Exploring Isolate Data
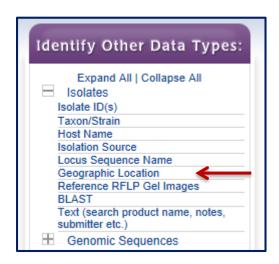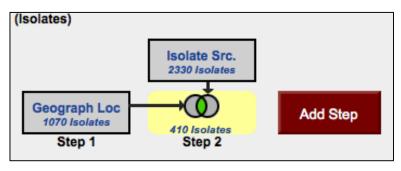
1. **Exploring isolates in *Cryptosporidium* and using the alignment tool. (http://www.cryptodb.org)**

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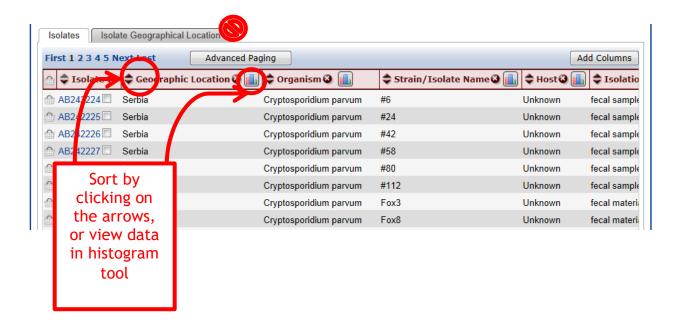

   b. How many of the *Cryptosporidium* isolates collected in Europe were isolated from feces? (hint: add another isolate search step – isolation source).
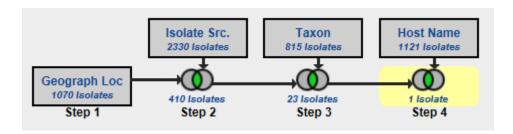

c. What is the general distribution of these isolates in Europe? (hint: you can do this quickly in two ways: sort the geographic location column by clicking on the sort arrows, then look at the represented countries; or use the histogram tool on the Geographic Location colum; or use the ~~"Isolate Geographic Location" tab to view a map and results summary table)~~.
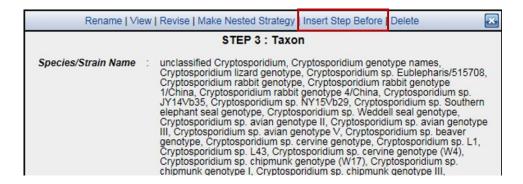
**For a sneak preview of our proposed new map functionality following our merger with Vectorbase, have a look at: https://www.vectorbase.org/popbio/map/ Try the filters (above the map) and tools to the left of the map.

d. Out of those in step 'b', how many are unclassified *Cryptosporidium* species? (hint: add another isolate search step and select taxon/strain then select the unclassified isolates)

e. How many of step 'd' isolates originated from humans?

f. How many of the isolates in step 'b' were typed using GlycoProtein GP40/15 (GP60)? (hint: you can insert a step within a strategy. Click "edit" on the step of interest then select "Insert step before").

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

STEP 3 : Taxon

Species/Strain Name : unclassified Cryptosporidium, Cryptosporidium genotype names, Cryptosporidium lizard genotype, Cryptosporidium sp. Eublepharis/515708, Cryptosporidium rabbit genotype, Cryptosporidium rabbit genotype 1/China, Cryptosporidium rabbit genotype 4/China, Cryptosporidium sp. JY14Vb35, Cryptosporidium sp. NY15Vb29, Cryptosporidium sp. Southern elephant seal genotype, Cryptosporidium sp. Weddell seal genotype, Cryptosporidium sp. avian genotype II, Cryptosporidium sp. avian genotype III, Cryptosporidium sp. avian genotype V, Cryptosporidium sp. beaver genotype, Cryptosporidium sp. cervine genotype, Cryptosporidium sp. L1, Cryptosporidium sp. L43, Cryptosporidium sp. cervine genotype (W4), Cryptosporidium sp. chipmunk genotype (W17), Cryptosporidium sp. chipmunk genotype I, Cryptosporidium sp. chipmunk genotype III,
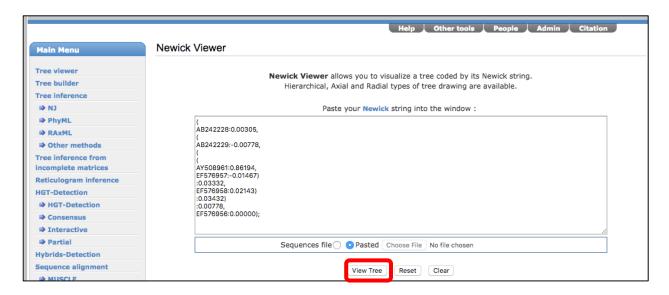
g. Compare some of these isolates using the multiple sequence alignment tool (ClustalW). Note: you must check the little square box next to the Isolate ID in order to include it in a Multiple Sequence Alignment. Do you see any sequences with insertions or deletions?

