# **Search Strategies**

***Note:*** *this exercise uses VectorBase* [*https://beta.vectorbase.org*](https://beta.vectorbase.org) *as an example database, but the same functionality is available on all VEuPathDB resources.*

Answerer Key (release 47 - June 2020)[[1]](#footnote-0)

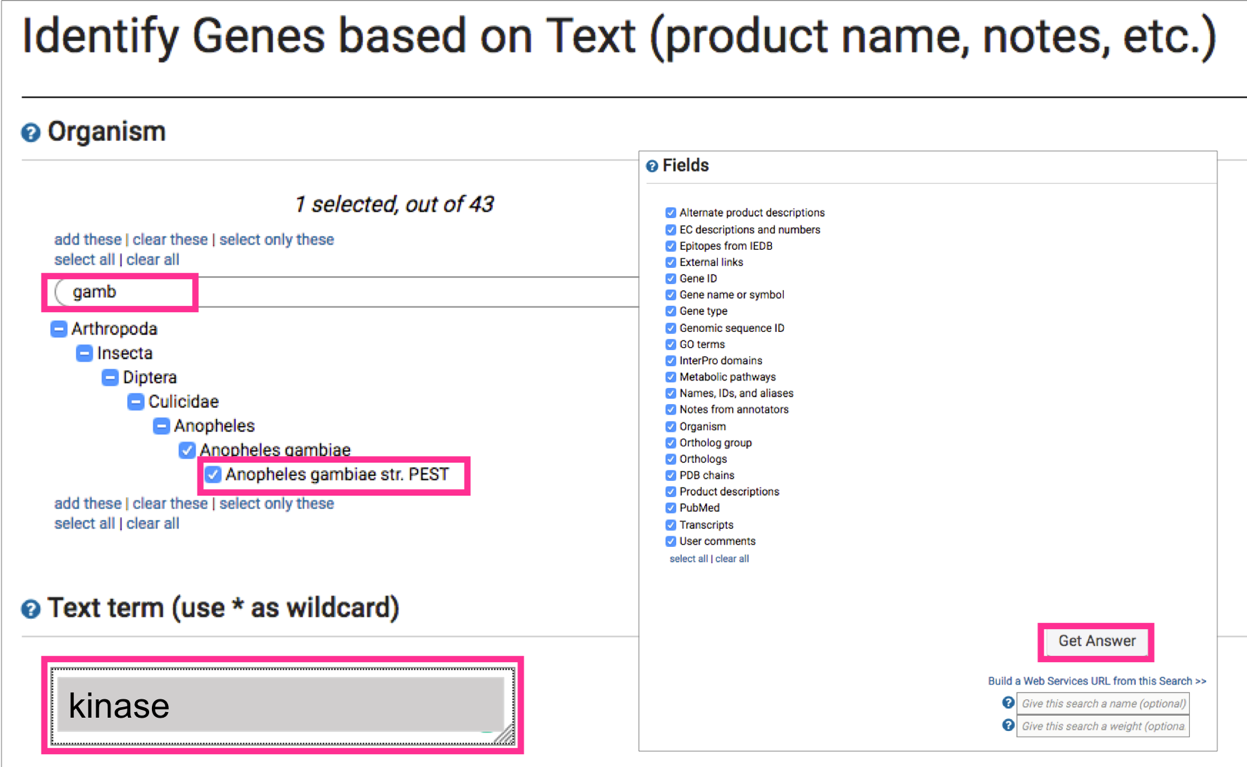
**Learning objectives**:

* Use keywords such as text and gene IDs in the search strategy system
* Configure a query: filters and wild card
* Download query results
* Extend a strategy with more than one step
* Combine strategies

