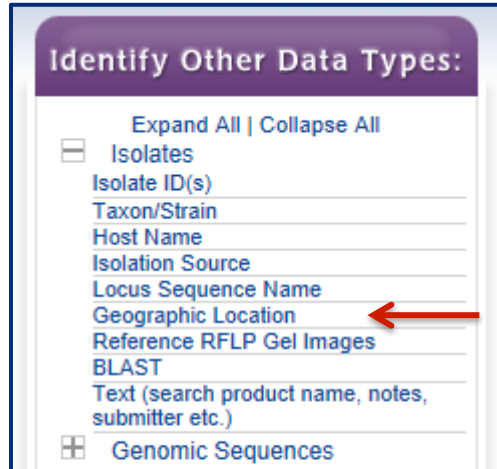
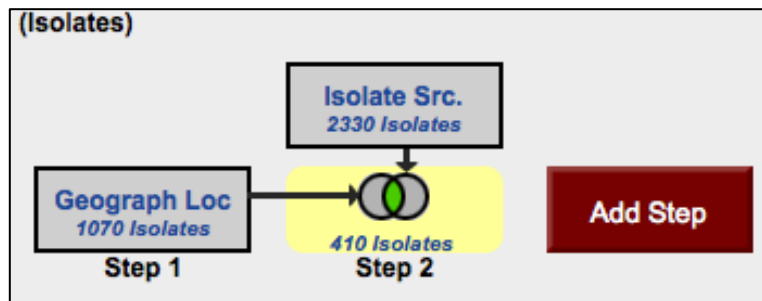


## Exploring Isolate Data Exercise 3

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).

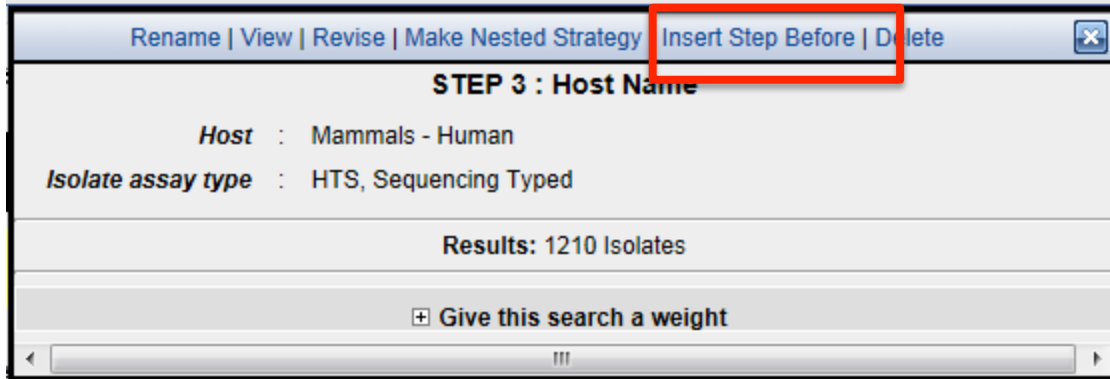


- c. What is the general distribution of these isolates in Europe? (hint: you can do this quickly in two ways: sort the geographic location column by clicking on the sort arrows, then look at the represented countries; or use the “Isolate Geographical Location” tab to view a map and results summary table).

Sort by clicking on the arrows

Country	Number of Isolates	Isolate Type	Latitude	Longitude
Belgium	1	Sequencing Typed	50.503887	4.469936
Czech Republic	73	Sequencing Typed	49.817492	15.472962
Germany	90	Sequencing Typed	51.165691	10.451526
Ireland	4	Sequencing Typed	53.41291	-8.24389
Italy	11	Sequencing Typed	41.87194	12.56738
Lithuania	1	Sequencing Typed	55.169438	23.881275
Netherlands	41	Sequencing Typed	52.132633	5.291266

- 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?

