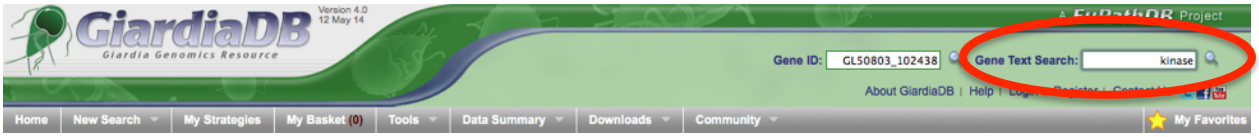
Finding Genes And Exploring The Gene Page
(Exercise 1)

1. **Finding a gene using text search (http://giardiadb.org).**

a. **Find all possible kinases in *Giardia*.**

Hint: use the keyword "kinase" (without quotations) in the "Gene Text Search" box.



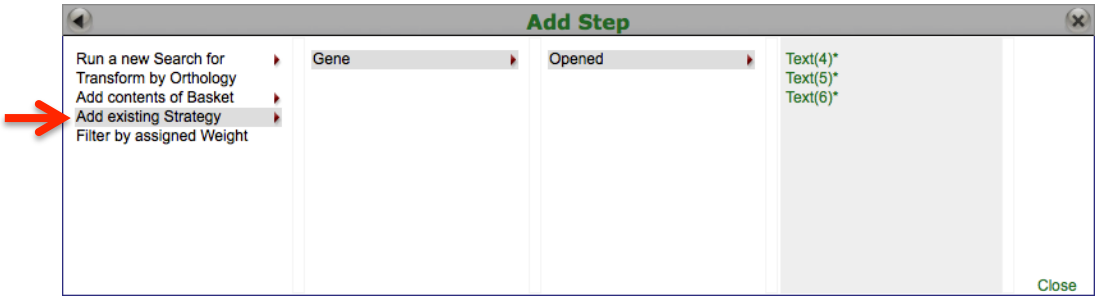
- How many genes did you get?
- How many of those are in assemblage E? How did you find this out?
- What happens if you search using the word "kinases"? How many results did you return?

b. **How can you increase the number of possible kinases in your results?**

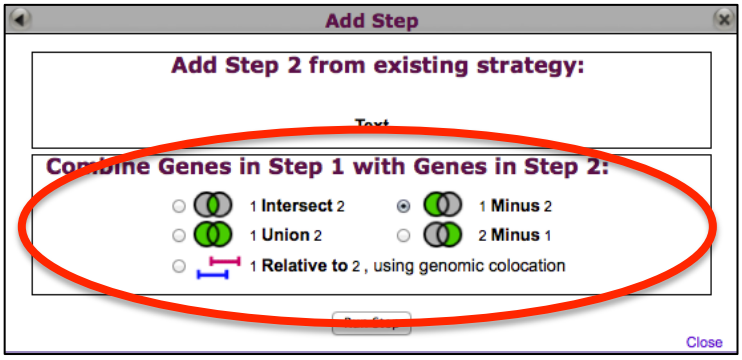
Hint: the search you did in 'a' will miss things like "pyrophosphokinase" or "kinases" so you need to use a wild card in your search – try "kinase*", "*kinase" and "*kinase*" (without quotations).

- Did you get more results?
- Which one of the above wild card combinations gave you the largest number of kinases?

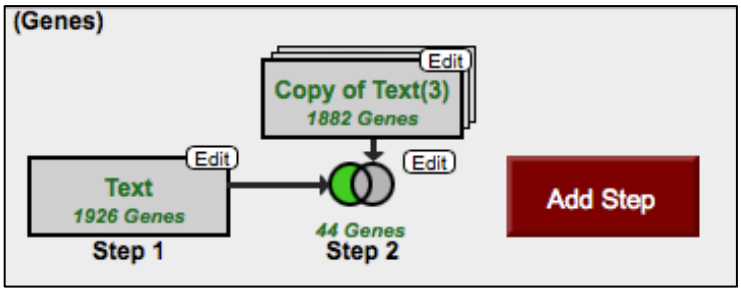
- How can you quickly examine the genes that were identified using the key word "*kinase*" but not with the word "kinase"? Hint: You can easily do this by combining search strategies. Click on "Add Step" then select "existing strategy":



- Select the right strategy from your list of Gene Strategies and combine the strategies with the correct operation:



Which operation did you choose?