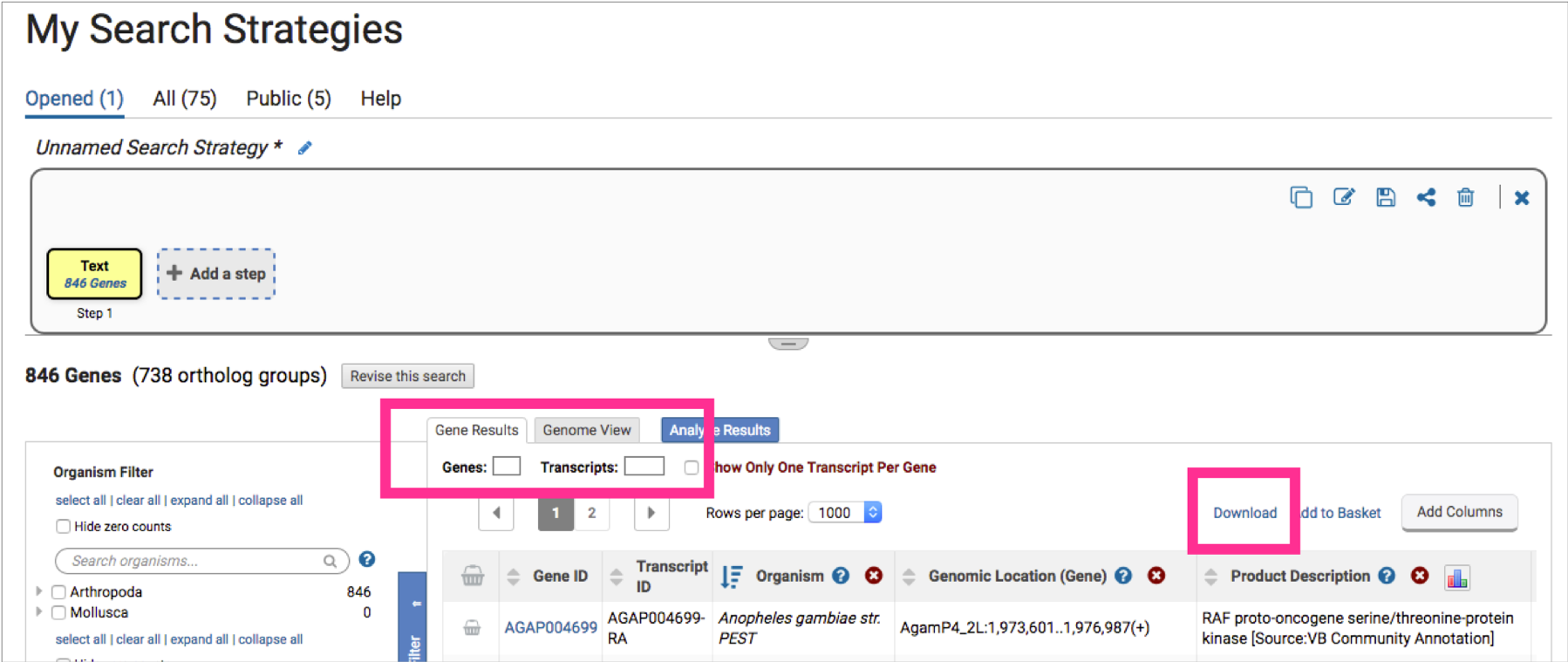
There are three options to start a Search Strategy. 1) From the “Site Search” box ---> Export as Search Strategy, 2) In the site header from the “Searches” menu and 3) In the home page (left hand side) from the “Search for …” section.

