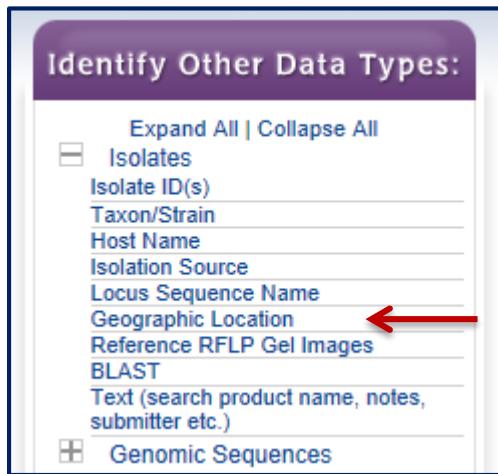
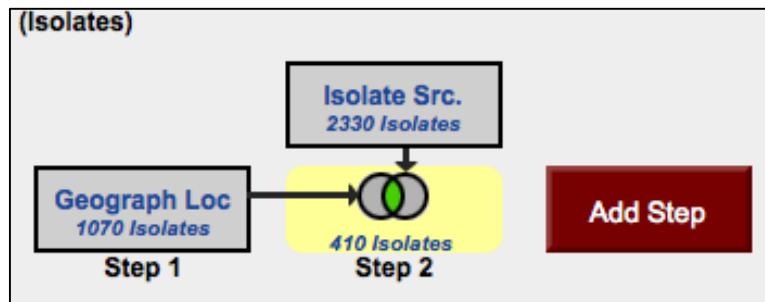


## Exploring Isolate Data

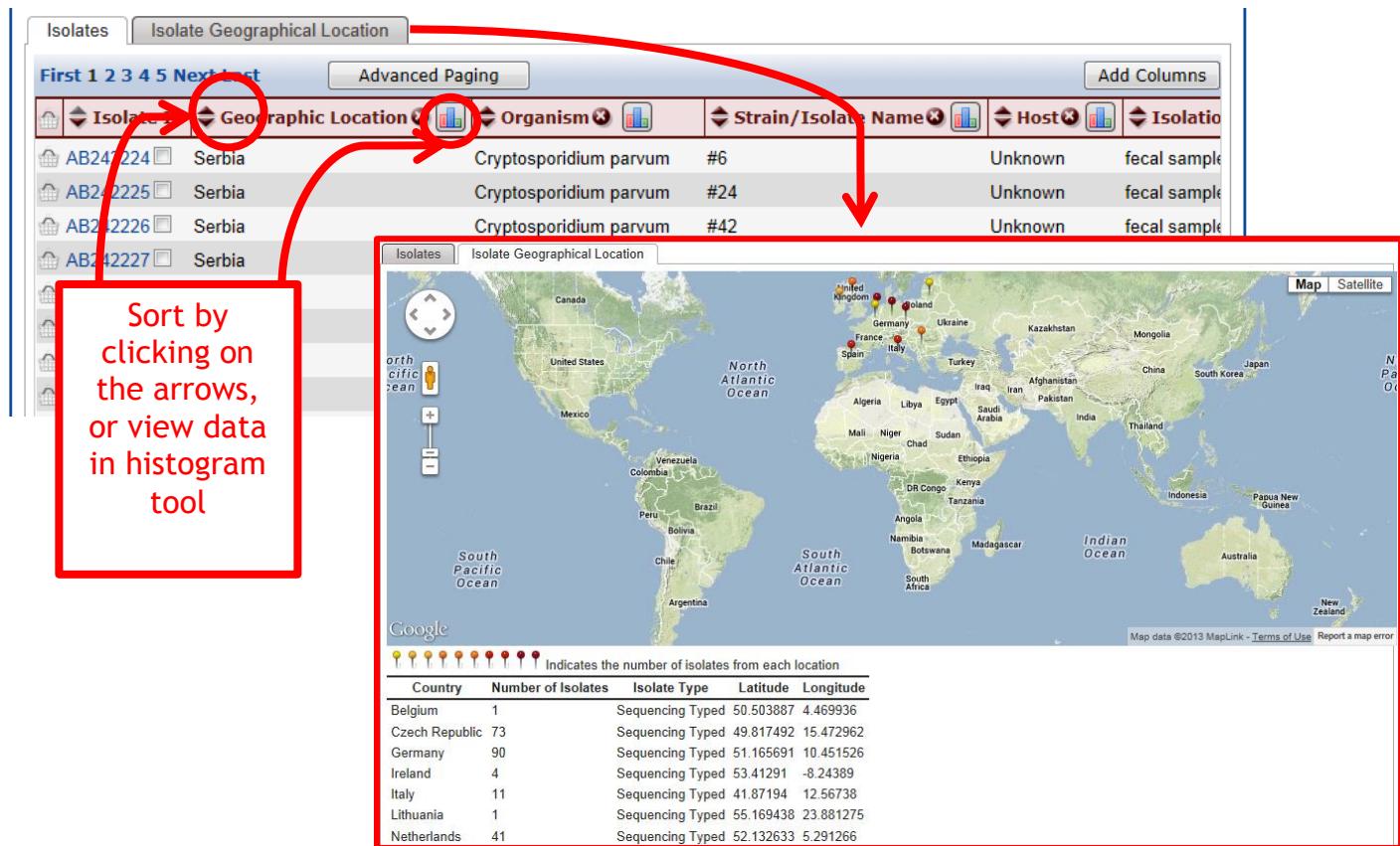
1. Exploring isolates in *Cryptosporidium* and using the alignment tool.  
[\(http://www.cryptodb.org\)](http://www.cryptodb.org)
  - a. Identify all *Cryptosporidium* isolates from Europe. (hint: search for isolates by geographic location in the “Identify Other Data Types” section).



