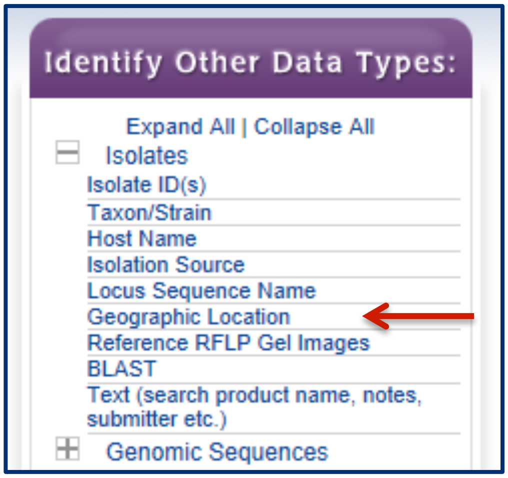
Exploring Isolate Data
Exercise 7

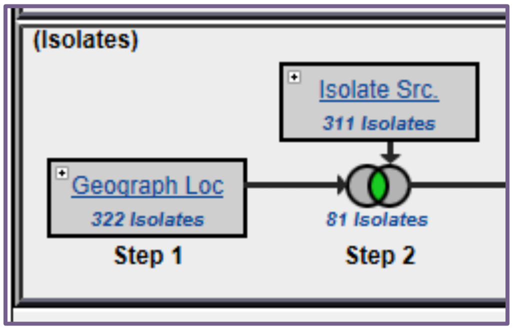
7.1 Exploring isolates in *Cryptosporidium* and using the alignment tool.

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

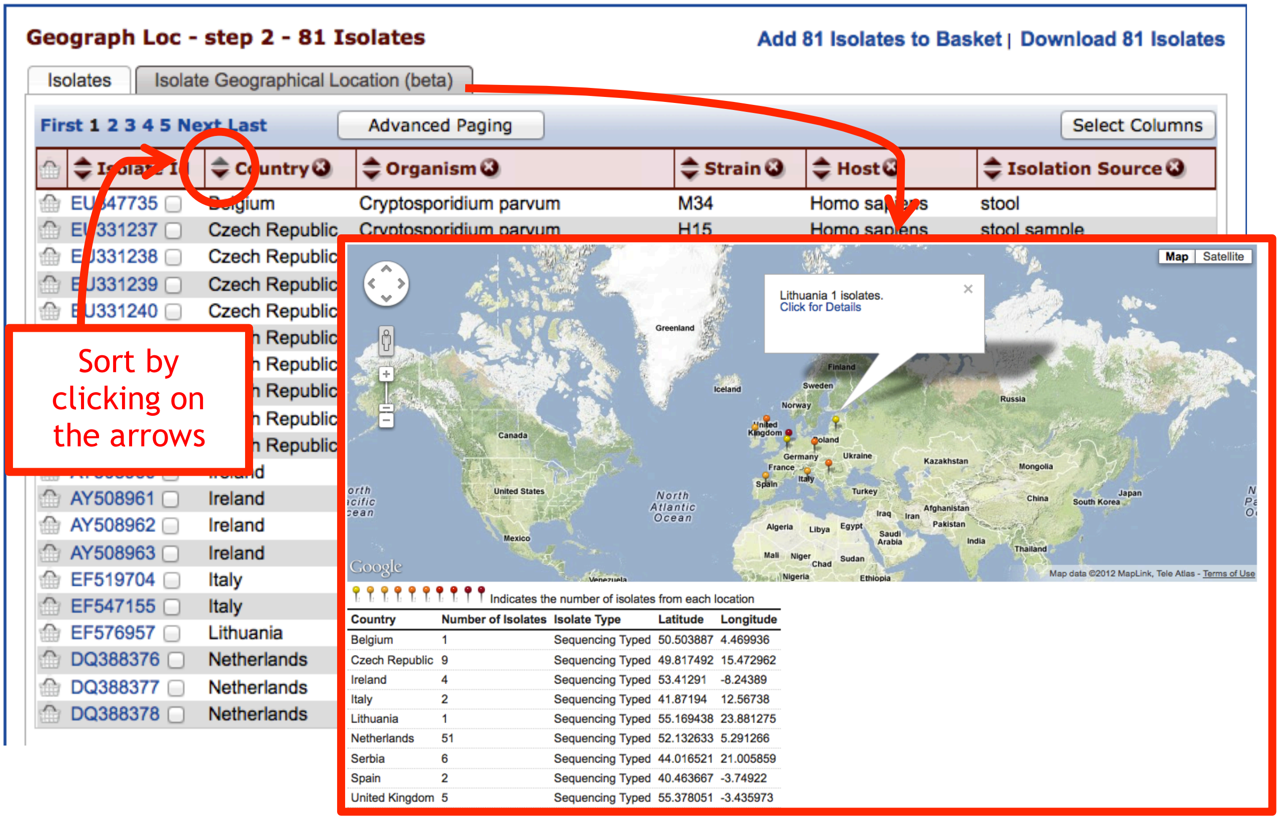
a. Identify all *Cryptosporidium* isolates from Europe.
Hint: search for isolates by geographic location in the "Identify Other Data Types" section.



