

# Search strategies in FungiDB

## 1. Finding genes using the Gene Text Search box

The Gene Text query searches gene records for the text you specify. You can use the text search from two places. At the top right side of any FungiDB webpage (this searches all genomes and a predefined set of fields within the gene record) or from the gene text search page where you are able to specify the organisms and fields to search.

The screenshot illustrates the search process in FungiDB. On the left, the 'Initiate a search' section shows the 'Gene Text Search' box with the keyword 'actin'. The search is processed through 'Records that are being searched', which includes 'Taxonomy' (125 selected, out of 125) and 'Associated records' (a list of fields like Alias, Community annotation, etc.). The 'Output' is a 'Results table' containing a list of species where the criteria were matched.

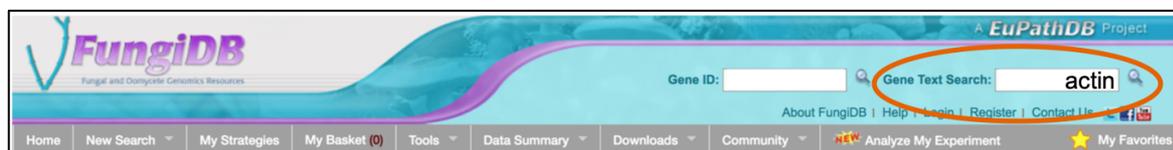
Gene ID	Transcript ID	Organism
NCU04173	NCU04173-t26_1	N. crassa OR74A
NCU04173	NCU04173-t26_2	N. crassa OR74A
SMAC_04416	SMAC_04416-t26_1	S. macrospora k-hell
PITG_02462	PITG_02462-t26_1	P. infestans T30-4
AN2756	AN2756-T	A. nidulans FGSC A4
AN6542	AN6542-T	A. nidulans FGSC A4
AN12237	AN12237-T	A. nidulans FGSC A4
AO90701000065	AO90701000065-T	A. oryzae RIB40
CNBB1210	CNBB1210-t26_1	C. neoformans var. neoformans B-3501A
CNBB1850	CNBB1850-t26_1	C. neoformans var. neoformans B-3501A
NCU01756	NCU01756-t26_1	N. crassa OR74A
NCU07171	NCU07171-t26_1	N. crassa OR74A
YDL029W	YDL029W-t26_1	S. cerevisiae S288c
YFL039C	YFL039C-t26_1	S. cerevisiae S288c
YIL034C	YIL034C-t26_1	S. cerevisiae S288c
YJR065C	YJR065C-t26_1	S. cerevisiae S288c
AN4998	AN4998-T	A. nidulans FGSC A4
AN8831	AN8831-T	A. nidulans FGSC A4
P174DRAFT_439960-137_1	P174DRAFT_439960-137_1	A. novofumigatus IBT 16906
AO90009000235	AO90009000235-T	A. oryzae RIB40

### a. Find all genes containing the term actin.

- Use the keyword **actin** in the **Gene Text Search** box (top right of the webpage) and click on the search icon or simply press the Enter key.

Note: When using this shortcut, the search identifies your input term in all FungiDB organisms and in a predefined lists of gene record fields (e.g. Gene name, description, user comments, GO terms, etc.).

The text search in FungiDB is an exact match search, which makes the wild card option particularly useful when trying to find complex terms (see below).



- How many genes were returned by your search?

The screenshot shows a search result box with a yellow background. It displays 'Text 11591 Genes' and a red 'Add Step' button. Below the text, it says 'Step 1'.

Note: Due to frequent database updates, the results shown here may not match exactly to your search result.

- How many of those genes are in *Aspergillus nidulans*?

Note: Use the filter table located between the strategy panel and the result table that shows the distribution of results across the various organisms. Scroll to the right, click on a number under the *Neurospora crassa* to 'filter' the results.

The screenshot shows a search interface with a filter table for Eurotiomycetes. The table has columns for various species and their corresponding gene counts. The 'Aspergillus nidulans' column is highlighted with an orange box, showing a count of 155. Below the filter table is a 'Gene Results' section with a table containing columns for Gene ID, Transcript ID, Organism, Genomic Location (Gene), Product Description, Found in, and Score. The first few rows of the Gene Results table are visible, showing genes from *A. candida* and *Acan2VRRR*.

