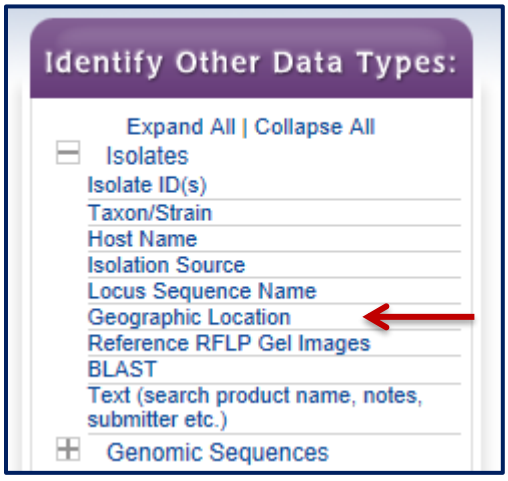
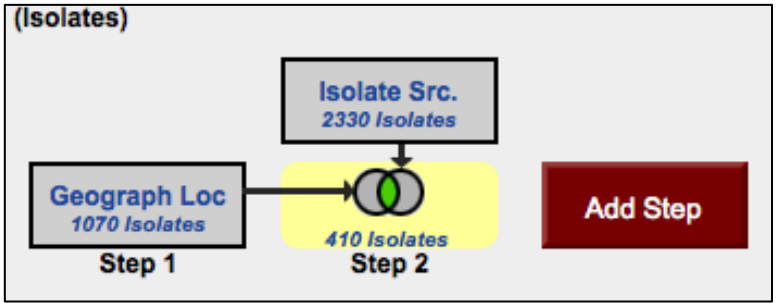
Exploring Isolate Data

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

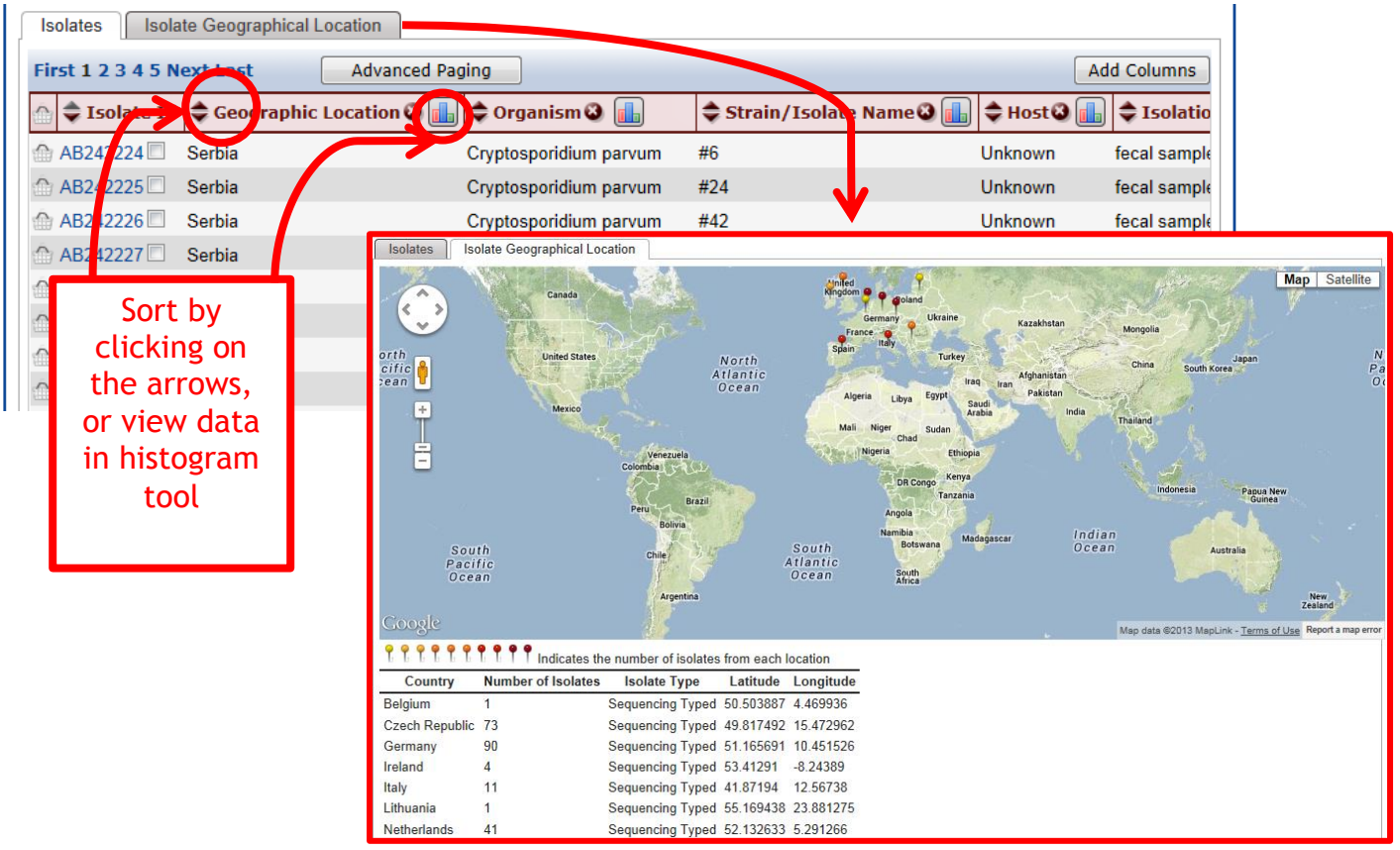
   a. Identify all *Cryptosporidium* isolates from Europe. (hint: search for isolates by geographic location in the "Identify Other Data Types" section).



