%	WP_012100846.1
☐	peptide ABC transporter ATP-binding protein [Clostridium botulinum B2 331]	628	628	100%	0.0	95%	KEI74276.1
☐	ABC transporter ATP-binding protein [Clostridium botulinum]	628	628	100%	0.0	95%	WP_024932851.1
☐	ABC transporter ATP-binding protein [Clostridium botulinum]	626	626	100%	0.0	95%	WP_012048144.1

Figure 1. The blast results using a non-filtered nr blast search for CLJ_B3418.

The blast search can be set up slightly differently to prevent this problem from occurring. As noted in figure 2, we can set the search up to exclude, in this case, the taxid: 1485 (Clostridium). The taxid number stands for the NCBI Taxonomy ID number. By excluding the taxid number 1485, all blast hits in that taxonomic classification will not be included. To do this we type the genus name Clostridium in the Organism textbox below the sequence input box. As you type a pulldown menu of options will appear which you can subsequently just click on to select (see highlighted menu item in Figure 2). The simply click the Exclude checkbox next to the organism name and then select blast.



Enter accession number(s), gi(s), or FASTA sequence(s)

```
MENENLIEVRNLKKHFKVGKNAILKAVDGVSFDIRKGETLGLVGESGCGKTTCGRTILGLYEAT
EGEVRFEGVNIHEFSKKEKREFTKKAQIIFQDPYASLNPRMTVADIIGEGIDIHEIYTEGERLN
KIYELLSLVGLNKEHASRFPHEFSGGQRQRIGIARALAIEPKFIVCDEPISALDVSIQAQVVNL
LIDLQNKLGLTYLFIAHDLSMVRHISDRVGVMYLGNVLELSNSQELYEKPLHPYTKALLSAIPV
PDPKVGRDRKRIMLHGEVPSPINPKPGCKFAARCKYAEEICKKEKPELKEIEKDHFVACHLFD
```

From

To

Or, upload file: no file selected

Job Title: CLJ_B3418

Enter a descriptive title for your BLAST search

☐ Align two or more sequences

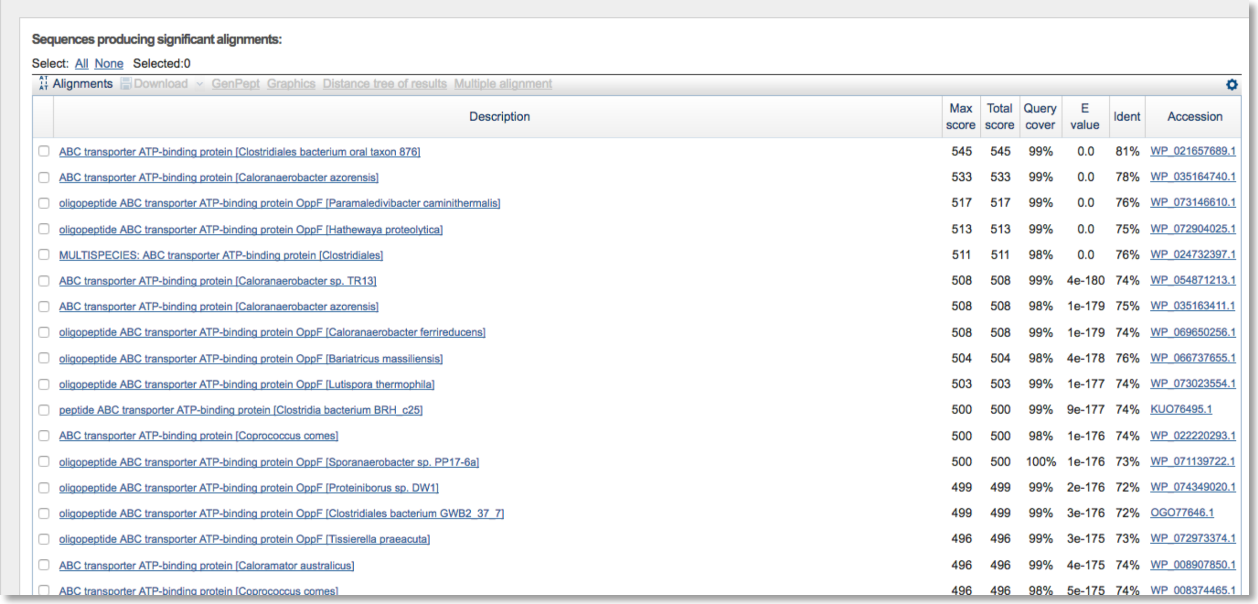
Database: Non-redundant protein sequences (nr)