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: A. sort the country column by clicking on the sort arrows, then look at the represented countries. B. Click on the “Geographic Location” tab to view a map and results summary table).



	Isolate ID	Country	Organism	Strain	Host	Isolation Source
	EU 47735 ☐	Belgium	Cryptosporidium parvum	M34	Homo sapiens	stool
	EU331237 ☐	Czech Republic	Cryptosporidium parvum	H15	Homo sapiens	stool sample
	EU331238 ☐	Czech Republic				
	EU331239 ☐	Czech Republic				
	EU331240 ☐	Czech Republic				
	AY508961 ☐	Ireland				
	AY508962 ☐	Ireland				
	AY508963 ☐	Ireland				
	EF519704 ☐	Italy				
	EF547155 ☐	Italy				
	EF576957 ☐	Lithuania				
	DQ388376 ☐	Netherlands				
	DQ388377 ☐	Netherlands				
	DQ388378 ☐	Netherlands				

Sort by clicking on the arrows

Lithuania 1 isolates.
Click for Details

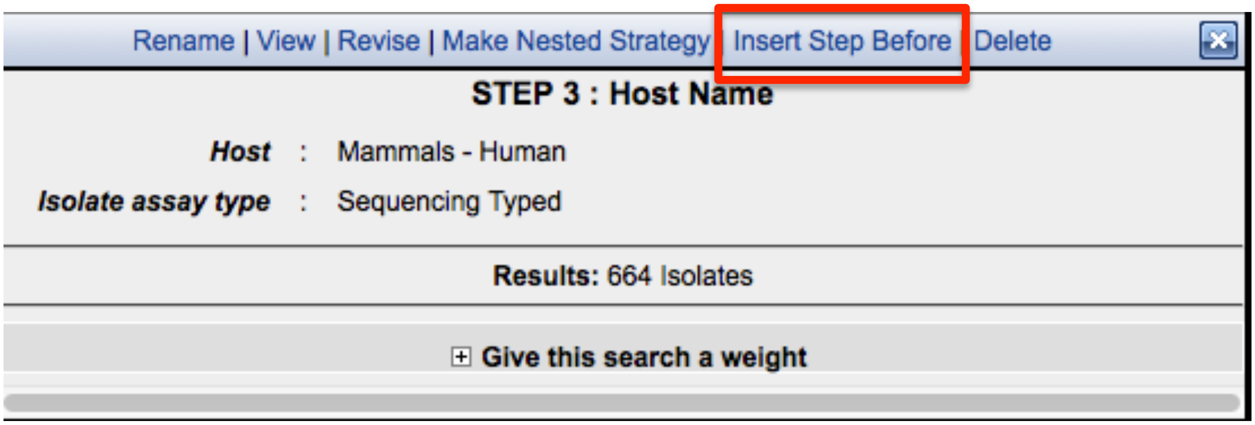
Indicates the number of isolates from each location

