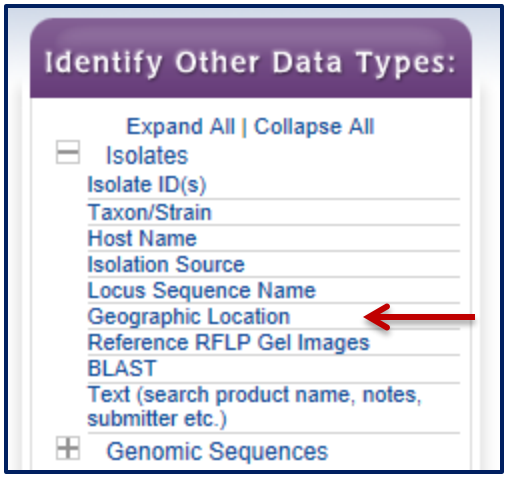
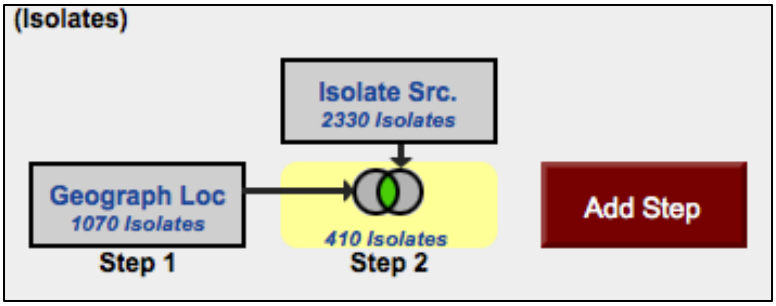
Exploring Isolate Data

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

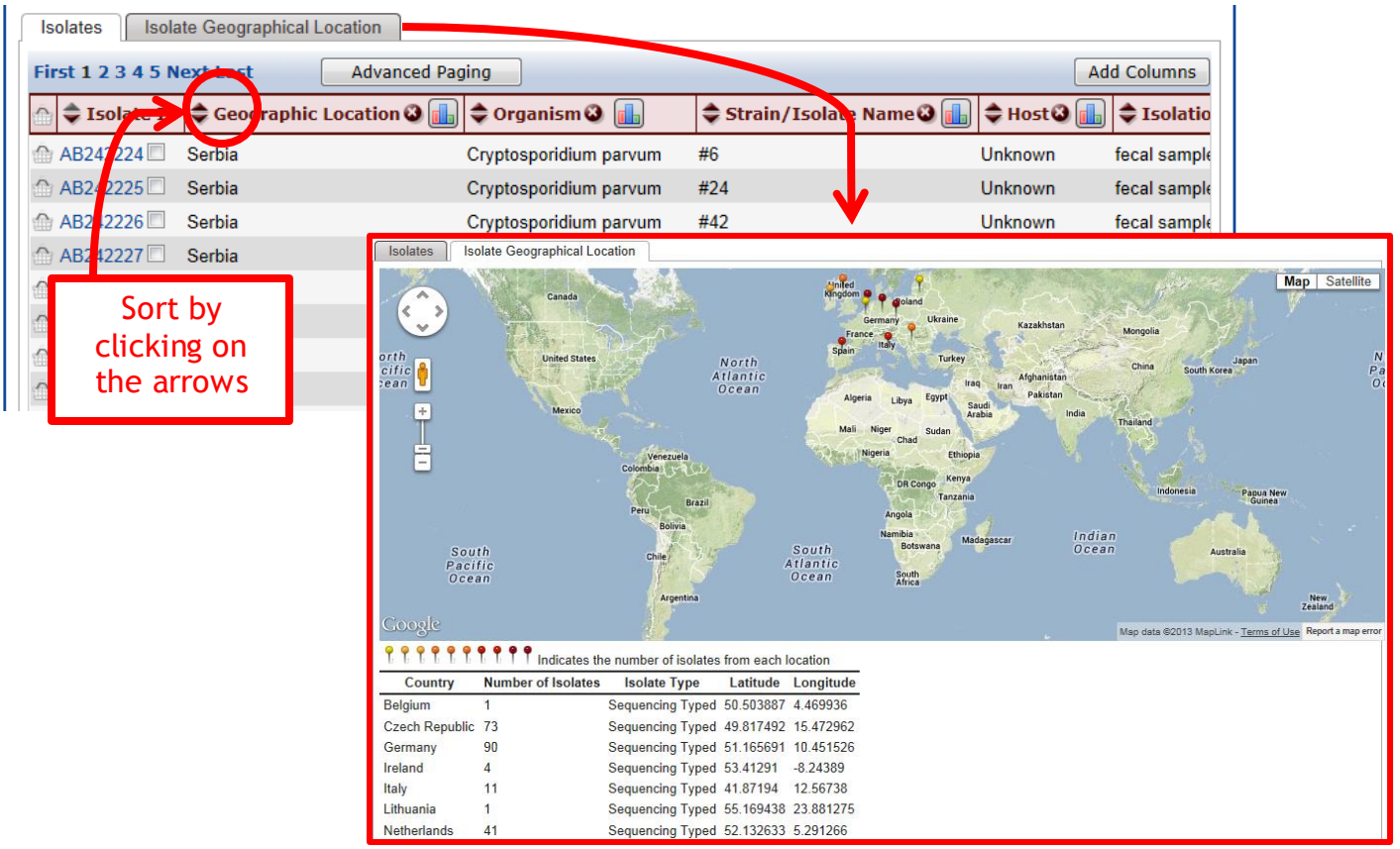
   a. Identify all *Cryptosporidium* isolates from Europe. (hint: search for isolates by geographic location in the "Identify Other Data Types" section).