Organism
Optional: Clostridium ☑ Exclude +

Figure 2. Setting up a blast search to exclude Costridium taxid: 1485.

Figure 3 shows the results of the blast result for the same protein AFTER excluding the Clostridium taxid 1485. Note the different names appearing in the search results. However, also note that the top hit is no longer the one that matches the protein under investigation in the species you are working on. **Thus, you would take the FIRST nr hit as the top hit in this case instead of skipping over the first one as was described in the instructions for module 2. Note also that if you use this blast result to select sequences for the T-Coffee alignment that you will subsequently do in the Sequence Based Similarity Module, that you will need to add the FASTA formatted sequence of the protein under investigation to the top of the list before constructing the alignment.** Students should also add a comment in their notebook of which taxid number was excluded from their search when recording their results.

Experiment with different levels of exclusion (only one species) or add multiple options for exclusion (i.e., the genus) or somewhere in between (different specific species excluded by adding additional organism boxes in which to enter choices by using the + option to the right of the exclude check box to add another).



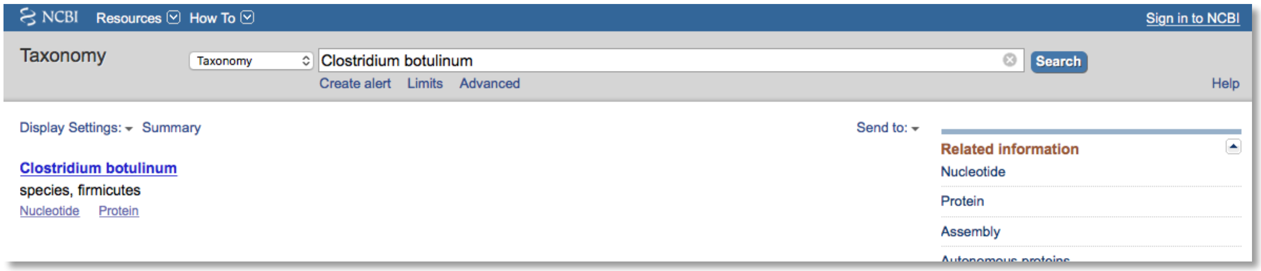
Sequences producing significant alignments:

Select: All None Selected:0

	Description	Max score	Total score	Query cover	E value	Ident	Accession
☐	ABC transporter ATP-binding protein [Clostridiales bacterium oral taxon 876]	545	545	99%	0.0	81%	WP_021657689.1
☐	ABC transporter ATP-binding protein [Caloranaerobacter azorensis]	533	533	99%	0.0	78%	WP_035164740.1
☐	oligopeptide ABC transporter ATP-binding protein OppF [Paramaledivibacter caminithermalis]	517	517	99%	0.0	76%	WP_073146610.1
☐	oligopeptide ABC transporter ATP-binding protein OppF [Hathewaya proteolytica]	513	513	99%	0.0	75%	WP_072904025.1
☐	MULTISPECIES: ABC transporter ATP-binding protein [Clostridiales]	511	511	98%	0.0	76%	WP_024732397.1
☐	ABC transporter ATP-binding protein [Caloranaerobacter sp. TR13]	508	508	99%	4e-180	74%	WP_054871213.1
☐	ABC transporter ATP-binding protein [Caloranaerobacter azorensis]	508	508	98%	1e-179	75%	WP_035163411.1
☐	oligopeptide ABC transporter ATP-binding protein OppF [Caloranaerobacter ferrireducens]	508	508	99%	1e-179	74%	WP_069650256.1
☐	oligopeptide ABC transporter ATP-binding protein OppF [Bariatricus massiliensis]	504	504	98%	4e-178	76%	WP_066737655.1
☐	oligopeptide ABC transporter ATP-binding protein OppF [Lutispora thermophila]	503	503	99%	1e-177	74%	WP_073023554.1
☐	peptide ABC transporter ATP-binding protein [Clostridia bacterium BRH_c25]	500	500	99%	9e-177	74%	KUO76495.1
☐	ABC transporter ATP-binding protein [Coprococcus comes]	500	500	98%	1e-176	74%	WP_022220293.1
☐	oligopeptide ABC transporter ATP-binding protein OppF [Sporanaerobacter sp. PP17-6a]	500	500	100%	1e-176	73%	WP_071139722.1
☐	oligopeptide ABC transporter ATP-binding protein OppF [Proteiniborus sp. DW1]	499	499	99%	2e-176	72%	WP_074349020.1
☐	oligopeptide ABC transporter ATP-binding protein OppF [Clostridiales bacterium GWB2_37_7]	499	499	99%	3e-176	72%	OGO77646.1
☐	oligopeptide ABC transporter ATP-binding protein OppF [Tissierella praeacuta]	496	496	99%	3e-175	73%	WP_072973374.1
☐	ABC transporter ATP-binding protein [Caloramator australicus]	496	496	99%	4e-175	74%	WP_008907850.1
☐	ABC transporter ATP-binding protein [Coprococcus comes]	496	496	98%	5e-175	74%	WP_008374465.1

Figure 3. The nr blast results for CLJ_B3418 AFTER excluding the Costridium taxid: 1485. Note the different genus and species names of the top hits.

You can also use the NCBI Taxonomy Browser to find different levels of taxid to use in your exclusion searches, especially if it is not clear what you should choose from the pulldown menu in BLAST. The use of the Taxonomy Browser is described in general terms in the Horizontal Gene Transfer section of the project manual. Briefly, go to: https://www.ncbi.nlm.nih.gov/taxonomy and enter the name of your organism's genus in the search window (Clostridium botulinum is the example used below) and click on Search. A result similar to Figure 4 will display. Click on the organism hyperlink in blue, and you will be taken to the full lineage of the organism (next figure).