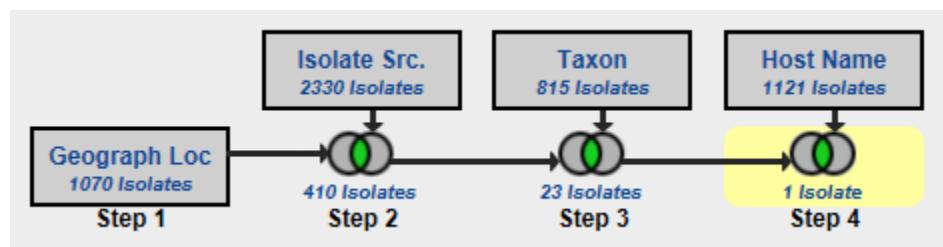
- b. How many of the *Cryptosporidium* isolates collected in Europe were isolated from feces? (hint: add another isolate search step - isolation source).



- 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).



- 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 GP40/15 (GP60)? (hint: you can insert a step within a strategy. Click “edit” on the step of interest then select “Insert step before”).

**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). Do you see any sequences with insertions or deletions?

**95 Isolates from Step 3**

**Strategy:** Geograph Loc

**Isolates** **Isolate Geographical Location** **Add 95 Isolates**

First	1	2	3	4	5	Next	Last	Advanced Paging
	<b>Isolate</b>	<b>Organism</b>	<b>Strain/Isolate Name</b>	<b>Host</b>	<b>Geographic Location</b>			
	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 dendrogram 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.

The screenshot shows the 'Newick Viewer' interface. On the left is a 'Main Menu' sidebar with various tree inference methods listed under 'Tree inference'. The main area is titled 'Newick Viewer' and contains a text input box for pasting Newick strings. Below the input box are buttons for 'Sequences file' (radio button), 'Pasted' (radio button, selected), 'Choose File', and 'No file chosen'. At the bottom right of the input area is a red-outlined 'View Tree' button, which is the target of the instruction. Above the input box, a descriptive text states: 'Newick Viewer allows you to visualize a tree coded by its Newick string. Hierarchical, Axial and Radial types of tree drawing are available.'

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: go to the BLAST page in CryptoDB and blast your sequence against the reference isolates).

```
AAGCTCGTAGTTGGATTCTGTTAATAATTATATAAAAATATTTGATGAATATTATAT  
AATATTAACATAATTCAATTACTATATATTAGTATATGAAATTACTTTGAGAAAA  
TTAGAGTGCTTAAAGCAGGCATATGCCCTGAATACTCCAGCATGGAATAATATTAAAGAT  
TTTATCTTCTTATTGGTTCTAACAGATAAGAATAATGATTAATAGGGACAGTTGGGGGCA  
TTGTATTAAACAGTCAGAGGTGAAATTCTTAGATTGTTAAAGACAAACTATGCGAAA  
GCATTTGCCAAGGATGTTTCATTAATCAAGAACGAAAGTTAGGGGATCGAAGACGATCA  
GATACCGTCGTAGTCTAACCATAAACTATGCCAACTAGAGATTGGAGGTTTCCTTAC  
TCCTTCAGCACCTTA
```

- b. You can get to the BLAST page from the home page (BLAST link under the tool section) or from the isolate searches and 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”.