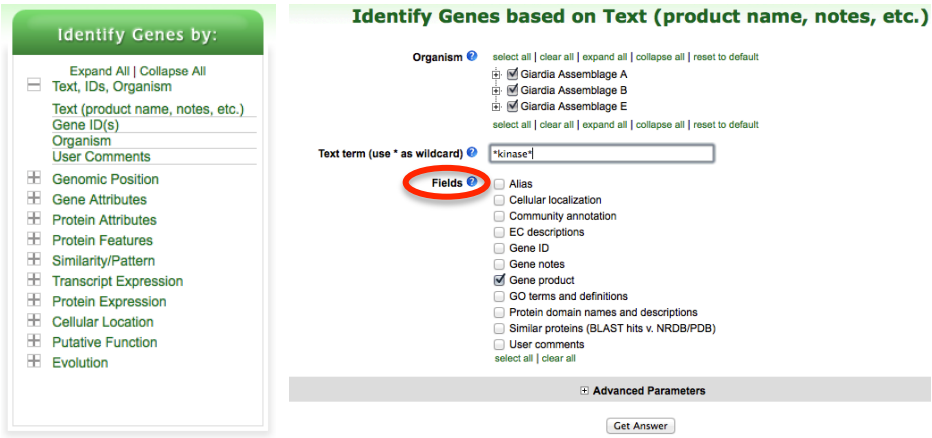


- Do the results make sense? Do all the product names contain the word kinase?

c. **Find only the kinases that specifically have the word "kinase" in the gene product name.**

Hint: Use the text search page, the specific page where you can define the fields to be. There are many ways to navigate to the Text Search page.

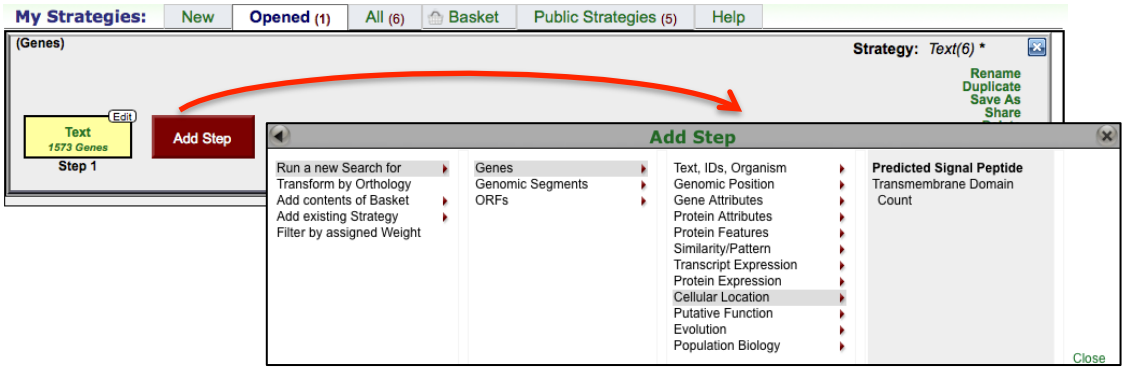
- How did you get there?
- How many kinases have the word kinase in their product names?
- Did you remember to use the wild card?

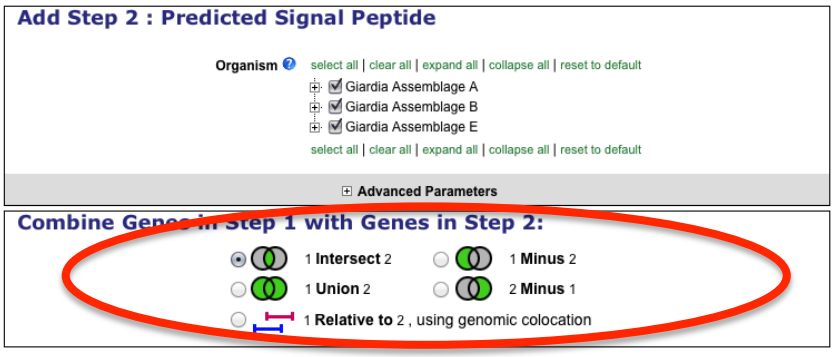


2. Combing text search results with results from other searches

a. In exercise 1 you identified genes that have the word "kinase" somewhere in their product name. Can you now find out how many of these kinases are likely secreted?

Hint: grow your search strategy by adding a step. Choose a search that identifies genes with likely secretory signal peptides. How did you combine the search results?