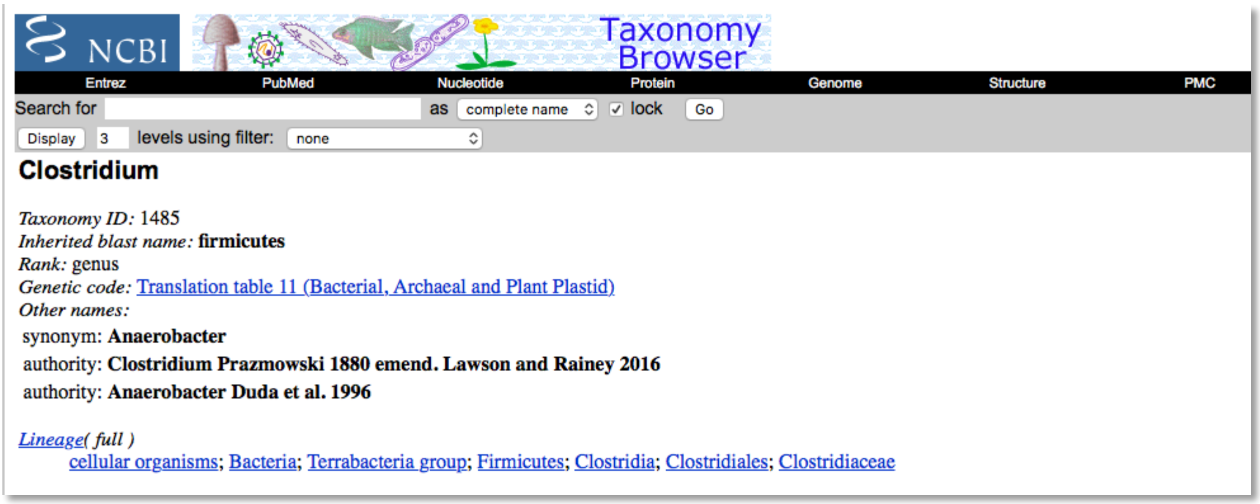
Taxonomy: Clostridium botulinum

Display Settings: Summary

Clostridium botulinum
species, firmicutes
Nucleotide Protein

Figure 4. Results of searching for Clostridium botulinum in the NCBI Taxonomy browser.

Figure 5 below shows a portion of the C. botulinum results. In the lineage line, the last entry is the genus (Clostridum), but you can hover the cursor over any of the levels of taxonomy and see the name of the level (i.e., family, order etc.).

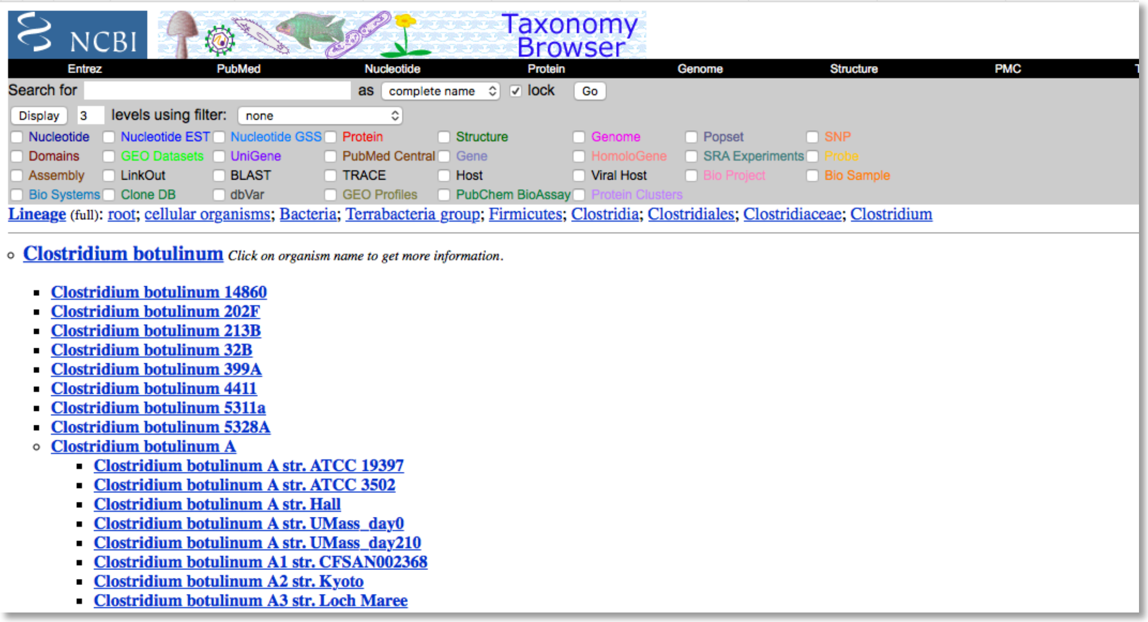


Clostridium

Taxonomy ID: 1485
Inherited blast name: **firmicutes**
Rank: genus
Genetic code: Translation table 11 (Bacterial, Archaeal and Plant Plastid)
Other names:
synonym: **Anaerobacter**
authority: **Clostridium Prazmowski 1880 emend. Lawson and Rainey 2016**
authority: **Anaerobacter Duda et al. 1996**

Lineage(full)
cellular organisms; Bacteria; Terrabacteria group; Firmicutes; Clostridia; Clostridiales; Clostridiaceae

Figure 6 shows what will display when the Clostridium hyperlink is selected.