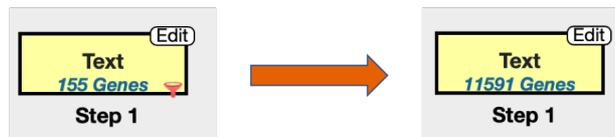
- Do you believe all of these genes encode actin?

- Find the **Product Description** and the **Found in** columns in the Gene Results table to understand your results.

Note: Some genes are annotated with GO terms (internal curation or uploaded from the GO consortium), while other genes have predicted protein function annotated with domains from InterPro scan, product description from the gene model functional annotations, etc. Understanding why these genes are associated with the term actin is important when interpreting your results.

- Can you revert the display back to showing all genes across all organisms?

Hint: Scroll back to the beginning of the horizontal display and click on the **All Results** number in blue. Note that the search strategy display has also changed (hour glass is no longer visible), which means that you are no longer displaying species-specific results.



**b. Search ubiquitin-related gene products (e.g. ubiquitin carboxyl-terminal hydrolase, ubiquitin-protein ligase, small ubiquitin-like modifier) using a wild card**

Rather than searching for individual terms, try using a wild card (\*).

- Search for **ubiquitin\*** in *Gene product in Oomycetes*
  - How many genes were returned?
  - What happens if you search for **\*ubiquitin\***?

The screenshot shows a search filter panel. Under the 'Organism' section, 'Oomycetes' is selected. Under the 'Text term (use \* as wildcard)' section, the search term '\*ubiquitin\*' is entered. Under the 'Fields' section, 'Gene product' is selected.

### c. Search for a two-word term

- Search for "**Histone H3**" (including quotations) using *Search for Genes > Text* approach and select *Oomycetes* in the *Organism* search parameter and leave *Fields* selections at default.
  - How many results are returned?
  - Notice that hypothetical and unspecified products are also returned in this search? Why do you think that is? How would you go about finding out where the "Histone H3" term may be mentioned for this gene to be included in your search results?

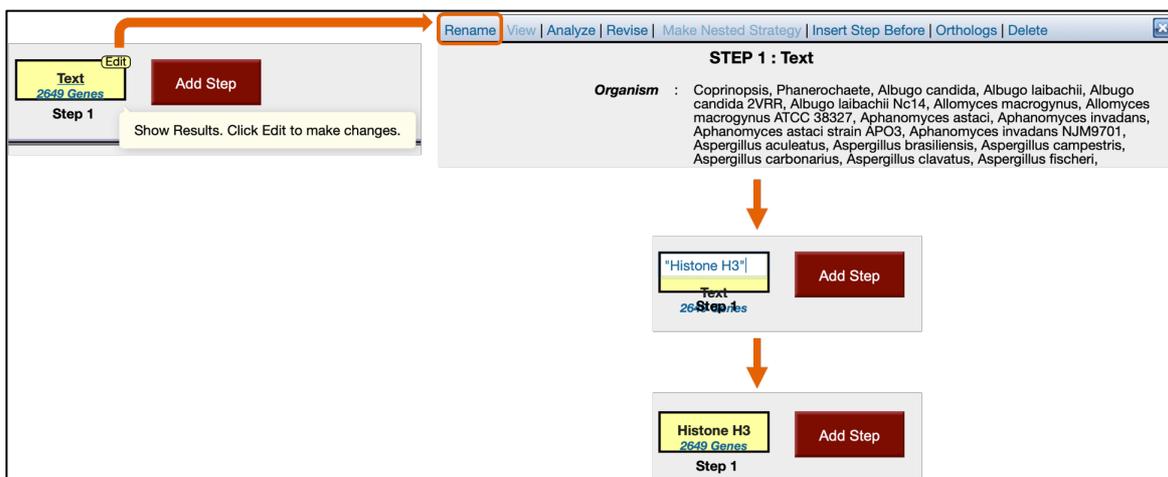
*Hint: Look in the Found in column for the source of Histone H3 annotation. Navigate to the gene record page by clicking on the GeneID to learn more about why this gene has been returned in your search. Try using web text search (Ctrl F (Windows); Command F (Mac)) to see where Histone H3 is mentioned within the gene records page.*

- How would you go about finding out where the "Histone H3" term may be mentioned for this gene to be included in your search results?
- Save your strategy with the name "Histone H3" (you would need to be logged in to your account to do this; also see below).