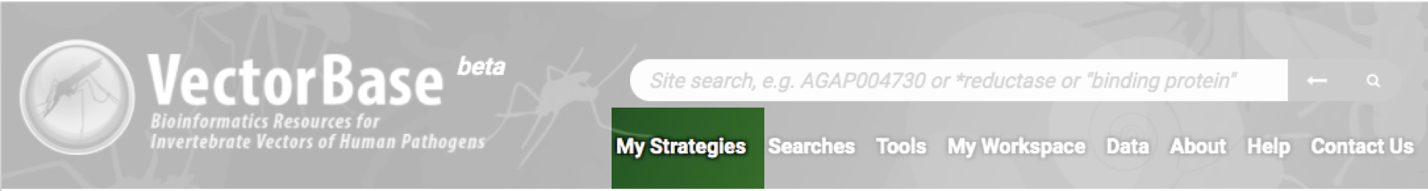
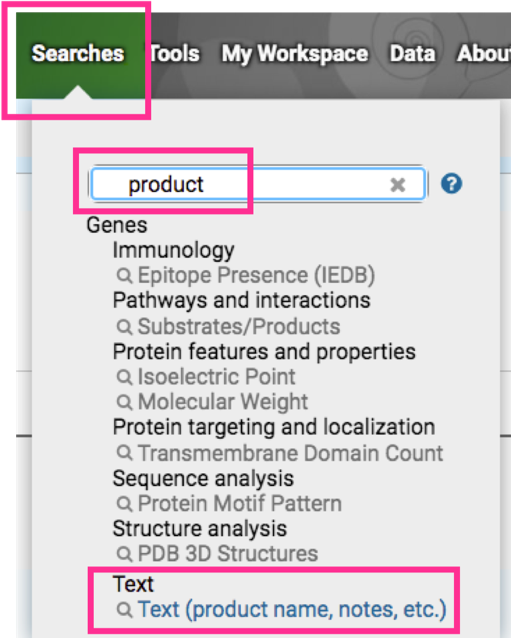
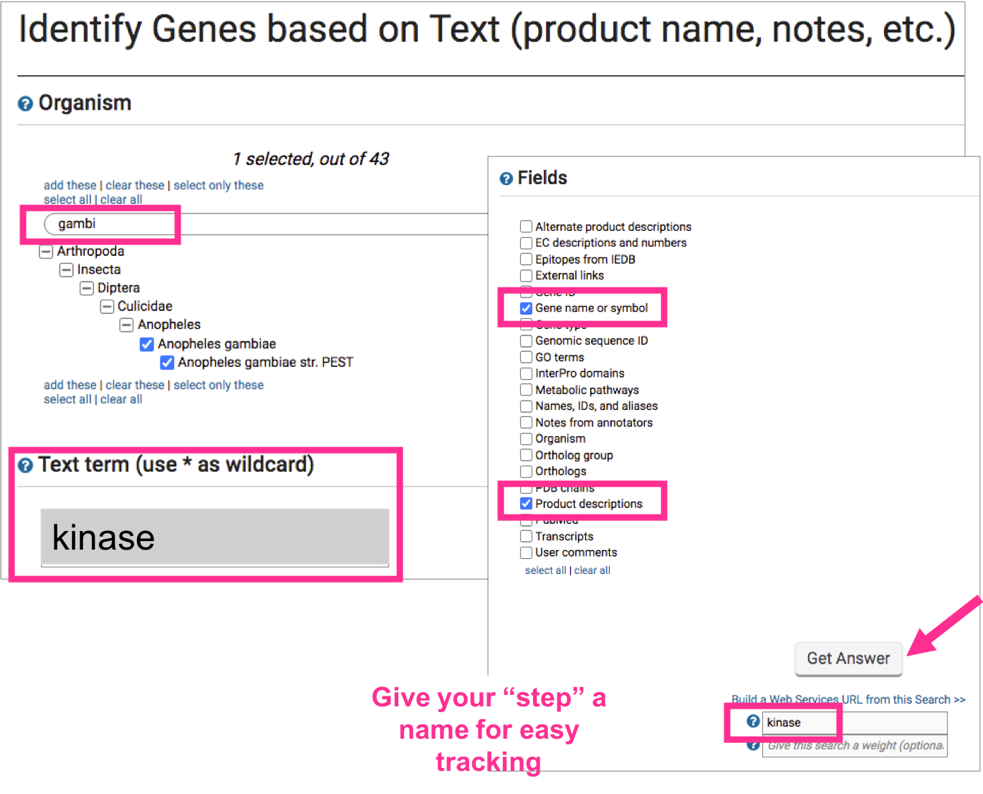