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

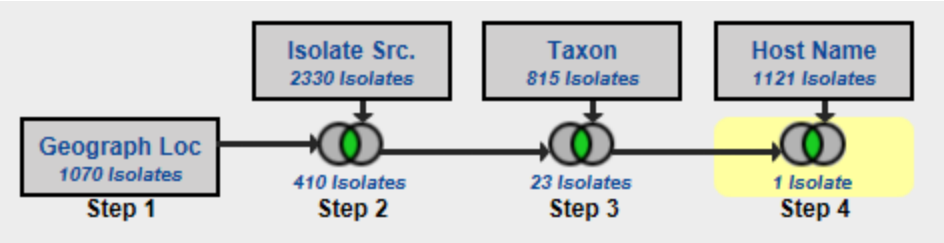


c. What is the general distribution of these isolates in Europe? (hint: you can do this quickly in two ways: sort the geographic location column by clicking on the sort arrows, then look at the represented countries; or use the "Isolate Geographical Location" tab to view a map and results summary table).

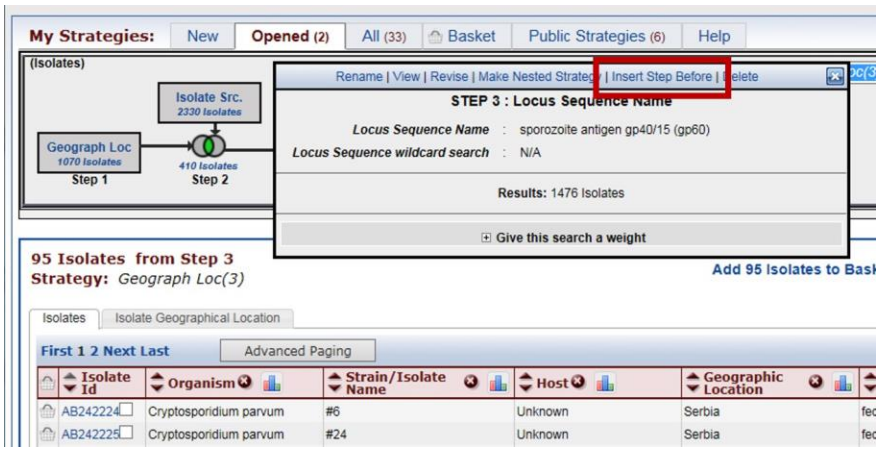


d. Out of those in step 'b', how many are unclassified *Cryptosporidium* species? (hint: add another isolate search step and select taxon/strain then select the unclassified isolates)