## 2. Modify an existing search strategy or add a new step

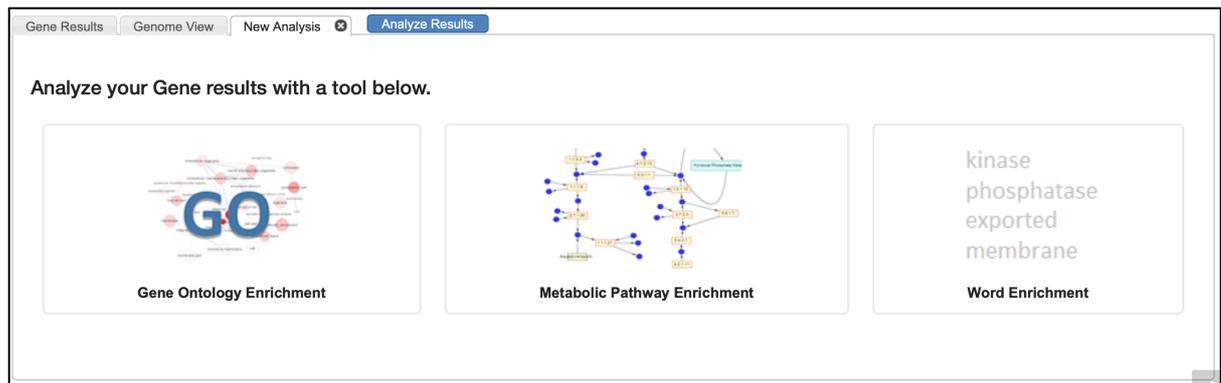
### a. Rename a text search step in the previously saved "Histone H3" strategy

- Click on the step to highlight it in yellow if you navigated away from it.
- Hover over the box to select to reveal the edit link.
- Click on the edit function within the yellow box, then click on 'Rename'
- Click on the Save As link again and then replace the strategy to keep your latest edits.



Other functions of the Edit tool include:

- **Analyze** – redirects you to the Analyse results section:



- **Revise** – revise your strategy. Practice revising your strategy (e.g. “Histone H3”) by limiting your search to *Aspergillus* (Fungi) and *Gene description* only:

#### Revise Step 1 : Text (product name, notes, etc.)

##### Organism

26 selected, out of 129

Filter list below...

Oomycetes

Fungi

[select all](#) | [clear all](#) | [expand all](#) | [collapse all](#)

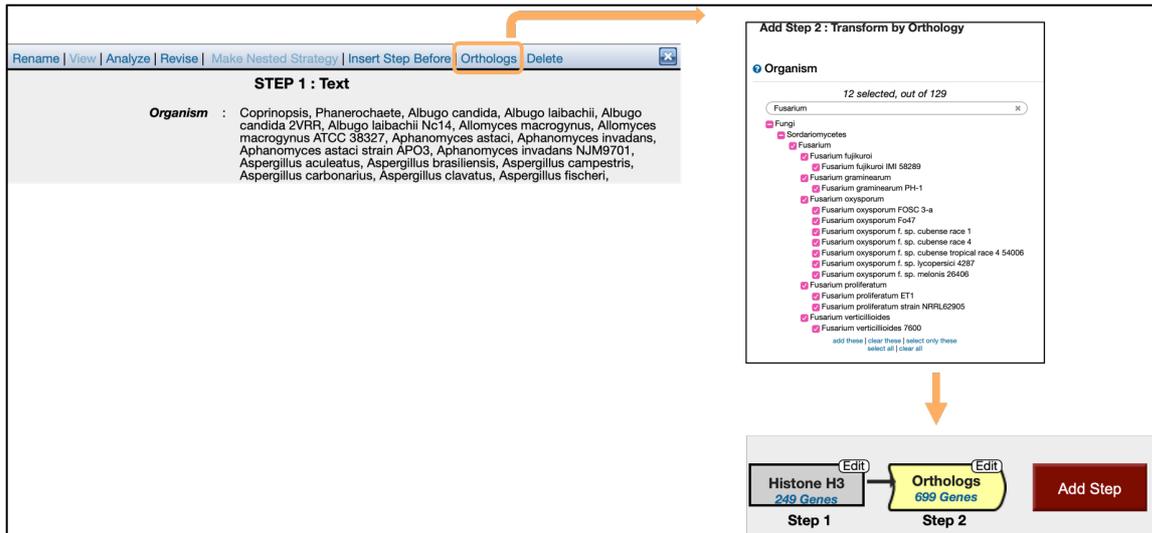
##### Text term (use \* as wildcard)