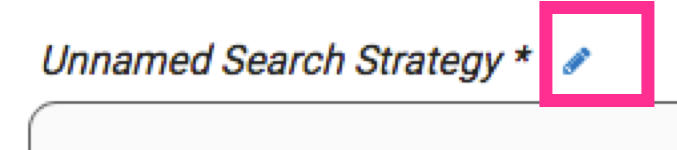
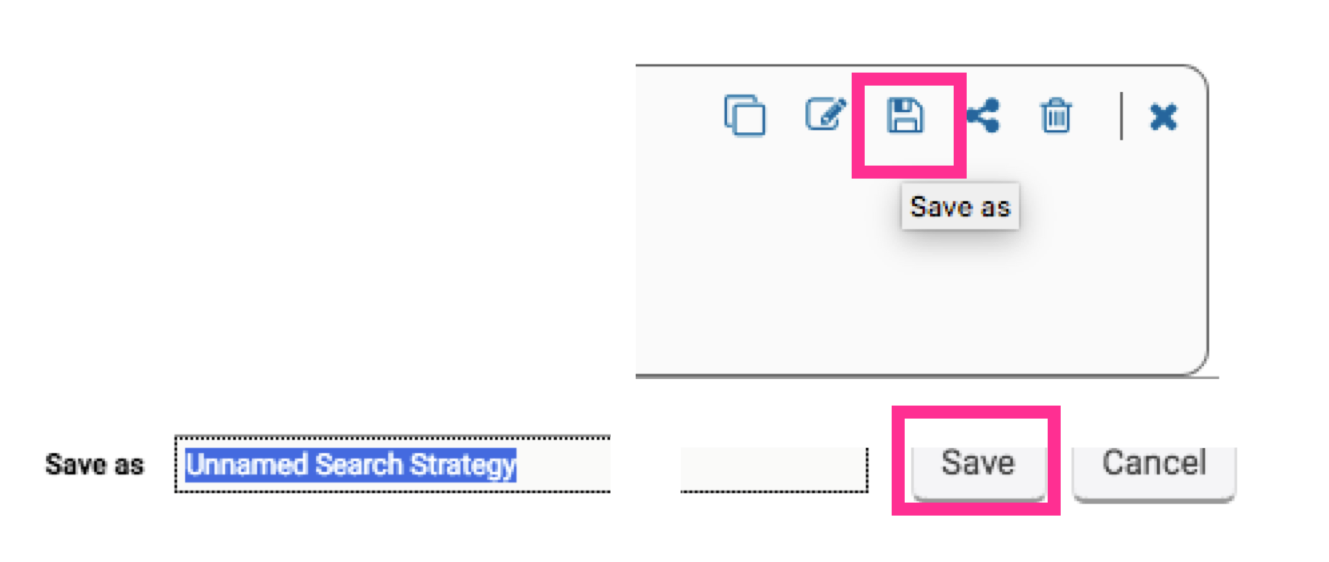


