Exploring Isolate Data
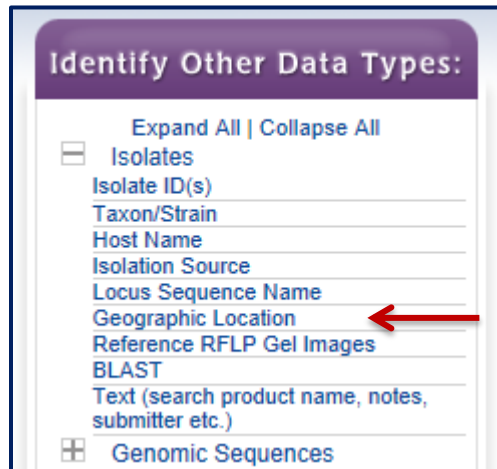## Exercise 8

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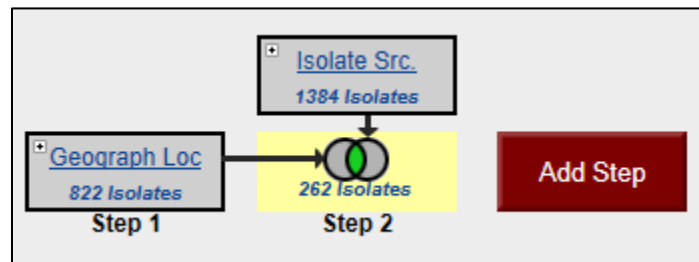
c. What is the general distribution of these isolates in Europe? (hint: you can do this quickly in two ways: sort the geographic location column by clicking on the sort arrows, then look at the represented countries; or use the "Isolate Geographic Location" tab to view a map and results summary table).

Isolates	Isolate Geographical Location

First 1 2 3 4 5 Next Last — Advanced Paging — Add Columns

Isolate	Geographic Location	Organism	Strain/Isolate Name	Host	Isolatio
AB247224	Serbia	Cryptosporidium parvum	#6	Unknown	fecal sample
AB242225	Serbia	Cryptosporidium parvum	#24	Unknown	fecal sample
AB242226	Serbia	Cryptosporidium parvum	#42	Unknown	fecal sample
AB242227	Serbia				

Sort by clicking on the arrows

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

Indicates the number of isolates from each location

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").

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

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

a. 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).

b. Go to the isolate record page and copy the DNA sequence.

c. Go to search for isolates based on BLAST, select isolates and make sure only the reference isolates are selected in the target organism window.

d. Paste the DNA sequence in the input window and select the Blastn program. Click on "Get Answer".

e. Explore your results. Based on the similarity which reference isolate is this one closest to?

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?