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

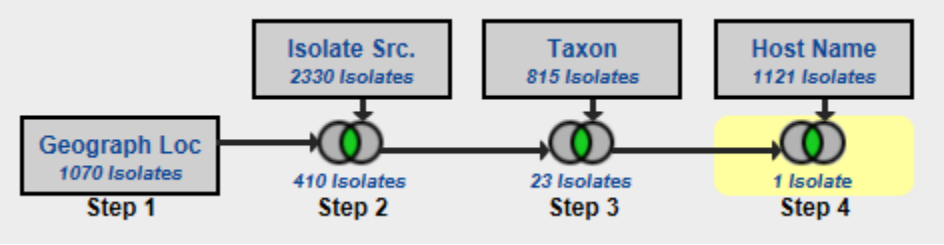


c. What is the general distribution of these isolates in Europe? (hint: you can do this quickly in two ways: sort the geographic location column by clicking on the sort arrows, then look at the represented countries; or use the histogram tool on the Geographic Location colum; or use the "Isolate Geographical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

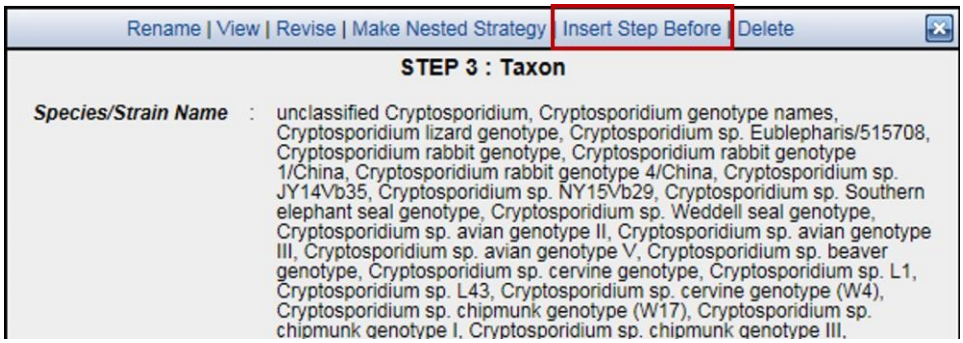


d. Out of those in step 'b', how many are unclassified *Cryptosporidium* species? (hint: add another isolate search step and select taxon/strain then select the unclassified isolates)

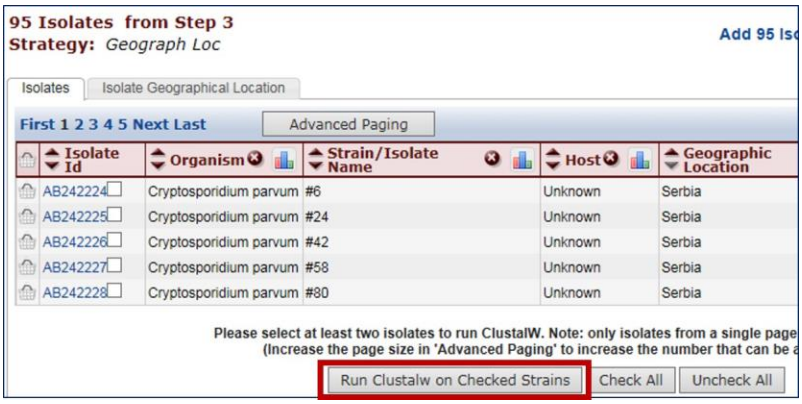
e. How many of step 'd' isolates originated from humans?



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”).