"Histone H3"

##### Fields

- Alias
- Community annotation
- EC descriptions
- Gene ID
- Gene notes
- Gene product

- **Insert Step Before** – allows to revise your strategy by adding steps before the step you are modifying
- **Orthologs** – allows to add a step after the one you are modifying that will find orthologs according to your selections in the orthologs parameter pop-up window. For example, choose to find all orthologs of “Histone H3” results in *Fusarium*:



- **Delete** function deletes the step that is being modified

### 3. Finding genes using the Gene ID search box

Let's say you are working with *Candida auris* and studying the process of biofilm formation. B9J08\_001462, annotated as hypothetical protein product, piqued your interest because it came up in your genetic screen and you would like to know more about the function of this gene.

- Enter B9J08\_001462 in the *Gene ID* search box and press Enter:



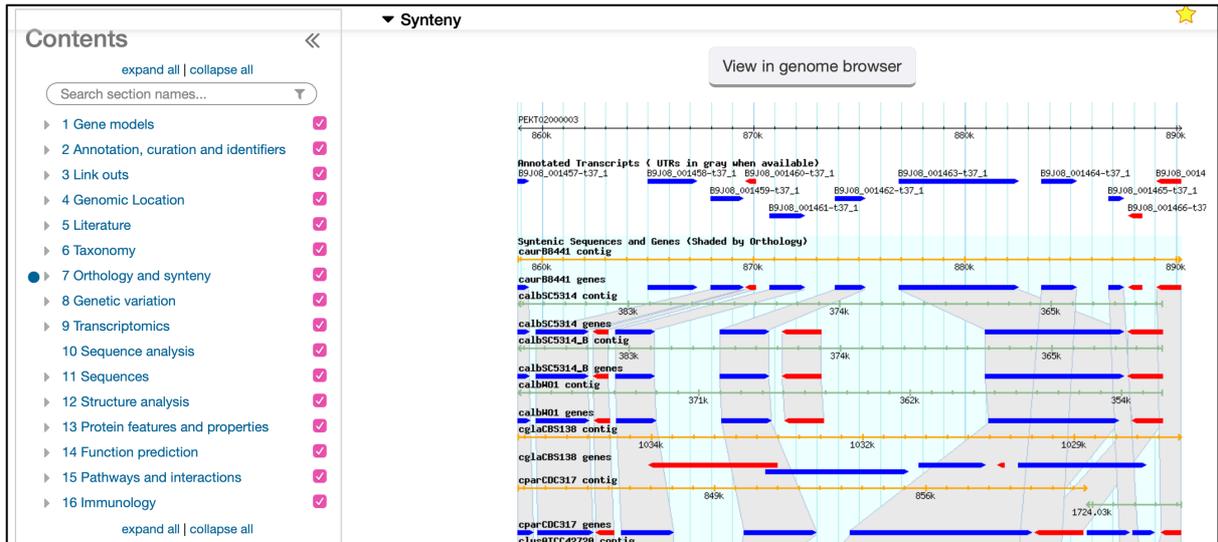
- Examine the **Orthology and Synteny** section of the gene record page.
  - Does the provided information support its putative function as a cyclin?