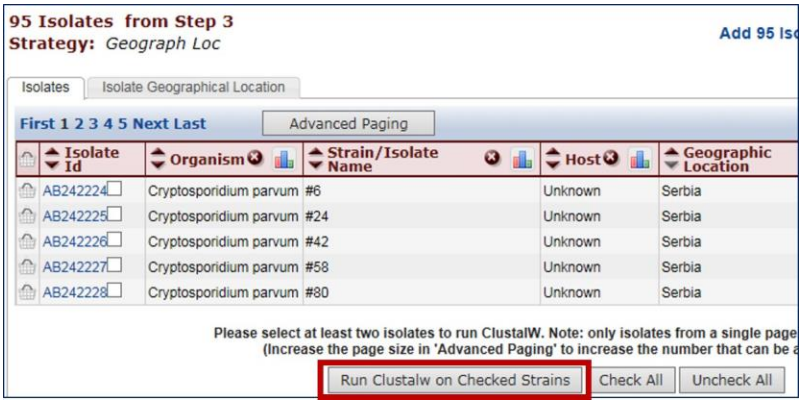
e. How many of step 'd' isolates originated from humans?



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”).

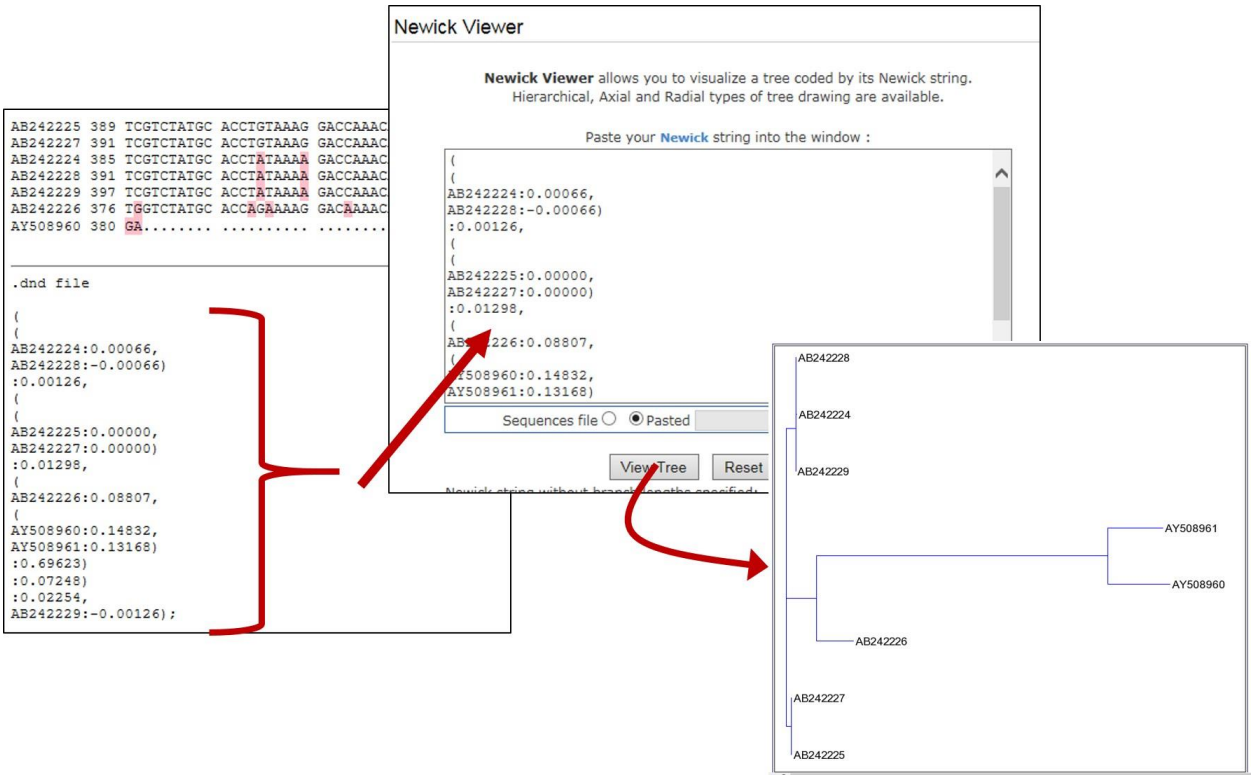


g. Compare the first 8 isolates using the multiple sequence alignment tool (ClustalW). Do you see any sequences with insertions or deletions?



h. Create a guide tree based on this alignment. Below the alignment in the output file, are the contents of a .dnd file. The Newick Viewer (link below) uses the dnd file to create a guide tree. Cut and paste the .dnd file beginning with the first open parenthesis into the Newick Viewer string box. Then click View Tree.

http://www.trex.uqam.ca/index.php?action=newick



i. Change the isolates that you selected for alignment – how does the tree change? Do isolates from the same country cluster together?