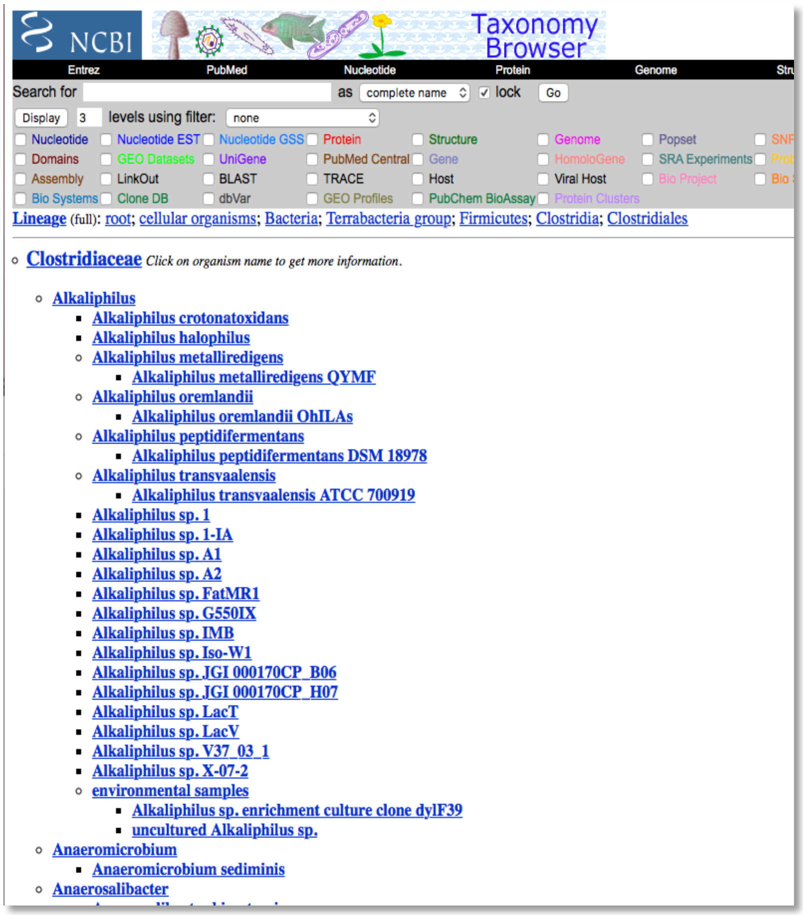
Lineage (full): root; cellular organisms; Bacteria; Terrabacteria group; Firmicutes; Clostridia; Clostridiales; Clostridiaceae; Clostridium

- **Clostridium botulinum** *Click on organism name to get more information.*
  - **Clostridium botulinum 14860**
  - **Clostridium botulinum 202F**
  - **Clostridium botulinum 213B**
  - **Clostridium botulinum 32B**
  - **Clostridium botulinum 399A**
  - **Clostridium botulinum 4411**
  - **Clostridium botulinum 5311a**
  - **Clostridium botulinum 5328A**
  - **Clostridium botulinum A**
    - **Clostridium botulinum A str. ATCC 19397**
    - **Clostridium botulinum A str. ATCC 3502**
    - **Clostridium botulinum A str. Hall**
    - **Clostridium botulinum A str. UMass_day0**
    - **Clostridium botulinum A str. UMass_day210**
    - **Clostridium botulinum A1 str. CFSAN002368**
    - **Clostridium botulinum A2 str. Kyoto**
    - **Clostridium botulinum A3 str. Loch Maree**

Figure 6. The *Clostridium botulinum* results from NCBI Taxonomy browser (not complete).

Of interest in the results from clicking on the Clostridium hyperlink displayed displayed in Figure 6 is the Taxonomy ID of 1485 (exactly the one we found by limiting the BLAST results from within the BLAST tool). You could use this information to simply type "Clostridium (taxid:1485)" – do not, however, include the quotation marks- in the organism window of the BLAST search and click exclude as before.

We can also go further "up" in taxonomic window to exclude more than one genus though you should not have to do that routinely). For example, Figure 7 shows the display that would come up if we clicked on the *Clostridiaceae* (i.e., the family to which the genus Clostridum belongs) hyperlink instead of the Clostridium hyperlink. Different genera will appear that are part of this family. Clicking on the Clostridiaceae link from this page will result in the information shown in the Figure 8.