*Hint: Navigate to the Orthology and Synteny section of the gene record page and examine orthologs' function.*

- What can you say about this gene's expression in biofilm in *C. auris*?

*Hint: Navigate to the Transcriptomics section of the gene record page to evaluate the expression of this gene in biofilm forming conditions (plank = planktonic cells; 4, 12, 24hr = biofilm forming conditions).*

- Does this gene have syntenic orthologs in *Candida* species?



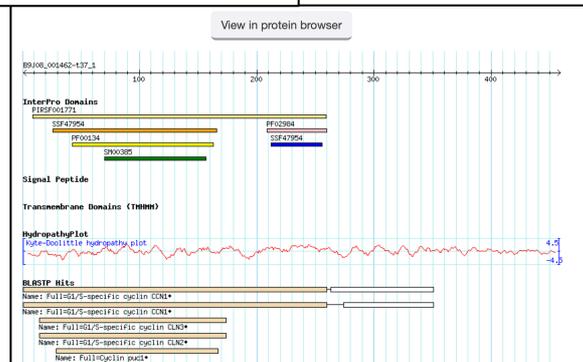
- Does the protein have any identifiable domains?

Hint: Use the Contents menu section to navigate to the **Protein features and properties** section of the gene record page.

▼ InterPro Domains [Download](#) [Data sets](#)

Search this table...  Showing 6 rows

Transcript ID(s)	Name	Interpro ID	Primary ID	Secondary ID	Description	Start Min	End Min	E-value
B9J08_001462-137_1	PFAM	IPR006671	PF00134	Cyclin_N	Cyclin, N-terminal domain	43	162	5.0E-39
B9J08_001462-137_1	PFAM	IPR004367	PF02984	Cyclin_C	Cyclin, C-terminal domain	209	259	6.7E-05
B9J08_001462-137_1	PIRSF		PIRSF001771		Cyclin, A/B/D/E types	9	258	1.7E-38
B9J08_001462-137_1	SMART	IPR013763	SM00385	cyclin_7		70	156	2.3E-22
B9J08_001462-137_1	SUPERFAMILY	IPR036915	SSF47954	0054835	Cyclin-like	26	165	1.7E-38
B9J08_001462-137_1	SUPERFAMILY	IPR036915	SSF47954	0054835	Cyclin-like	212	255	2.8E-05



- Cross-reference the search results with what is known in other model organisms.
  - Navigate to the *Orthology and Syteny* section
  - Click on the *CR\_02920C\_A* in *Candida albicans SC5314* to be redirected to an ortholog gene record page.
  - Navigate to the *Link outs* section on the *C5\_01680C\_A* gene record page and click on the *CGD External ID* link

### 3 Link outs

External Links [Download](#) [Data sets](#)

Search this table...  Showing 3 rows

Database	External ID
CGD	<a href="#">CR_02920C_A</a>
Entrez Gene	<a href="#">3642587</a>
The Universal Protein Resource (UniProt)	<a href="#">Q5A1Z4</a>

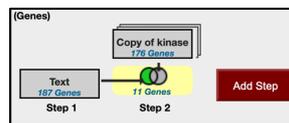
- Navigate to the **Phenotype** tab on the CGD record page for CR\_02920C\_A.
- Is the deletion mutant viable?
- What can you say about growth phenotype (filamentous and invasive growth) or biofilm formation in the null mutant?

*Note: you can also download the phenotype table and open it in Excel or save it for later use.*

## 4. Combining text search results with results from other types of searches

### a. Find kinase genes that are likely secreted.

- Let's go back to the \*kinase\* - **kinase** search strategy.



- **Insert a step** before current Step 2 to identify kinases that have a predicted Signal Peptide.