1. Go to the home page and in the “Search for ...” section on the left filter the searches by typing the word *text* to find the ***gene text search*** in the filtered results.A screenshot of a cell phone

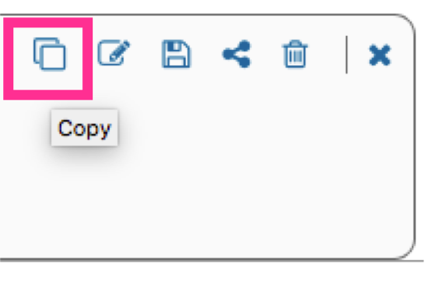
   Description automatically generated
2. You arrive at the “Identify Genes based on Text (product name, notes, etc.)” page. Configure the query as indicated: Select *Anopheles gambiae* str. PEST as the organism, type the word **kinase** in the search box, make sure all fields are selected then click on Get Answer (See image below).
3. How many genes and transcripts did you obtain? (hint: look at the number results section below the search strategy).

|  |  |
| --- | --- |
| Genes: 846 | Transcripts: 1,069 |



1. Click on “Download”. In the “Download Genes” page there are options to configure your file. Select each one of the reports to display its corresponding features.
2. Go back to your strategy with a click on “My Strategies”
3. Now let’s start a query from the “Searches” menu. Search for genes that have the word *kinase* (singular) specifically in the gene name or symbol and product description or function (hint: click on Searches > type the word product > select the Text search) 
4. Notice how in the image above with a different keyword (product) we arrive at the same query as before (step 2, keyword *text*).
5. Configure the search parameters as described below. 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 data sets used by the search.
6. Compare your results from the two gene text searches you ran (step 3 and 9 in this exercise: Notice how in the second one we named only the “step” (kinase), not the strategy. 
7. Giving names to steps is different from saving a strategy. There is an asterisk at the end of an unsaved strategy.
8. Click on “Save as” and give it a name, *e.g*., kinase 
9. How many genes did you get after running the second gene text search? Why is this number different from the first text search?

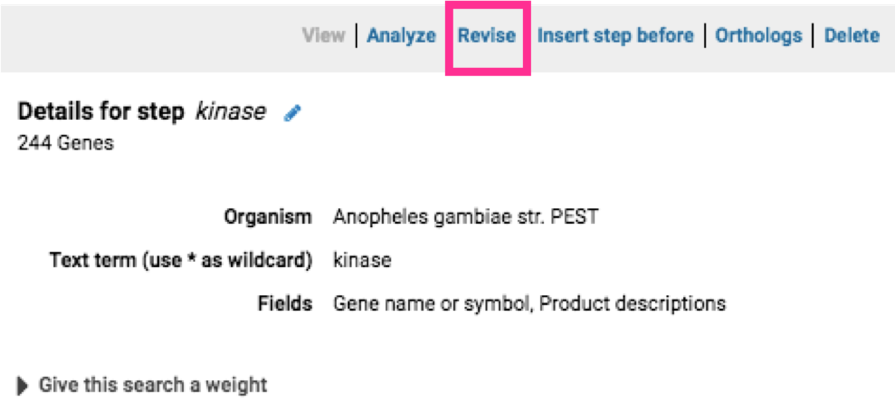
|  |
| --- |
| 244, because in the second search we did not select all fields |



1. How can you find your text term in plural form *kinases* or in compound words like *6-phosphofructokinase*? You can do this by copying the strategy.

In the newly copied strategy, hover with the mouse and the edit option will be visible.



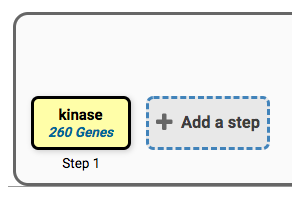
To modify the parameters, click on revise.

Add a wild card (wild card = asterisk and means any character) in your search term will broaden your search (you will need them later). Give each new search a name to help you keep track of the searches. Repeat (copy, edit and revise) to complete the table

|  |  |  |
| --- | --- | --- |
| kinase | \*kinase | \*kinase\* |
| 244 | 259 | 260 |

