Finding Genes, Building Search Strategies and Visiting a Gene Page

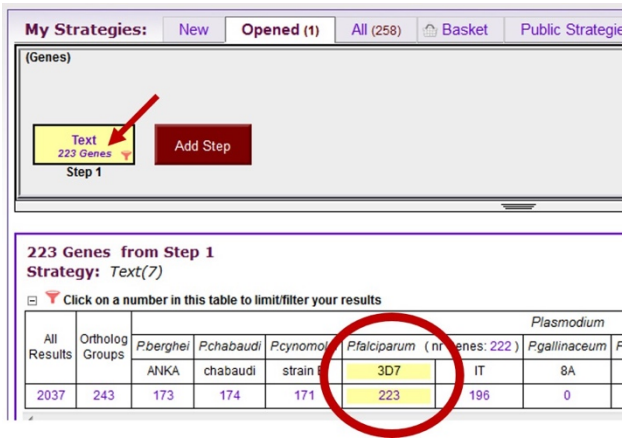
1. Finding a gene using text search.
   For this exercise use http://www.plasmodb.org

   a. Find all possible kinases in *Plasmodium*.

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

   - How many genes did you get?
   - Look closely at the sections of the result page. How many of those are in *P. falciparum*? How did you find this out?

   Hint – the filter table is located between the strategy panel and the result table and shows the distribution of results across the organisms that you searched. Click on a number to display on that species' portion of the results.



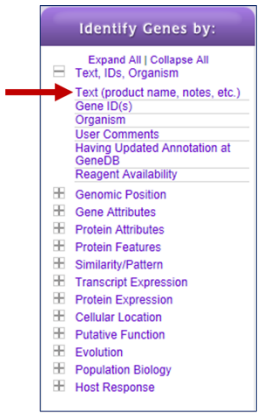
   - What happens if you search using the term **kinases** in the Gene Text Search box? How many results are returned?

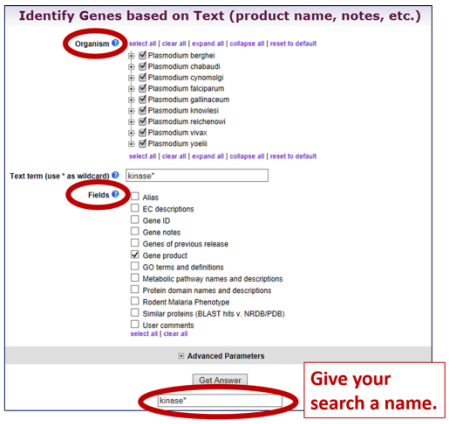
   b. Find only the kinases that specifically have the word "kinase" in the gene product name.

   The search you ran in step 1a using the Gene Text Search box initiates a preconfigured search. Initiating the search from the full text search form - **Identify Genes based on Text**, allows you to configure the search yourself, choosing parameters that best meet your

needs. Use the search form to search for genes that have the word kinase in their **gene product** name/description.

- There are several ways to navigate to the **Identify Genes based on Text** page. Notice the sections of the search page. At the top are parameters and the Get Answer button followed by a search description and a list of datasets used by the search.