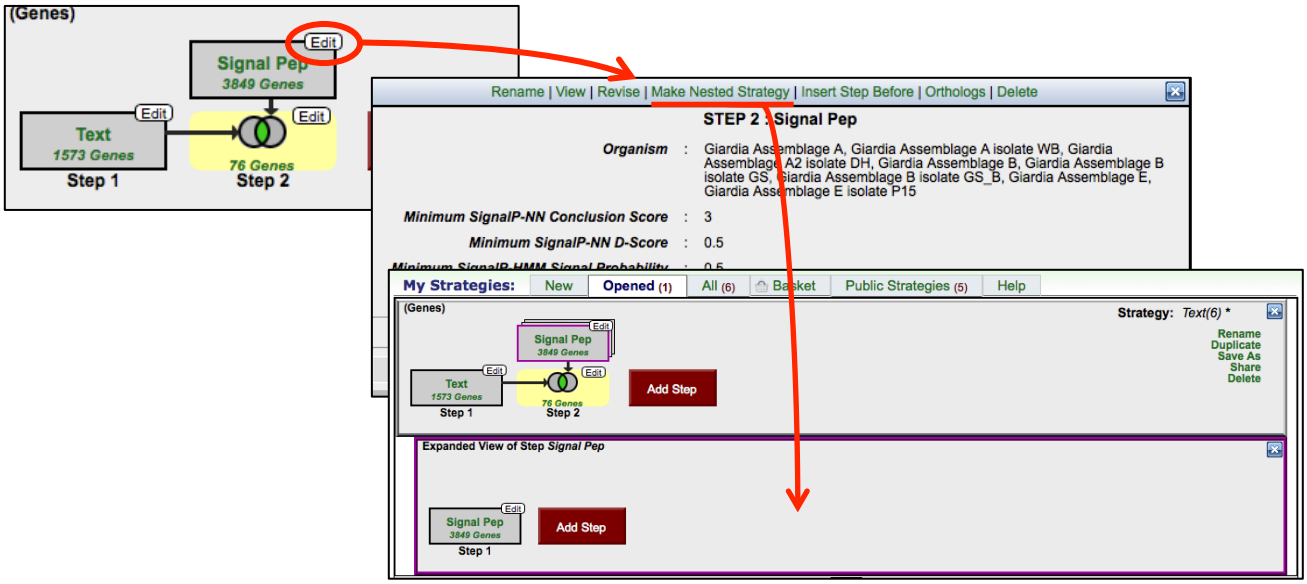
b. **Now that you have a list of possible secreted kinases, how would you expand this strategy even further?**

Which operation did you choose?

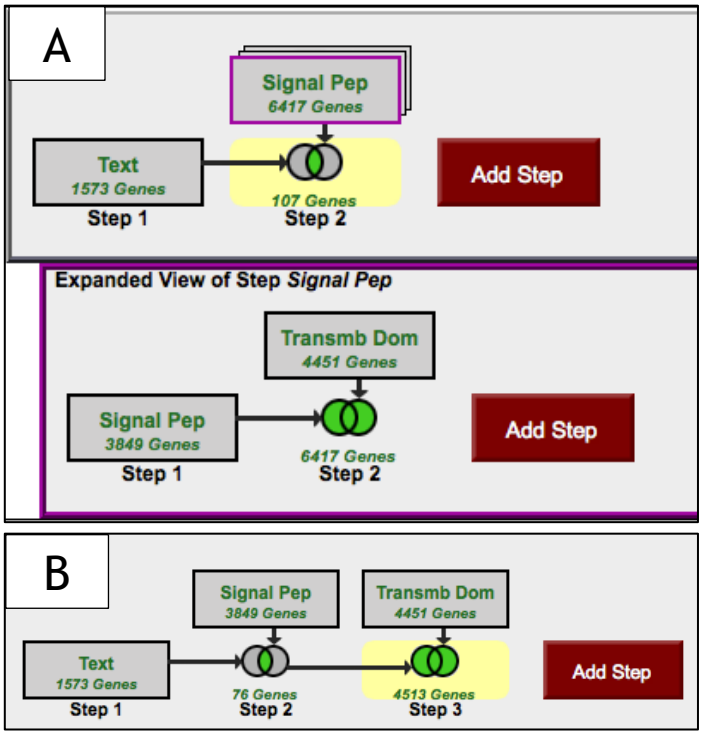
Hint: there is no wrong answer here….
- From a biological standpoint what else would be interesting to know about these kinases? Add more searches to grow this strategy.
- For example, how many of these secreted kinases also have transmembrane domains?

c. **In the above example, how can you define kinases that have either a secretory signal peptide AND/OR a transmembrane domain(s)?**

Hint: to do this properly you will have to employ the "Nested Strategy" feature. Why?