The screenshot shows the CryptoDB BLAST search interface. A red arrow points to the "Target Data Type" section, where "Isolates" is selected. Another red arrow points to the "BLAST Program" section, where "blastn" is selected. A third red arrow points to the "Target Organism" section, where the "Cryptosporidiidae SSU\_18srRNA Reference isolates" checkbox is checked. A fourth red arrow points to the "Input Sequence" section, which contains the DNA sequence provided in the task. A large red arrow points to the "Get Answer" button at the bottom right.

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  
-  Chromerida Popset/Genbank Isolates  
-  Cryptosporidiidae Popset/Genbank Isolates  
-  Cryptosporidiidae SSU\_18srRNA Reference isolates  
-  Greganidae Popset/Genbank Isolates  
select all | clear all | expand all | collapse all | reset to default

Input Sequence ⓘ  
GAAATTCTTAGATTGTTAAAGA  
CAAACTAATGCGAA  
GCATTGCCAACGATGTTCA  
TTAACCAAGAACGAAAGTTAGG  
GGATCGAACGAGATCA  
GATACCGTCGTAGTCTTAACCA  
TAAACTATCCCAACTAGAGATT  
GGAGGTTGTTCCCTAC  
TCCTTCAGCACCTTA

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

Expectation value ⓘ 10

Maximum descriptions/alignments (V=B) ⓘ 50

Low complexity filter ⓘ no

Advanced Parameters

Get Answer

Give this search a name

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

<b>AF093490</b>	organism=Cryptosporidium_parvum	description=Cryptosporidium_parvum	862	0.0
<b>AF093491</b>	organism=Cryptosporidium_parvum	description=Cryptosporidium_parvum	817	0.0
<b>AF112571</b>	organism=Cryptosporidium_parvum	description=Cryptosporidium_parvum	813	0.0
<b>AF112572</b>	organism=Cryptosporidium_parvum	description=Cryptosporidium_parvum	809	0.0
<b>AF115378</b>	organism=Cryptosporidium_wrairi	description=Cryptosporidium_wrairi	809	0.0
<b>AF159110</b>	organism=Cryptosporidium_parvum	description=Cryptosporidium_parvum	801	0.0
<b>AF112574</b>	organism=Cryptosporidium_meleagridis	description=Cryptosporidium_meleagridis	799	0.0
<b>AY737573</b>	organism=Cryptosporidium_environmental_sequence	description=Cryptosporidium_environmental_sequence	781	0.0
<b>AY737560</b>	organism=Cryptosporidium_environmental_sequence	description=Cryptosporidium_environmental_sequence	761	0.0
<b>AF262330</b>	organism=Cryptosporidium_sp.	description=Cryptosporidium_sp.	753	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 = 862 bits (435), Expect = 0.0
  Identities = 435/435 (100%)
  Strand = Plus / Plus

  Query: 1      aagctcgtagttggatttctgttaataatttatataaaaatatttgatgaatatttatat 60
  Sbjct: 601    aagctcgtagttggatttctgttaataatttatataaaaatatttgatgaatatttatat 660

  Query: 61     aatattaacataattcatattactatatatatttagtatatgaaattttactttgagaaaa 120
  Sbjct: 661    aatattaacataattcatattactatatatatttagtatatgaaattttactttgagaaaa 720

  Query: 121    ttagagtgtttaaaggcaggcatatgccttgaataactccagcatggaataatattaaagat 180
  Sbjct: 721    ttagagtgtttaaaggcaggcatatgccttgaataactccagcatggaataatattaaagat 780

  Query: 181    ttttatcttcattgggttctaagataagaataatgattaatagggacagttggggca 240
  Sbjct: 781    ttttatcttcattgggttctaagataagaataatgattaatagggacagttggggca 840

  Query: 241    tttgtatccaacagtcaagaggtaaaattcttagattttgtttaagacaaaactaatgcgaaa 300
  Sbjct: 841    tttgtatccaacagtcaagaggtaaaattcttagattttgtttaagacaaaactaatgcgaaa 900

  Query: 301    gcatttgccaaggatgtttcattaatcaagaacgaaaatggggatcgaaagacgatca 360
  Sbjct: 901    gcatttgccaaggatgtttcattaatcaagaacgaaaatggggatcgaaagacgatca 960

```