Country	Number of Isolates	Isolate Type	Latitude	Longitude
Belgium	1	Sequencing Typed	50.503887	4.469936
Czech Republic	9	Sequencing Typed	49.817492	15.472962
Ireland	4	Sequencing Typed	53.41291	-8.24389
Italy	2	Sequencing Typed	41.87194	12.56738
Lithuania	1	Sequencing Typed	55.169438	23.881275
Netherlands	51	Sequencing Typed	52.132633	5.291266
Serbia	6	Sequencing Typed	44.016521	21.005859
Spain	2	Sequencing Typed	40.463667	-3.74922
United Kingdom	5	Sequencing Typed	55.378051	-3.435973

d. Out of those in step ‘b’, how many are unclassified *Cryptosporidium* species? Hint: add another isolate search step.

e. How many of step ‘c’ isolates originated from humans?

f. How many of the isolates in step ‘b’ were typed using GP15/40 (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”).



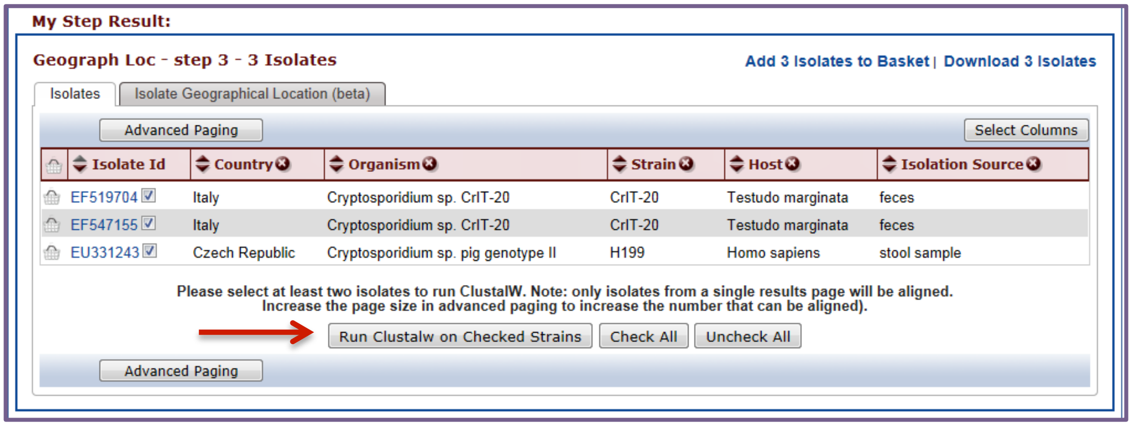
STEP 3 : Host Name

Host : Mammals - Human

Isolate assay type : Sequencing Typed

Results: 664 Isolates

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



My Step Result:

Geograph Loc - step 3 - 3 Isolates

Isolate Id	Country	Organism	Strain	Host	Isolation Source
EF519704 ☑	Italy	Cryptosporidium sp. CrIT-20	CrIT-20	Testudo marginata	feces
EF547155 ☑	Italy	Cryptosporidium sp. CrIT-20	CrIT-20	Testudo marginata	feces
EU331243 ☑	Czech Republic	Cryptosporidium sp. pig genotype II	H199	Homo sapiens	stool sample

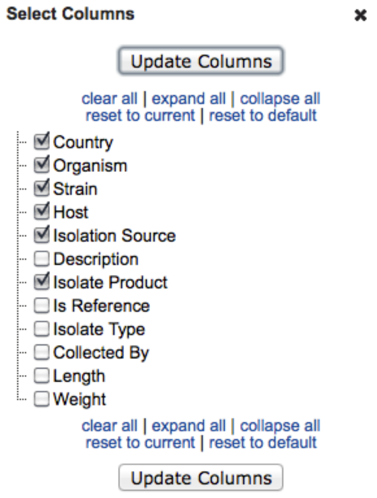
Please select at least two isolates to run ClustalW. Note: only isolates from a single results page will be aligned. Increase the page size in advanced paging to increase the number that can be aligned).

