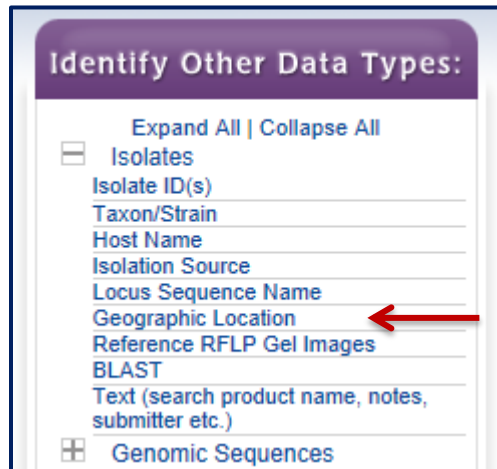


Exploring Isolate Data Exercise 8

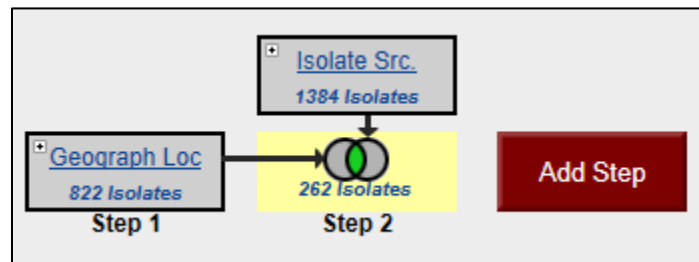
8.1 Exploring isolates in *Cryptosporidium* and using the alignment tool.

Note: For this exercise use <http://www.cryptodb.org>

- a. Identify all *Cryptosporidium* isolates from Europe.
Hint: search for isolates by geographic location in the “Identify Other Data Types” section.



- b. How many of the *Cryptosporidium* isolates collected in Europe were isolated from feces?
Hint: add another isolate search step.



- c. What is the general distribution of these isolates in Europe? (hint: you can do this quickly in two ways: sort the geographic location column by clicking on the sort arrows, then look at the represented countries; or use the “Isolate Geographical Location” tab to view a map and results summary table).

Sort by clicking on the arrows

Isolate ID	Geographic Location	Organism	Strain/Isolate Name	Host	Isolation
AB247224	Serbia	Cryptosporidium parvum	#6	Unknown	fecal sample
AB247225	Serbia	Cryptosporidium parvum	#24	Unknown	fecal sample
AB247226	Serbia	Cryptosporidium parvum	#42	Unknown	fecal sample
AB247227	Serbia				

Country	Number of Isolates	Isolate Type	Latitude	Longitude
Belgium	1	Sequencing Typed	50.503887	4.469936
Czech Republic	73	Sequencing Typed	49.817492	15.472962
Germany	90	Sequencing Typed	51.165691	10.451526
Ireland	4	Sequencing Typed	53.41291	-8.24389
Italy	11	Sequencing Typed	41.87194	12.56738
Lithuania	1	Sequencing Typed	55.169438	23.881275
Netherlands	41	Sequencing Typed	52.132633	5.291266

- d. Out of those in step ‘b’, how many are unclassified *Cryptosporidium* species? Hint: add another isolate search step.
- e. How many of step ‘b’ isolates originated from humans?
- f. How many of the isolates in step ‘b’ were typed using GP40/15 (GP60)? (hint: you can insert a step within a strategy. Click on the name of the step you want to insert a step before, then click on “Insert step before”).

Rename | View | Revise | Make Nested Strategy | **Insert Step Before** | Delete

STEP 3 : Host Name

Host : Mammals - Human
Isolate assay type : HTS, Sequencing Typed

Results: 1210 Isolates

+ Give this search a weight

- g. Compare some of these isolates using the multiple sequence alignment tool (ClustalW). Do you see any sequences with insertions or deletions?

My Step Result:

Geograph Loc - step 3 - 3 Isolates Add 3 Isolates to Basket | Download 3 Isolates

Isolates Advanced Paging Select Columns

Isolate Id	Country	Organism	Strain	Host	Isolation Source
EF519704 <input checked="" type="checkbox"/>	Italy	Cryptosporidium sp. CrIT-20	CrIT-20	Testudo marginata	feces
EF547155 <input checked="" type="checkbox"/>	Italy	Cryptosporidium sp. CrIT-20	CrIT-20	Testudo marginata	feces
EU331243 <input checked="" type="checkbox"/>	Czech Republic	Cryptosporidium sp. pig genotype II	H199	Homo sapiens	stool sample

Please select at least two isolates to run ClustalW. Note: only isolates from a single results page will be aligned. Increase the page size in advanced paging to increase the number that can be aligned.

Run Clustalw on Checked Strains
Check All
Uncheck All

Advanced Paging

- h. Take a look at the 'guide tree' that was built using this alignment. Change the isolates that you selected for alignment - how does the tree change? Do isolates from the same country cluster together?

8.2 Typing an unclassified isolate.

Note: For this exercise use <http://www.cryptodb.org>

- Run a search to find all unclassified *Cryptosporidium* isolates and find one that was typed using 18S small subunit ribosomal RNA. (Hint: Identify Isolates based on Taxon/Strain and choose 'unclassified' under *Cryptosporidium*. Add a column for Gene Product and sort the column).

Select Columns

Update Columns

clear all | expand all | collapse all
reset to current | reset to default

- Organism
- Strain/Isolate Name
- Host
- Isolation Source
- Geographic Location
- EuPathDB Annotated Host
- EuPathDB Annotated Isolation Source
- EuPathDB Annotated Geographic Location
- Gene Product (Sequence Typed Only)
- Description
- Is Reference
- Isolate Type
- Weight

clear all | expand all | collapse all
reset to current | reset to default

Update Columns

Gene Product (Sequence Typed Only)
18S rRNA
COWP
COWP

- Go to the isolate record page and copy the DNA sequence.
- Go to search for isolates based on BLAST, select isolates and make sure only the reference isolates are selected in the target organism window.
- Paste the DNA sequence in the input window and select the Blastn program. Click on "Get Answer".
- Explore your results. Based on the similarity which reference isolate is this one closest to?

Target Data Type ? Transcripts
 Proteins
 Genome
 EST
 ORF
 Isolates

BLAST Program ? blastn
 blastp
 blastx
 tblastn
 tblastx

Target Organism ? select all | clear all | expand all | collapse all | reset to default
 Cryptosporidium Isolates minus Reference
 Cryptosporidium Reference Isolates

Input Sequence ?

```
AATTAGAGTGCCTTAAGCAGGC
ATTGCGCTTGAATACTCCAGCATGGATAATATTTAAG
ATTTTATCTTTCTTATTGGTT
CTAAGATAAAAATAATGATTAATAGGGACAGTTGGGG
CATTTCATTTAA
```

Note: only one input sequence allowed.
maximum allowed sequence length is 31K bases.

Expectation value ? 10

Maximum descriptions (V) ? 50

Maximum alignments (B) ? 50

Low complexity filter ? yes

Give this search a weight

Give this search a name

8.3 Exploring isolates in *Plasmodium*.

Note: For this exercise use <http://www.plasmodb.org>

- a. Identify all isolates from Mexico.
- b. How many of those are *P. falciparum*? How many *P. vivax*?
- c. What about all of North and South America?
Hint: revise the first step in your strategy to include all countries in both continents.
- d. For these results, add columns such as isolate product and length. Sort these columns and explore your results. For example, what product is mainly used in typing *P. falciparum* isolates? What about *P. vivax* isolates?