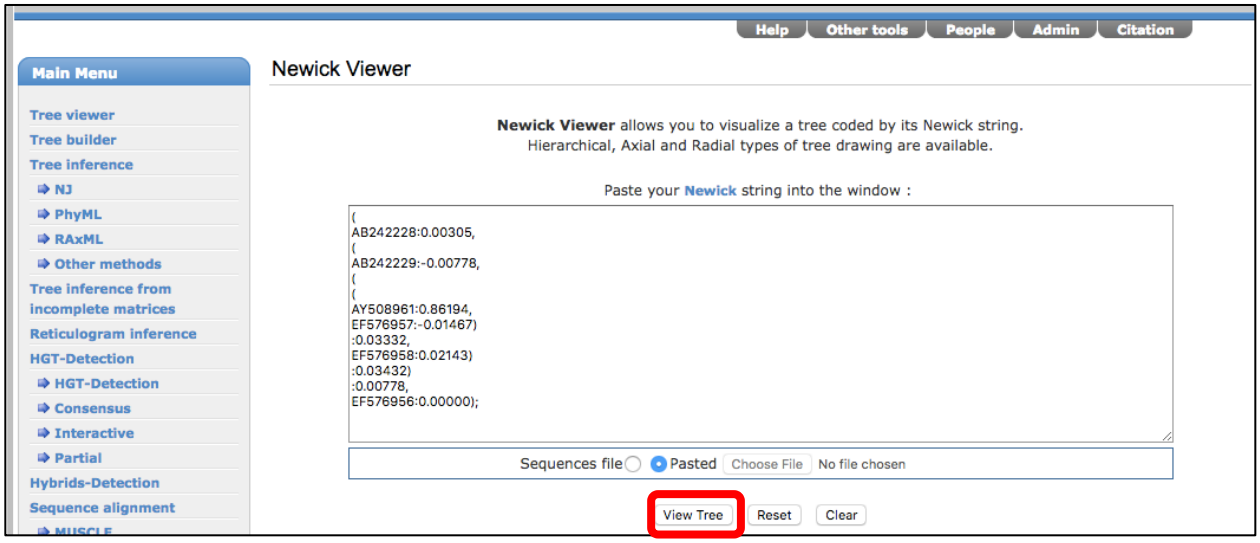
g. Compare some of these isolates using the multiple sequence alignment tool (ClustalW). Do you see any sequences with insertions or deletions?



h. Take a look at the ‘guide tree’ that was built 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?

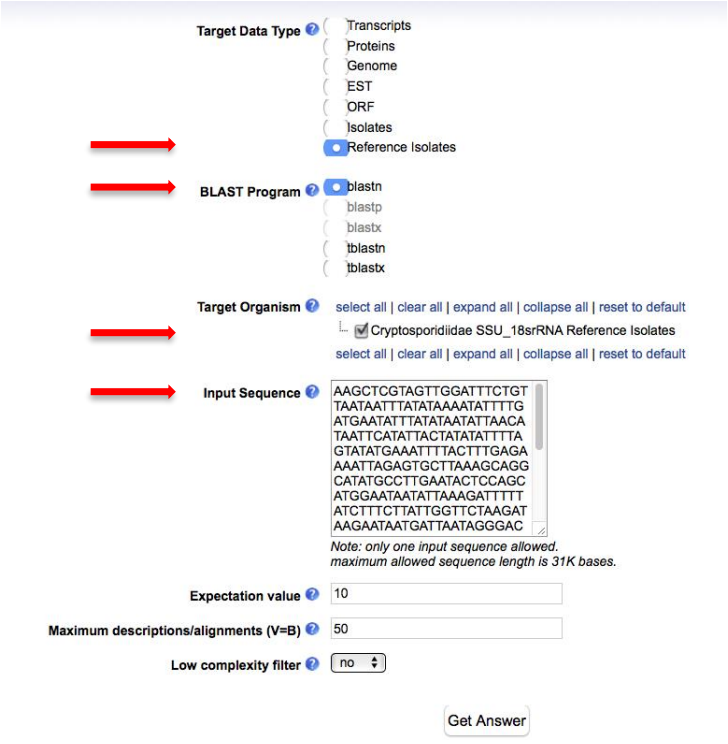
2. Typing an unclassified *Cryptosporidium* isolate. (http://www.cryptodb.org)

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: 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?

		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
AF115378	organism=Cryptosporidium_wrairi | description=Crypto...	756	0.0
AF112574	organism=Cryptosporidium_meleagridis | description=C...	749	0.0
EF641022	organism=Cryptosporidium sp. beaver genotype | desc...	742	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

Query  361   GATACCGTCGTAGTCTTAACCATAAACTATGCCAACTAGAGATTGGAGGTTGTTCCTTAC  420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  961   GATACCGTCGTAGTCTTAACCATAAACTATGCCAACTAGAGATTGGAGGTTGTTCCTTAC  1020

Query  421   TCCTTCAGCACCTTA  435
             |||||||||||||||
Sbjct  1021  TCCTTCAGCACCTTA  1035
```