Lineage (full): root; cellular organisms; Bacteria; Terrabacteria group; Firmicutes; Clostridia; Clostridiales

- **Clostridiaceae** *Click on organism name to get more information.*
  - **Alkaliphilus**
    - **Alkaliphilus crotonatoxidans**
    - **Alkaliphilus halophilus**
    - **Alkaliphilus metalliredigens**
      - **Alkaliphilus metalliredigens QYMF**
    - **Alkaliphilus oremlandii**
      - **Alkaliphilus oremlandii OhILAs**
    - **Alkaliphilus peptidifermentans**
      - **Alkaliphilus peptidifermentans DSM 18978**
    - **Alkaliphilus transvaalensis**
      - **Alkaliphilus transvaalensis ATCC 700919**
    - **Alkaliphilus sp. 1**
    - **Alkaliphilus sp. 1-IA**
    - **Alkaliphilus sp. A1**
    - **Alkaliphilus sp. A2**
    - **Alkaliphilus sp. FatMR1**
    - **Alkaliphilus sp. G550IX**
    - **Alkaliphilus sp. IMB**
    - **Alkaliphilus sp. Iso-W1**
    - **Alkaliphilus sp. JGI 000170CP_B06**
    - **Alkaliphilus sp. JGI 000170CP_H07**
    - **Alkaliphilus sp. LacT**
    - **Alkaliphilus sp. LacV**
    - **Alkaliphilus sp. V37_03_1**
    - **Alkaliphilus sp. X-07-2**
    - **environmental samples**
      - **Alkaliphilus sp. enrichment culture clone dylF39**
      - **uncultured Alkaliphilus sp.**
  - **Anaeromicrobium**
    - **Anaeromicrobium sediminis**
  - **Anaerosalibacter**

Figure 7. The display resulting from selection of the Family Clostridiaceae in the taxonomy browser.