*Hint: Select to revise the strategy and then navigate to the Insert Step Before tab.*

*Note: Signal Prediction is made with the SignalP program that predicts the presence and location of signal peptide cleavage sites in amino acid sequences.*

STEP 2 - Copy of kinase  
The nested strategy gets opened below.  
Results: 176 Genes

Give this search a weight  
10  
Assign

Run a new Search for  
Transform by Orthology  
Add contents of Basket  
Add existing Strategy  
Filter by assigned Weight

- Genes
- Genomic Segments
- SNPs
- ORFs
- Text
- Gene models
- Annotation, curation and identifiers
- Genomic Location
- Orthology and synteny
- Phenotype
- Genetic variation
- Transcriptomics
- Sequence analysis
- Structure analysis
- Protein features and properties
- Protein targeting and localization
- Function prediction
- Pathways and interactions
- Proteomics
- Immunology
- Predicted Signal Peptide
- transmembrane domain count

Insert Step 2: Predicted Signal Peptide

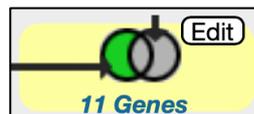
Organism  
1 selected out of 125  
kn  
Fungi  
Tremellomycetes  
Cryptococcus  
Cryptococcus neoformans  
Cryptococcus neoformans var. grubii KN99

Advanced Parameters

Combine Genes in Step 1 with Genes in Step 2:  
 1 Intersect 2  
 1 Union 2  
 1 Minus 2  
 2 Minus 1  
 1 Relative to 2, using genomic colocation

(Genes)  
Step 1: Text (187 Genes)  
Step 2: Signal Pep (975 Genes)  
Step 3: Copy of kinase (176 Genes)  
Add Step

Search operator selection can be edited at any time by clicking on the edit link within the operator window:



Operator	:	Combined Result will contain:
<input type="radio"/> 1 INTERSECT 2	:	IDs in common between the two lists
<input checked="" type="radio"/> 1 UNION 2	:	IDs from list 1 and list 2
<input type="radio"/> 1 MINUS 2	:	IDs unique to 1
<input type="radio"/> 2 MINUS 1	:	IDs unique to 2
<input type="radio"/> 1 Relative to 2	:	IDs whose features are near each other (colocated) in the genome

1

2

The top four operators (1) are used when both searches produce the same type of genomic feature results and therefore are used for direct comparison of similar types of data (e.g. gene list 1 and gene list 2 obtained from proteomics and RNA-Seq evidence searches, respectively).

When two search steps are run against different types of data (e.g. genes and chromosomal sequences or SNPs), the colocation feature (2) allows combining these different types of data (e.g. searching for a list of SNPs or transposable elements colocating to the upstream/downstream regions of genes).

- How many kinases are predicted to have a signal peptide?
- Can you identify genes that have predicted signal peptides but are not kinases?

Hint: Hover over the operator selection to select the Edit function.