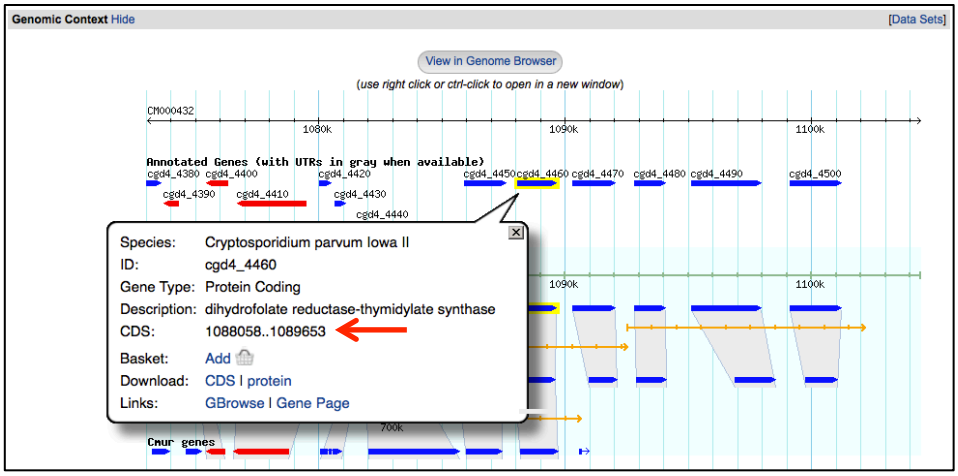
Notice the different results obtained in figures A (with nesting) and B (without nesting) below:



3. **Visiting a specific gene page (http://cryptodb.org).**

a. Find the bifunctional dihydrofolate reductase-thymidylate synthase (DHFR-TS) gene *C. parvum*.

- How did you navigate to this gene? What other ways could you get there? (hint: You can run a text search and look for DHFR or you can search using a specific gene ID: cgd4_4460)
- What chromosome is this gene on?
- How many exons does this gene have? (hint: Look at the gene graphic in the genomic context section, you can also mouse over the graphic to get more information).

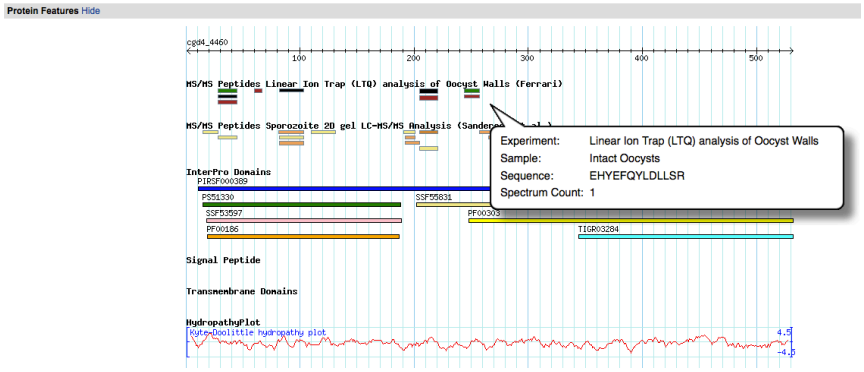


- What direction is the gene relative to the chromosome (ie. left to right or right to left)?
- How many nucleotides of coding sequence does this gene contain?

b. Exploring functional data on the gene page.

- Loosely defined functional data is any data that provides you with information about the function of a gene. This could essentially any data or analysis that adds information about a gene of interest.
- For the *C. parvum* DHFR, scroll down the gene page and explore the various sections (hint: many section are hidden by default. Click on the show link next to a section title to reveal its contents).
    - *Annotation section:* What information can you get from here. Does the *C. parvum* DHFR have orthologs in other *Cryptosporidium* species? What kinds of GO terms are associated with this gene?

- *Protein section:* What does this section tell you? Is the protein expressed? When? Does this gene have any InterPro domains? Does this protein have a similarity to a protein that has been crystallized (PDB = protein data bank)?



- *Expression section:* Is this gene expressed at the RNA level? Does the expression change? Do the RNA data coincide with the protein expression data?

c. Assigning function for a hypothetical protein.

- Using the same logic in the above section visit one of the following gene pages (or both):

  - GL50803_10524: http://giardiadb.org/gene/GL50803_10524
  - cgd5_4250: http://cryptodb.org/gene/cgd5_4250

- Based on the data presented on the gene page can you assign a function to the gene?