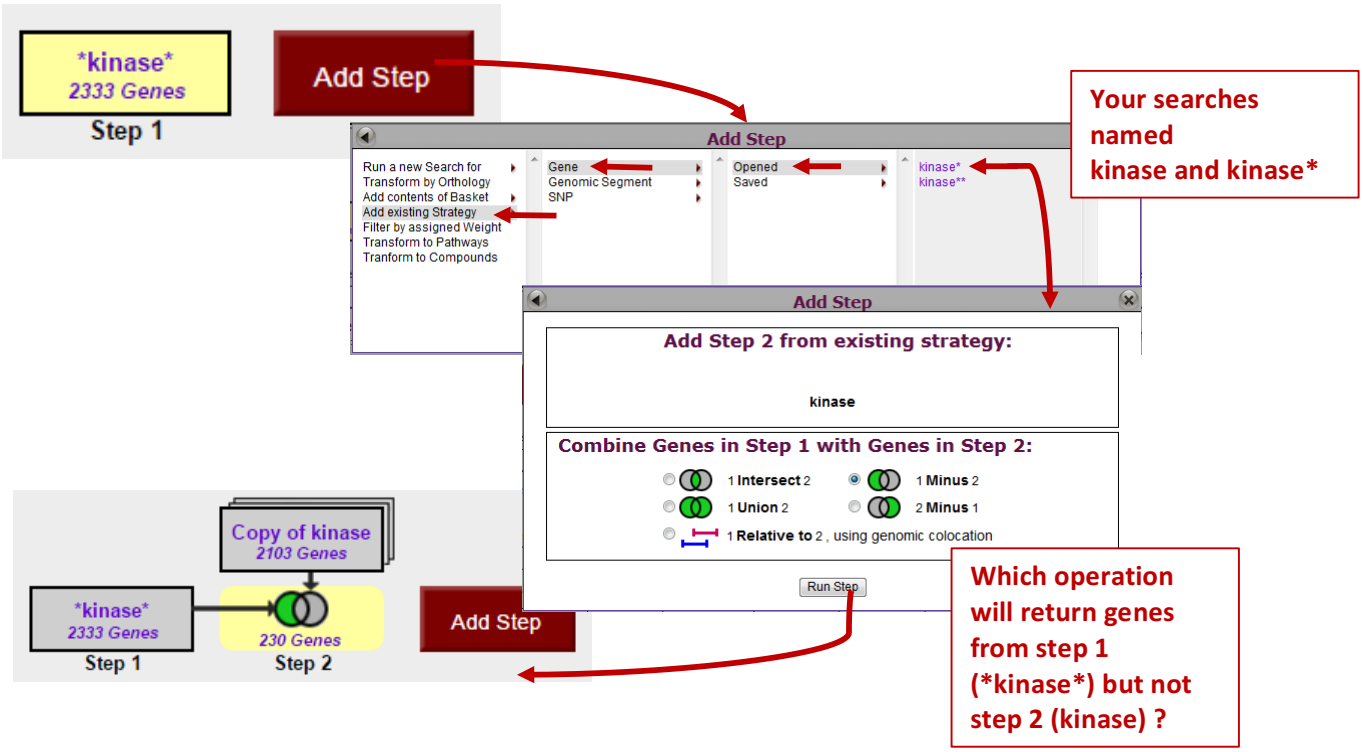
- How can you make sure to find your text term in plural form or in compound words like “kinases” or “6-phosphofructokinase”. Adding a wild card (wildcard = asterisk and means any character) in your search term will broaden your search. Use the full text search, the specific page where you can define the fields to be searched (Fields = Gene Product).

  Try kinase *kinase *kinase*

- **Give each new search a name** to help you keep track of the searches.
- How did you get to the Text Search page?
- How does limiting the number of fields searched affect your results?
- Did you remember to use the wild card?
- How many genes have the word kinase in their product names?

c. **Combine the results of two text searches.**
Find genes that were identified using the key word *kinase* but not the word kinase?

- Here we will build a search strategy that combines 2 of your searches. If you are not displaying the results of the ***kinase*** search (the strategy box will be highlighted in yellow), return to it by clicking on that step box in the strategy panel. To add your **kinase** search to this strategy, click on “Add Step” and select “existing strategy”:
- Select the right strategy from your list of Gene Strategies and combine the strategies with the correct operation. Notice that there is an extra asterisk at the end of an unsaved strategy name. The list of available searches will have an * at the end of the name.



- Do the results make sense? Do all the product names contain the word kinase? From the result page look at the table of gene IDs returned by the search. The Product Description column contains the gene product name.