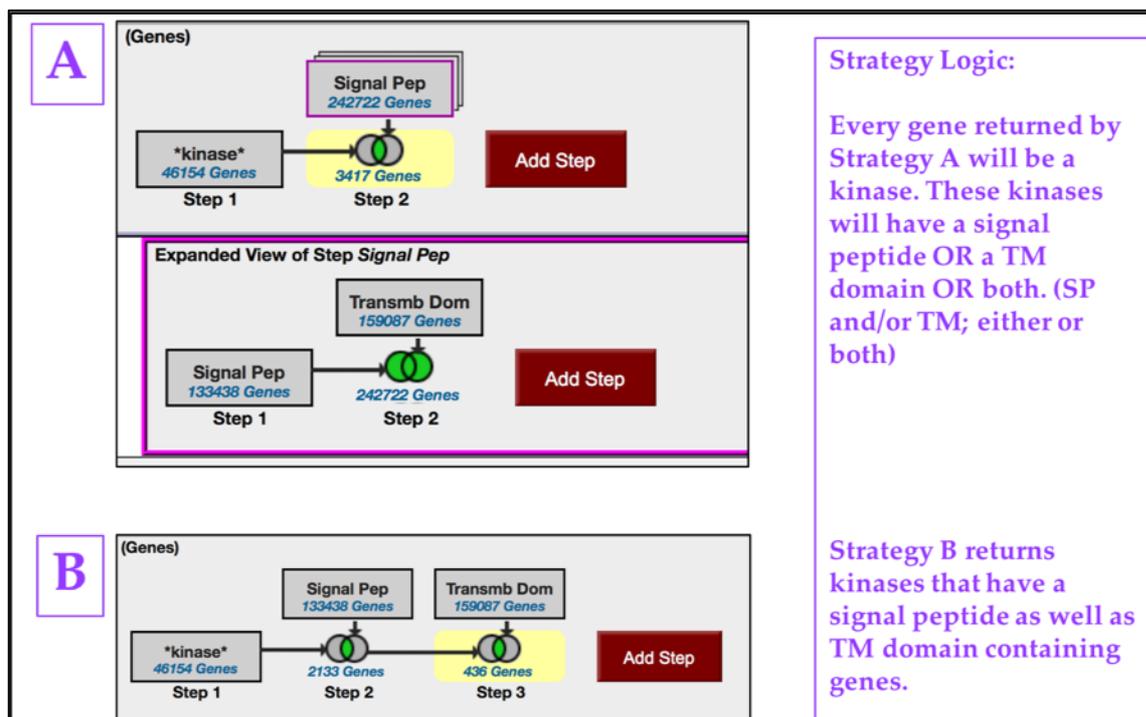
## 5. Creating Nested Strategies

Nesting a strategy allows you to control the order in which your search result sets are combined. Consider the difference between two mathematical equations:

A. Equation with nesting:  $2 \times (3 + 5) = 16$

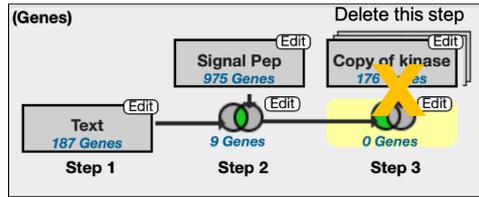
B. Equation without nesting:  $2 \times 3 + 5 = 11$

Here is how the sample equations above translate into search results in FungiDB queries:



Note that nested (A) and non-nested (B) strategies return results that are quite different. The final number of genes returned in nested vs non-nested strategy is **3417** and **436**, respectively.

Let's go back to the search in *C. neoformans* KN99 we just ran and delete the very last step to display the search results for \*kinase\* and signal peptide query:



- How can you define kinases that have either a secretory signal peptide **AND/OR** a transmembrane domain(s) (option A above)?

*Hint: You would need to create a nested strategy because search operators allow combining steps in subsequent fashion only. Start by revising the Step 2 to create a nested strategy as shown below.*

**Add Step 2: Transmembrane Domain Count**

Organism: 1 selected, out of 125  
 Cryptococcus neoformans var. grubii KN99

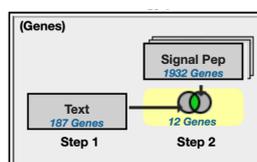
Minimum Number of Transmembrane Domains: 1

Maximum Number of Transmembrane Domains: 99

Combine Genes in Step 1 with Genes in Step 2:  
 1 Union 2  
 1 Intersect 2  
 1 Minus 2  
 2 Minus 1  
 1 Relative to 2, using genomic colocalization

**To select kinases that have either a secretory signal peptide **AND/OR** a transmembrane domain use a union operator**

- Close the **Expanded View of Step Signal Pep** window as shown above, and you should have a nested strategy as follows:



- Save your strategy (rename it if necessary)
- Modify your strategy to scenario B to return kinases that have predicted signal peptide as well as a transmembrane domain.
  - First duplicate the strategy and then delete the second step in the duplicated strategy:

My Strategies: [New](#) **Opened (1)** [All \(664\)](#) [Basket](#) [Public Strategies \(48\)](#) [Help](#)

Hide search strategy panel

(Genes)

Text 187 Genes Step 1

Signal Pep 1932 Genes Step 2

Add

Rename | View | Analyze | Revise | Open Nested Strategy | Unnest Strategy | Insert Step Before | Orthologs | Delete

**STEP 2 : Signal Pep**  
The nested strategy gets opened below.

Results: 1932 Genes

*Hint: To return kinases that have predicted signal peptide as well as a transmembrane domain, add two steps (searching for secreted and transmembrane proteins, respectively).*

- Compare the A and B strategies. Do the results make sense?