95 Isolates from Step 3
Strategy: *Geograph Loc*   Add 95 Isc

Isolates | Isolate Geographical Location

First 1 2 3 4 5 Next Last   Advanced Paging

▲ Isolate ▼ Id	▲ Organism ▼	▲ Strain/Isolate ▼ Name		▲ Host ▼	▲ Geographic ▼ Location
AB242224	Cryptosporidium parvum	#6		Unknown	Serbia
AB242225	Cryptosporidium parvum	#24		Unknown	Serbia
AB242226	Cryptosporidium parvum	#42		Unknown	Serbia
AB242227	Cryptosporidium parvum	#58		Unknown	Serbia
AB242228	Cryptosporidium parvum	#80		Unknown	Serbia

Please select at least two isolates to run ClustalW. Note: only isolates from a single page (Increase the page size in 'Advanced Paging' to increase the number that can be a

Run Clustalw on Checked Strains | Check All | Uncheck All

h. Take a look at the 'guide tree' that was built to help generate this alignment. The guide tree is located below the ".dnd" text located at the end of your multiple sequence alignment file. It may look something like the text below. The dendogram is in a "newick" file format.

```
(
AB242228:0.00305,
(
AB242229:-0.00778,
(
(
AY508961:0.86194,
EF576957:-0.01467)
:0.03332,
EF576958:0.02143)
:0.03432)
:0.00778,
EF576956:0.00000);
```

Note: the beginning " (" and closing "; " are important parts of the file format. You can use your mouse to select the text at the end of your file, copy it, and paste it into the box at the tree viewer site (remove the sample file in the box before adding your own sequence). Click on "view tree" to visualize the tree encoded in the text.
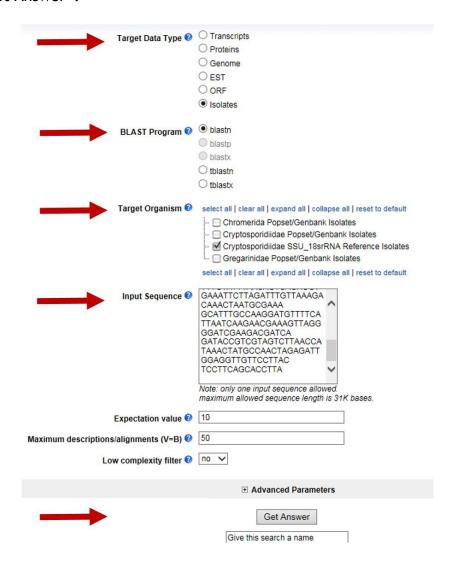
Change the isolates that you selected for alignment – how does the tree change?
Do isolates from the same country cluster together?

2. **Typing an unclassified *Cryptosporidium* isolate. (http://www.cryptodb.org)**

a. You have just finished sequencing part of the 18S small subunit ribosomal RNA gene from isolates you retrieved from a *Cryptosporidium* outbreak at a public swimming pool in Uppsala. The sequence was identical from all the isolates and is pasted below. Can you use CryptoDB to get an idea of which reference isolate this is most similar to? (hint: blast your sequence against the reference isolates using the BLAST tool found in Popset).