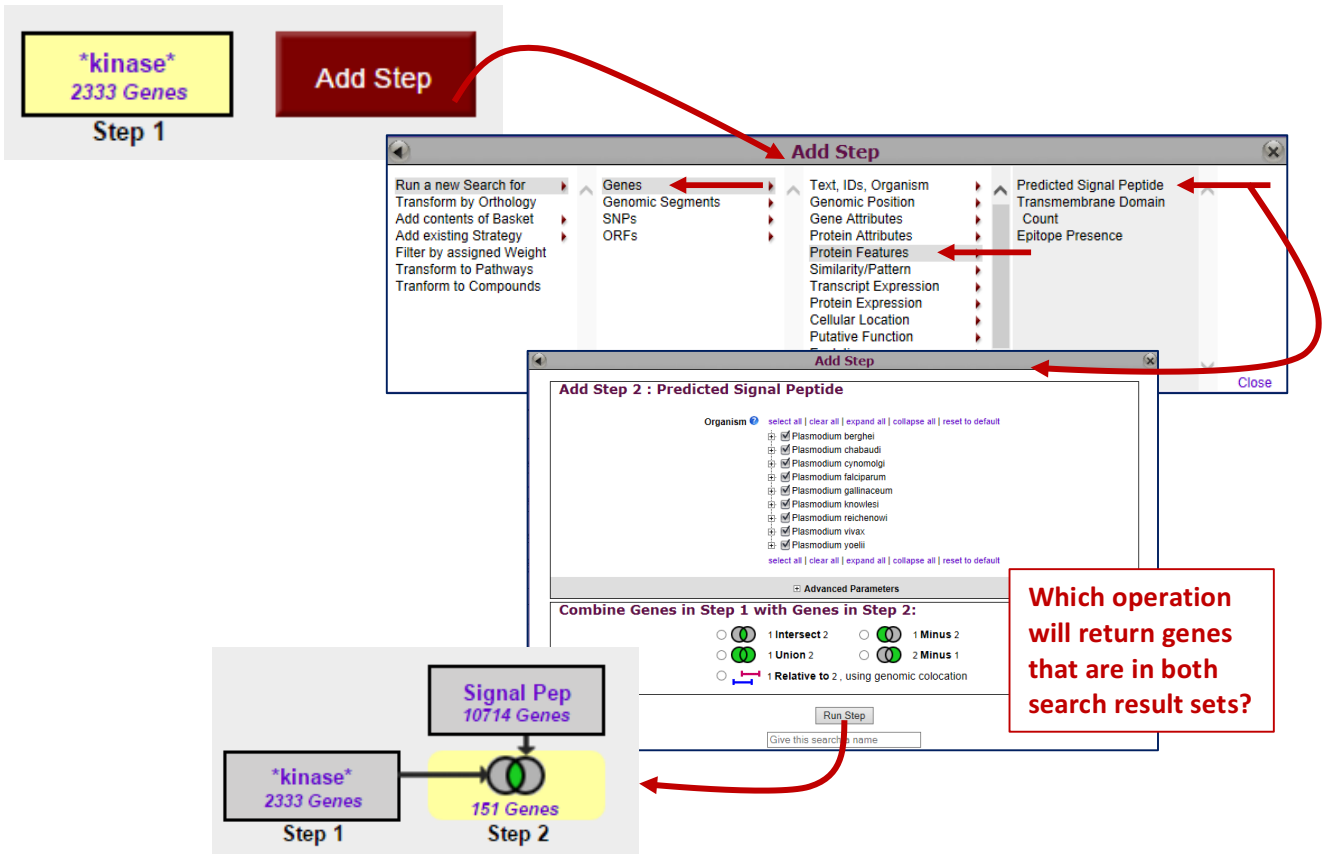
2. Combing text search results with results from other searches

a. Find kinase genes that are likely secreted.

In exercise 1b. you identified genes that have the word **kinase** somewhere in their gene product name (searching *kinase* in gene product field). Grow your search strategy by adding a step that returns genes whose protein products are predicted to have a signal peptide. In this search you are querying the results of our genome-wide analysis that used the SignalP program to predict the presence and location of signal peptide cleavage sites in amino acid sequences.
http://www.cbs.dtu.dk/services/SignalP/

Focus your Strategies section on the ***kinase*** search and click Add Step. For the second search choose **Identify Genes based on Protein Features, Predicted Signal Peptide**

- How did you combine the search results?
- How many kinases are predicted to have a signal peptide?