Display 3 levels using filter: none

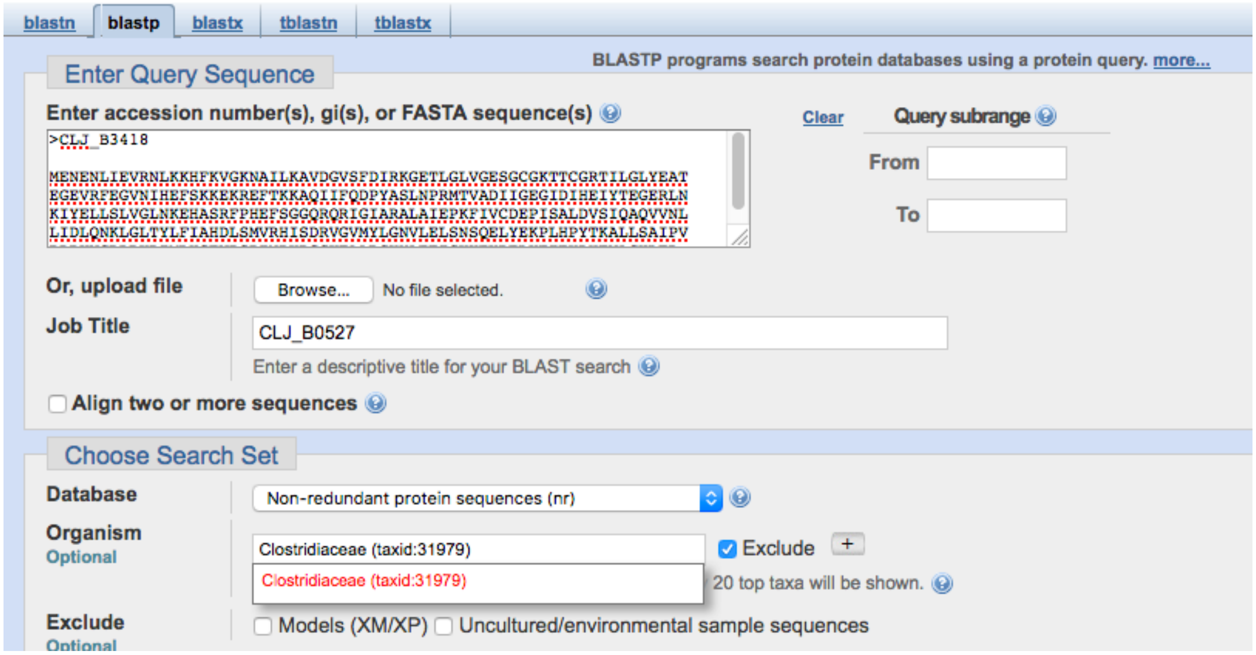
Clostridiaceae

Taxonomy ID: 31979
Inherited blast name: **firmicutes**
Rank: family
Genetic code: Translation table 11 (Bacterial, Archaeal and Plant Plastid)
Other names:
authority: **Clostridiaceae Pribram 1933**

Lineage(full)
cellular organisms; Bacteria; Terrabacteria group; Firmicutes; Clostridia; Clostridiales

Figure 8. The taxonomy identification number of the Family Clostridiaceae.

Here we see that the Clostridiaceae family has the Taxonomy ID of 31979. To exclude this Family from the BLAST results, we would simply type in "Clostridiaceae (taxid:31979)" into the organism box in the BLAST search and click exclude. The next image will show how the autofill option will highlight once we paste in the taxid.



```
>CLJ_B3418

MENENLIEVRNLKKHFKVGKNAILKAVDGVSFDIRKGETLGLVGESGCGKTTCGRTILGLYEAT
EGEVRFEGVNIHEFSKKEKREFTKKAQIIFQDPYASLNPRMTVADIIGEGIDIHEIYTEGERLN
KIYELLSLVGLNKEHASRFPHEFSGGQRQRIGIARALAIEPKFIVCDEPISALDVSIQAQVVNL
LIDLQNKLGLTYLFIAHDLSMVRHISDRVGVMYLGNVLELSNSQELYEKPLHPYTKALLSAIPV
```

Or, upload file: No file selected.

Job Title: CLJ_B0527

Enter a descriptive title for your BLAST search

Database: Non-redundant protein sequences (nr)

Organism: Clostridiaceae (taxid:31979) ☑ Exclude +

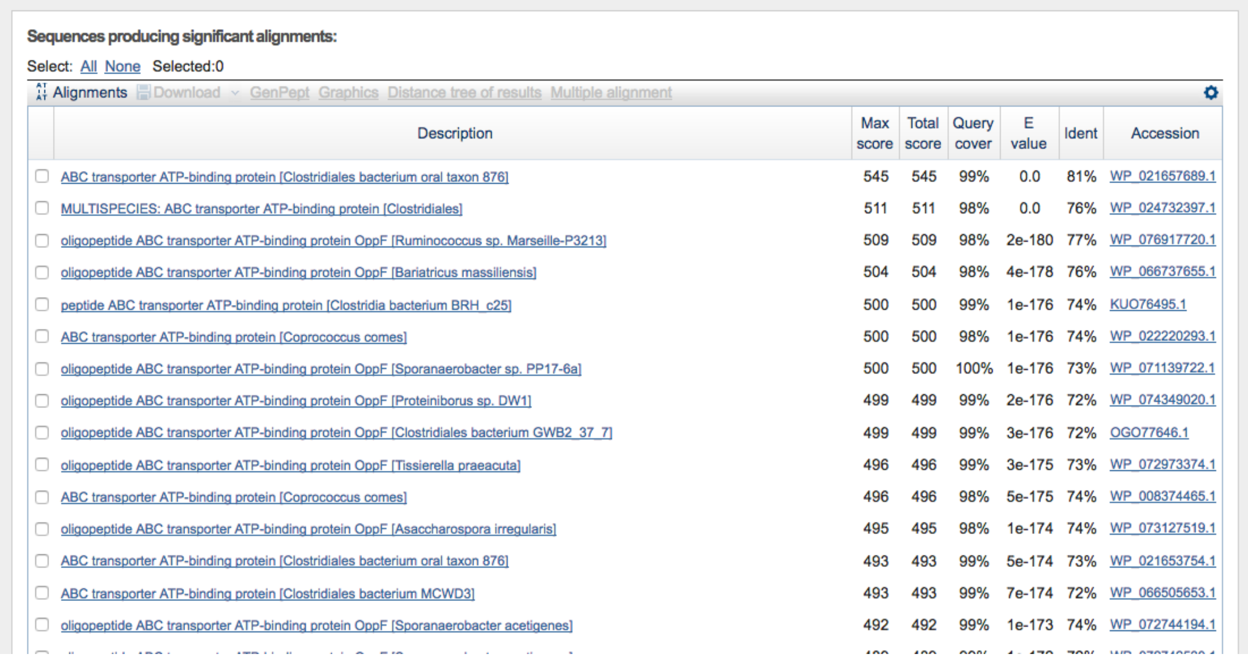
Clostridiaceae (taxid:31979)