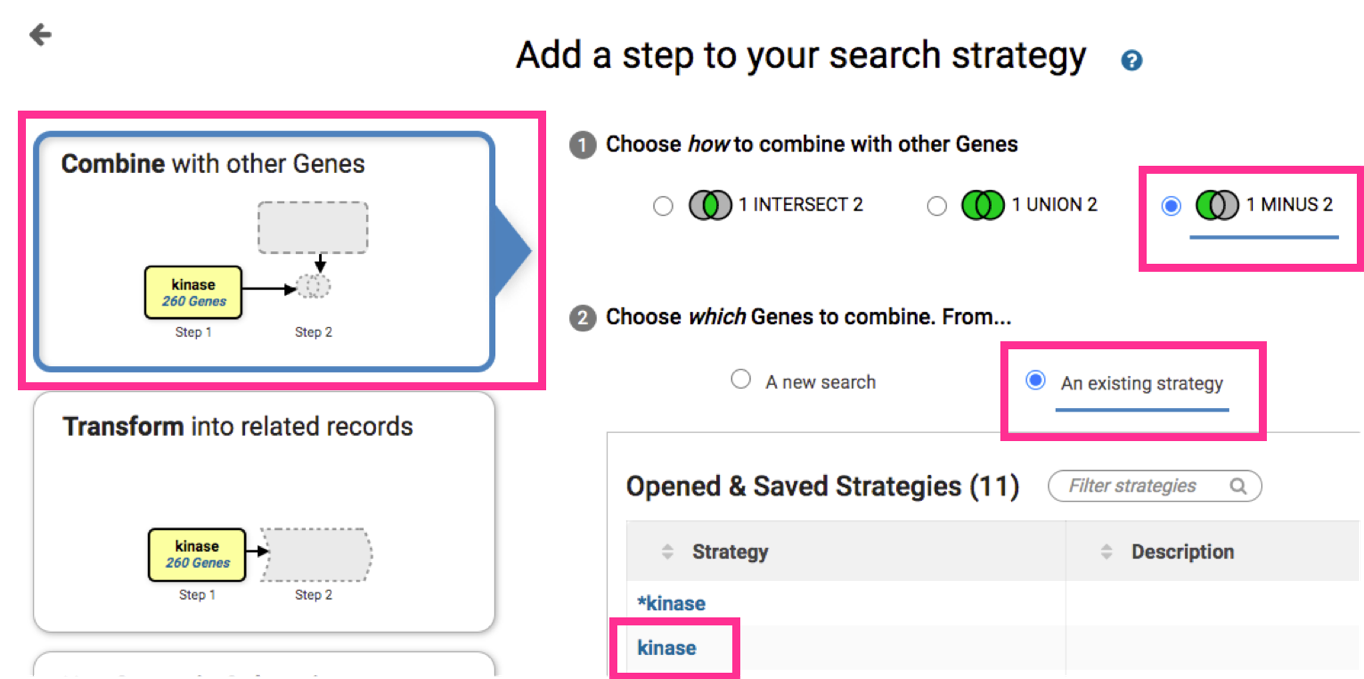
1. How does limiting and increasing the number of fields searched affect your results?

|  |
| --- |
| Hits are narrowed to categories of interest |

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

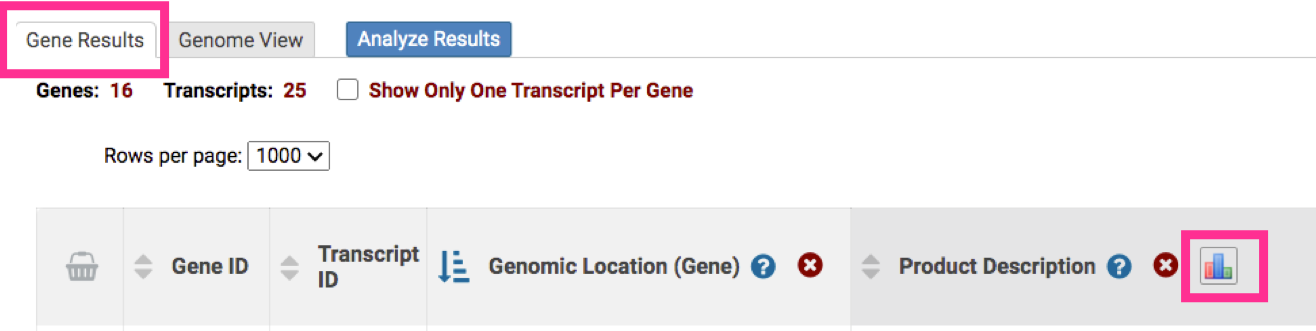
To add your kinase search to this strategy, click on “Add Step”

Configure the search parameters as described below. The “1 minus 2” operator will return genes from step 1 (\*kinase\*) but not from step 2 (kinase)