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

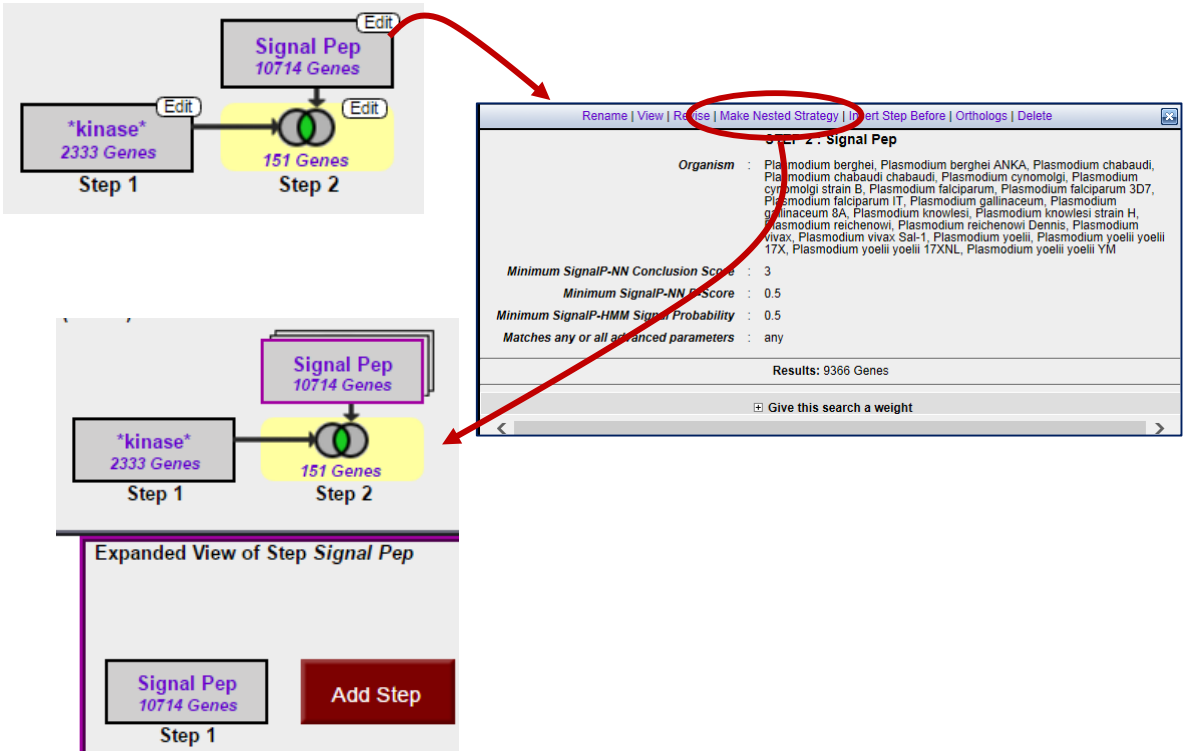
There is no wrong answer here!!

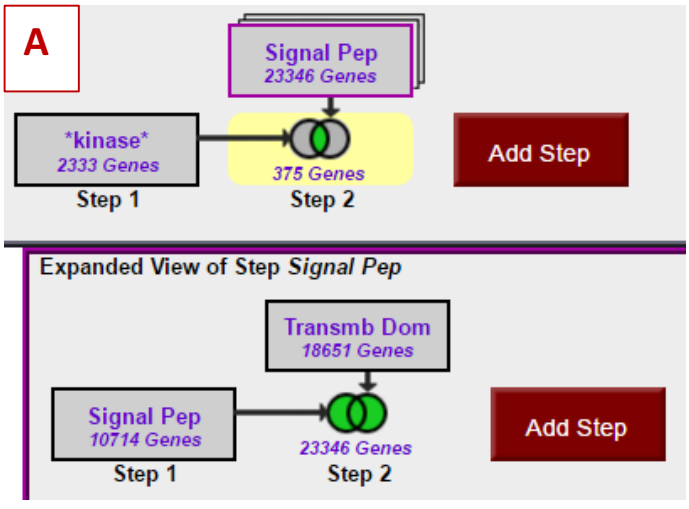
- From a biological standpoint what else would be interesting to know about these kinases? Add more searches to grow this strategy. Open the categories under Identify Genes By: on the home page and explore the types of searches that are available. You can reduce (or expand) your result set by adding searches that are based on many types of data.
- For example, how many of the 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. Nesting a strategy allows you to control the order in which your result sets are combined. Think about the difference between two mathematical equations.

Equation without nesting: 2 x 3 + 5 = 11
Equation with nesting: 2 x (3 + 5) = 16





Strategy Logic:

Every gene returned by Strategy A will be a kinase. These kinases with have a signal peptide OR a TM domain OR both. (SP and/or TM; either or both)

Strategy B returns kinases that have a signal peptide as well as TM domain containing genes.