20 top taxa will be shown.

Exclude: ☐ Models (XM/XP) ☐ Uncultured/environmental sample sequences

Figure 9. A BLAST search set up to exclude members of the Family Clostridiaceae from the search results.

Finally, Figure 10 shows the BLAST results from doing the exclusion at this level.



Sequences producing significant alignments:

Select: All None Selected:0

	Description	Max score	Total score	Query cover	E value	Ident	Accession
☐	ABC transporter ATP-binding protein [Clostridiales bacterium oral taxon 876]	545	545	99%	0.0	81%	WP_021657689.1
☐	MULTISPECIES: ABC transporter ATP-binding protein [Clostridiales]	511	511	98%	0.0	76%	WP_024732397.1
☐	oligopeptide ABC transporter ATP-binding protein OppF [Ruminococcus sp. Marseille-P3213]	509	509	98%	2e-180	77%	WP_076917720.1
☐	oligopeptide ABC transporter ATP-binding protein OppF [Bariatricus massiliensis]	504	504	98%	4e-178	76%	WP_066737655.1
☐	peptide ABC transporter ATP-binding protein [Clostridia bacterium BRH_c25]	500	500	99%	1e-176	74%	KUO76495.1
☐	ABC transporter ATP-binding protein [Coprococcus comes]	500	500	98%	1e-176	74%	WP_022220293.1
☐	oligopeptide ABC transporter ATP-binding protein OppF [Sporanaerobacter sp. PP17-6a]	500	500	100%	1e-176	73%	WP_071139722.1
☐	oligopeptide ABC transporter ATP-binding protein OppF [Proteiniborus sp. DW1]	499	499	99%	2e-176	72%	WP_074349020.1
☐	oligopeptide ABC transporter ATP-binding protein OppF [Clostridiales bacterium GWB2_37_7]	499	499	99%	3e-176	72%	OGO77646.1
☐	oligopeptide ABC transporter ATP-binding protein OppF [Tissierella praeacuta]	496	496	99%	3e-175	73%	WP_072973374.1
☐	ABC transporter ATP-binding protein [Coprococcus comes]	496	496	98%	5e-175	74%	WP_008374465.1
☐	oligopeptide ABC transporter ATP-binding protein OppF [Asaccharospora irregularis]	495	495	98%	1e-174	74%	WP_073127519.1
☐	ABC transporter ATP-binding protein [Clostridiales bacterium oral taxon 876]	493	493	99%	5e-174	73%	WP_021653754.1
☐	ABC transporter ATP-binding protein [Clostridiales bacterium MCWD3]	493	493	99%	7e-174	72%	WP_066505653.1
☐	oligopeptide ABC transporter ATP-binding protein OppF [Sporanaerobacter acetigenes]	492	492	99%	1e-173	74%	WP_072744194.1

Figure 10. BLAST results after excluding the Family Clostridiaceae from the search.