1. In the results page look at the “Gene Results” Tab with the table of gene IDs returned by the search (image below). The “Product Description” column contains the gene function. Do the results make sense? Do all the product names contain the word kinase?

|  |
| --- |
| yes |

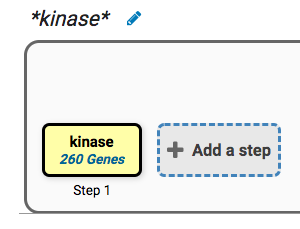
1. In this case the number of hits is 16 and we can quickly and easily look at the product description for all genes. For longer lists use the word cloud in the same column.
2. Previously you identified genes that have the word *kinase* somewhere in their gene product name (searching \*kinase\* in the gene product field). Grow your search strategy by adding a step that returns genes whose protein products are predicted to have a signal peptide.

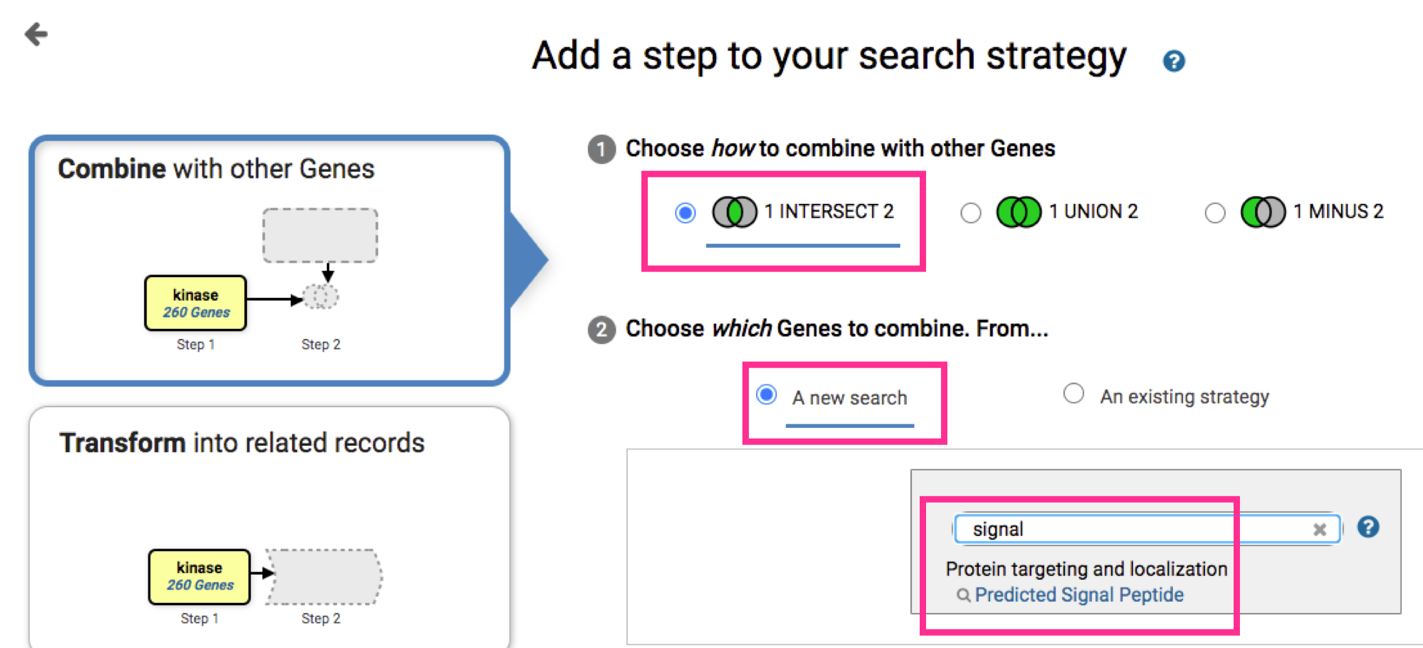
VectorBase (and all other VEuPathDB sites) run pipelines to (*in silico*) predict features of interest for all our genomes. For example, the SignalP program is used to predict the presence and location of signal peptide cleavage sites in amino acid sequences <http://www.cbs.dtu.dk/services/SignalP/> . For more details about other programs used visit our methods page:

<https://beta.vectorbase.org/vectorbase.beta/app/static-content/landing.html>

1. Click on My Search Strategies > All 

Scroll to the “Saved Strategies” and select \*kinase\* (which you saved in step 13)



1. Focus your strategies section on the \*kinase\* search and click “+Add a step”
2. For the second strategy step, configure the search parameters described below and select “Predicted Signal Peptide”. The operator “1 intersect 2” returns genes that are in both search results sets, step 1 and step 2, respectively 
3. How did you combine the search results? Think of the rationale behind it, especially the intersect operator, pretending you are explained it to somebody else

|  |
| --- |
| Imagine for a moment that you have an excel file with two columns, step 1 is column A and step 2 is column B; column C is step 3, the genes in common among both columns |

How many kinases in *A. gambiae* are predicted to have a signal peptide?

|  |
| --- |
| 9 |

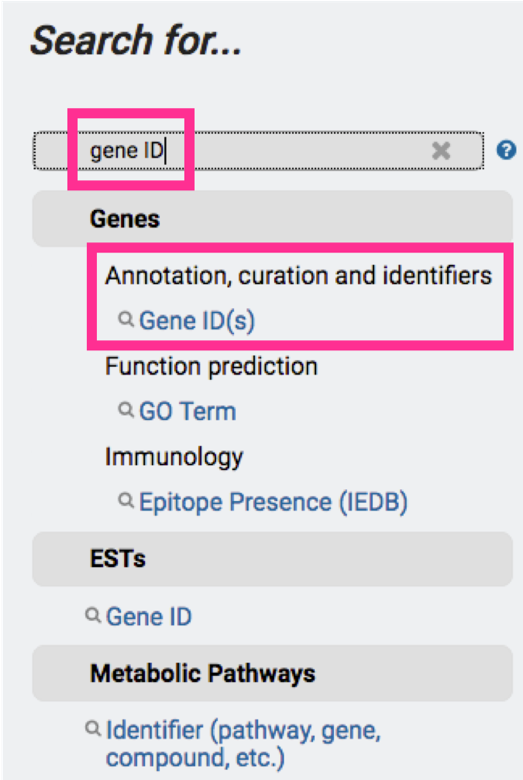
Help guide

|  |  |
| --- | --- |
| **Operator** | **Combined Result will contain** |
|  | IDs in common between the two lists |
|  | IDs from list 2 and list 3 |
|  | IDs unique to 2 |
|  | IDs unique to 3 |
|  | IDs whose features are near each other (collocated) in the genome |

1. How many of the secreted kinases also have transmembrane domains?

|  |
| --- |
| 3 |

Copy this strategy. Edit or delete previous steps. Add more searches to grow this strategy. Open all the drop-down menus and explore all possibilities. You can reduce (or expand) your result set by adding searches that are based on many types of data.



1. You can also use a list of gene IDs to extend or create a new strategy.

Upload your list of genes (AAEL000598), transcripts (AAEL000598-RA) or protein (AAEL000598-PA) IDs in a plain text format (\*txt)

Try this search with this list of IDs:

AGAP000094

AGAP000124

AGAP000141

AGAP000203

AGAP000203

AGAP000243

AGAP000248

AGAP000292

AGAP000421

AGAP000470

AGAP000483

AGAP000483

AGAP000483

AGAP000601

AGAP000601

AGAP000617

AGAP000617

AGAP000776

1. The answers might change in future releases. VectorBase (and all other VEuPathDB sites) data are updated every two months. For information about the most recent release please visit: <https://beta.vectorbase.org/vectorbase.beta/app/static-content/VectorBase/news.html#VectorBase47Released> [↑](#footnote-ref-0)