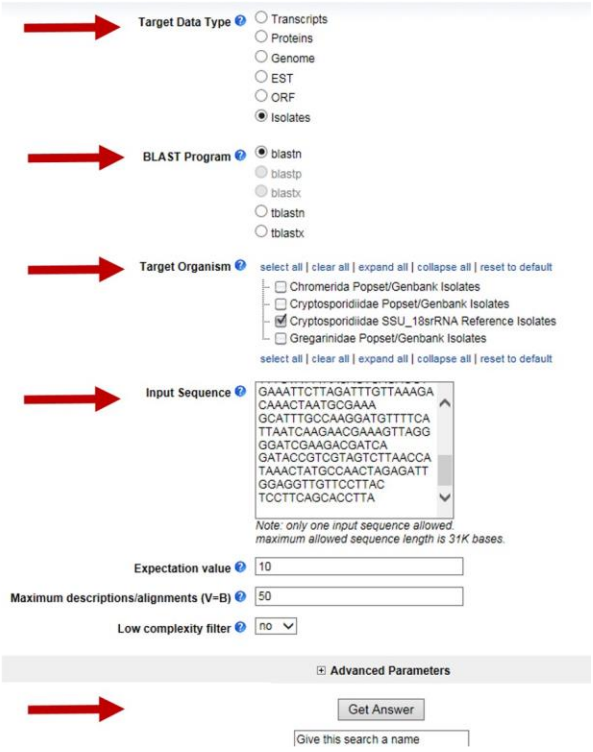
2. Typing an unclassified *Cryptosporidium* isolate. (http://www.cryptodb.org)

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).

```
AAGCTCGTAGTTGGATTTCTGTTAATAATTTATATAAAATATTTTGATGAATATTTATAT
AATATTAACATAATTCATATTACTATATATTTTAGTATATGAAATTTTACTTTGAGAAAA
TTAGAGTGCTTAAAGCAGGCATATGCCTTGAATACTCCAGCATGGAATAATATTAAAGAT
TTTTATCTTTCTTATTGGTTCTAAGATAAGAATAATGATTAATAGGGACAGTTGGGGGCA
TTTGTATTTAACAGTCAGAGGTGAAATTCTTAGATTTGTTAAAGACAAACTAATGCGAAA
GCATTTGCCAAGGATGTTTTCATTAATCAAGAACGAAAGTTAGGGGATCGAAGACGATCA
GATACCGTCGTAGTCTTAACCATAAACTATGCCAACTAGAGATTGGAGGTTGTTCCTTAC
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".



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

```
Query= MySeq1

Length=435
                                                                      Score     E

GQ983352.1| organism=Cryptosporidium hominis | description=Cryp...   785    0.0



> GQ983352.1| organism=Cryptosporidium hominis | description=Cryptosporidium
hominis isolate W15271 small subunit ribosomal
RNA gene, partial sequence
Length=92932

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

Query  1     AAGCTCGTAGTTGGATTTCTGttaataatttatataaaatattttgatgaatatttatat  60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1501  AAGCTCGTAGTTGGATTTCTGTTAATAATTTATATAAAATATTTTGATGAATATTTATAT  1560

Query  61    aatattaacataattcatattactatatattttagtatatGAAATTTTACTTTGAGAAAA  120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1561  AATATTAACATAATTCATATTACTATATATTTTAGTATATGAAATTTTACTTTGAGAAAA  1620

Query  121   TTAGAGTGCTTAAAGCAGGCATATGCCTTGAATACTCCAGCATGGAATAATATTAAAGAT  180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1621  TTAGAGTGCTTAAAGCAGGCATATGCCTTGAATACTCCAGCATGGAATAATATTAAAGAT  1680

Query  181   TTTTATCTTTCTTATTGGTTCTAAGATAAGAATAATGATTAATAGGGACAGTTGGGGGCA  240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  1681  TTTTATCTTTCTTATTGGTTCTAAGATAAGAATAATGATTAATAGGGACAGTTGGGGGCA  1740
```