**My Step Result:**

Geograph Loc - step 3 - 3 Isolates [Add 3 Isolates to Basket](#) | [Download 3 Isolates](#)

Isolates

<input type="checkbox"/>	Isolate Id	Country	Organism	Strain	Host	Isolation Source
<input checked="" type="checkbox"/>	EF519704	Italy	Cryptosporidium sp. CrIT-20	CrIT-20	Testudo marginata	feces
<input checked="" type="checkbox"/>	EF547155	Italy	Cryptosporidium sp. CrIT-20	CrIT-20	Testudo marginata	feces
<input checked="" type="checkbox"/>	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?

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
AATATTAACATAATTCATATTACTATATATTTTAGTATATGAAATTTACTTTGAGAAAA
TTAGAGTGCTTAAAGCAGGCATATGCCTTGAATACTCCAGCATGGAATAATATTAAGAT
TTTTATCTTTCTTATTGGTTCTAAGATAAGAATAATGATTAATAGGGACAGTTGGGGGCA
TTTGTATTTAACAGTCAGAGGTGAAATTCCTAGATTTGTTAAAGACAACTAATGCGAAA
GCATTTGCCAAGGATGTTTTTCATTAATCAAGAACGAAAGTTAGGGGATCGAAGACGATCA
GATACCGTCGTAGTCTTAACCATAAACTATGCCAACTAGAGATTGGAGGTTGTTCCCTTAC
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. Red arrows point to the following elements:

- Target Data Type:** The 'Isolates' radio button is selected.
- BLAST Program:** The 'blastn' radio button is selected.
- Target Organism:** The 'Cryptosporidiidae Reference Isolates' checkbox is checked.
- Input Sequence:** The DNA sequence from the previous block is pasted into the input field.

Other visible settings include:

- Expectation value: 10
- Maximum descriptions/alignments (V=B): 10
- Low complexity filter: no
- Advanced Parameters: expanded
- Get Answer button: highlighted with a red arrow.

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

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

```

AF093490 | organism=Cryptosporidium parvum | description=Cryptos... 862 0.0
AF093491 | organism=Cryptosporidium parvum | description=Cryptos... 817 0.0
AF112571 | organism=Cryptosporidium parvum | description=Cryptos... 813 0.0
AF112572 | organism=Cryptosporidium parvum | description=Cryptos... 809 0.0
AF115378 | organism=Cryptosporidium wrairi | description=Cryptos... 809 0.0
AF159110 | organism=Cryptosporidium parvum | description=Cryptos... 801 0.0
AF112574 | organism=Cryptosporidium meleagridis | description=Cr... 799 0.0
AY737573 | organism=Cryptosporidium environmental_sequence | des... 781 0.0
AY737560 | organism=Cryptosporidium environmental_sequence | des... 761 0.0
AF262330 | organism=Cryptosporidium_sp. | description=Cryptospor... 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 aagctcgtagttggatttctgttaataatttatataaaatattttgatgaatatttatat 60
|||||
Sbjct: 601 aagctcgtagttggatttctgttaataatttatataaaatattttgatgaatatttatat 660

Query: 61 aatattaacataattcatattactatatatatttttagtatatgaaattttactttgagaaaa 120
|||||
Sbjct: 661 aatattaacataattcatattactatatatatttttagtatatgaaattttactttgagaaaa 720

Query: 121 ttagagtgccttaagcaggcatatgccttgaatactccagcatggaataatattaaagat 180
|||||
Sbjct: 721 ttagagtgccttaagcaggcatatgccttgaatactccagcatggaataatattaaagat 780

Query: 181 ttttatctttcttattggttctaagataagaataatgattaatagggacagttgggggca 240
|||||
Sbjct: 781 ttttatctttcttattggttctaagataagaataatgattaatagggacagttgggggca 840

Query: 241 tttgtatttaacagtcagaggtgaaattcttagatttggttaagacaaactaatgcgaaa 300
|||||
Sbjct: 841 tttgtatttaacagtcagaggtgaaattcttagatttggttaagacaaactaatgcgaaa 900

Query: 301 gcatttgccaaggatgttttcattaatcaagaacgaaagttaggggatcgaagacgatca 360
|||||
Sbjct: 901 gcatttgccaaggatgttttcattaatcaagaacgaaagttaggggatcgaagacgatca 960

```