> AAGCTCGTAGTTGGATTTCTGTTAATAATTTATATAAAATATTTTGATGAATATTTATAT
> AATATTAACATAATTCATATTACTATATATTTTAGTATATGAAATTTTACTTTGAGAAAA
> TTAGAGTGCTTAAAGCAGGCATATGCCTTGAATACTCCAGCATGGAATAATATTAAAGAT
> TTTTATCTTTCTTATTGGTTCTAAGATAAGAATAATGATTAATAGGGACAGTTGGGGGCA
> TTTGTATTTAACAGTCAGAGGTGAAATTCTTAGATTTGTTAAAGACAAACTAATGCGAAA
> GCATTTGCCAAGGATGTTTTCATTAATCAAGAACGAAAGTTAGGGGATCGAAGACGATCA
> GATACCGTCGTAGTCTTAACCATAAACTATGCCAACTAGAGATTGGAGGTTGTTCCTTAC
> TCCTTCAGCACCTTA

b. You can get to the isolate BLAST page from the home page (BLAST link under popset isolate sequences) select "BLAST". Configure the BLAST search page: select isolates and make sure only the reference isolates are selected in the target organism window.

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

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

```
BLASTN 2.2.28+


Reference: Stephen F. Altschul, Thomas L. Madden, Alejandro A.
Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J.
Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of
protein database search programs", Nucleic Acids Res. 25:3389-3402.



Database:
Cryptosporidiidae_SSU_18SrRNA_Reference_Isolates.fasta
          73 sequences; 78,063 total letters



Query= MySeq1

Length=435
                                                    Score      E


  AF093490 | organism=Cryptosporidium_parvum | description=Crypto...   785    0.0

  AF164102 | organism=Cryptosporidium parvum strain IOWA | descri...   785    0.0

  AF093491 | organism=Cryptosporidium_hominis  renamed from C. pa...   762    0.0

  AF112571 | organism=Cryptosporidium tyzzeri - renamed from C. p...   760    0.0

  AF112572 | organism=Cryptosporidium parvum  ferret genotype | d...   756    0.0


> AF093490 | organism=Cryptosporidium_parvum | description=Cryptosporidium
parvum strain Bovine C. parvum genotype (BOH6) small
subunit ribosomal RNA gene, complete sequence. | length=1746
Length=1746

 Score =  785 bits (870),  Expect = 0.0
 Identities = 435/435 (100%), Gaps = 0/435 (0%)
 Strand=Plus/Plus

Query  1     AAGCTCGTAGTTGGATTTCTGTTAATAATTTATATAAAATATTTTGATGAATATTTATAT  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  601   AAGCTCGTAGTTGGATTTCTGTTAATAATTTATATAAAATATTTTGATGAATATTTATAT  660

Query  61    AATATTAACATAATTCATATTACTATATATTTTAGTATATGAAATTTTACTTTGAGAAAA  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  661   AATATTAACATAATTCATATTACTATATATTTTAGTATATGAAATTTTACTTTGAGAAAA  720

Query  121   TTAGAGTGCTTAAAGCAGGCATATGCCTTGAATACTCCAGCATGGAATAATATTAAAGAT  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  721   TTAGAGTGCTTAAAGCAGGCATATGCCTTGAATACTCCAGCATGGAATAATATTAAAGAT  780

Query  181   TTTTATCTTTCTTATTGGTTCTAAGATAAGAATAATGATTAATAGGGACAGTTGGGGGCA  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  781   TTTTATCTTTCTTATTGGTTCTAAGATAAGAATAATGATTAATAGGGACAGTTGGGGGCA  840

Query  241   TTTGTATTTAACAGTCAGAGGTGAAATTCTTAGATTTGTTAAAGACAAACTAATGCGAAA  300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  841   TTTGTATTTAACAGTCAGAGGTGAAATTCTTAGATTTGTTAAAGACAAACTAATGCGAAA  900

Query  301   GCATTTGCCAAGGATGTTTTCATTAATCAAGAACGAAAGTTAGGGGATCGAAGACGATCA  360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  901   GCATTTGCCAAGGATGTTTTCATTAATCAAGAACGAAAGTTAGGGGATCGAAGACGATCA  960
```