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?

7.2 Typing an unclassified isolate.

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

a. Select one of the unclassified isolates from step 7.1.d that was typed using 18S small subunit ribosomal RNA. (Hint: Add a column for isolate product).



Select Columns

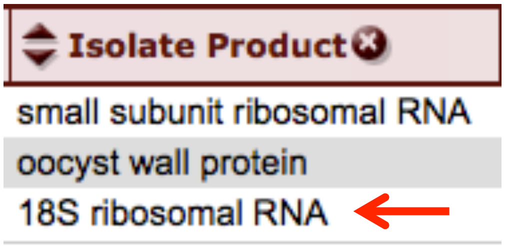
Update Columns

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

- ☑ Country
- ☑ Organism
- ☑ Strain
- ☑ Host
- ☑ Isolation Source
- ☐ Description
- ☑ Isolate Product
- ☐ Is Reference
- ☐ Isolate Type
- ☐ Collected By
- ☐ Length
- ☐ Weight

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

Update Columns



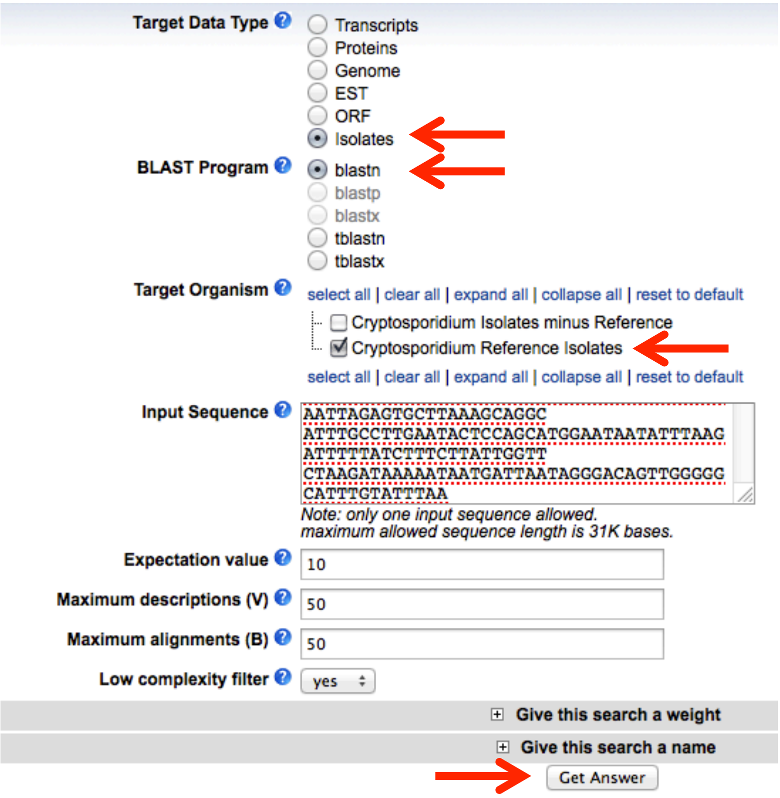
Isolate Product
small subunit ribosomal RNA
oocyst wall protein
18S ribosomal RNA

b. Copy the DNA sequence from the isolate record page.

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?



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

Input Sequence

```
AATTAGAGTGCTTAAAGCAGGC
ATTTGCCTTGAATACTCCAGCATGGAATAATATTTAAG
ATTTTTATCTTTCTTATTGGTT
CTAAGATAAAAATAATGATTAATAGGGACAGTTGGGGG
CATTTGTATTTAA
```

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

7.3 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?