

Monday, 10th May 2021

Database queries

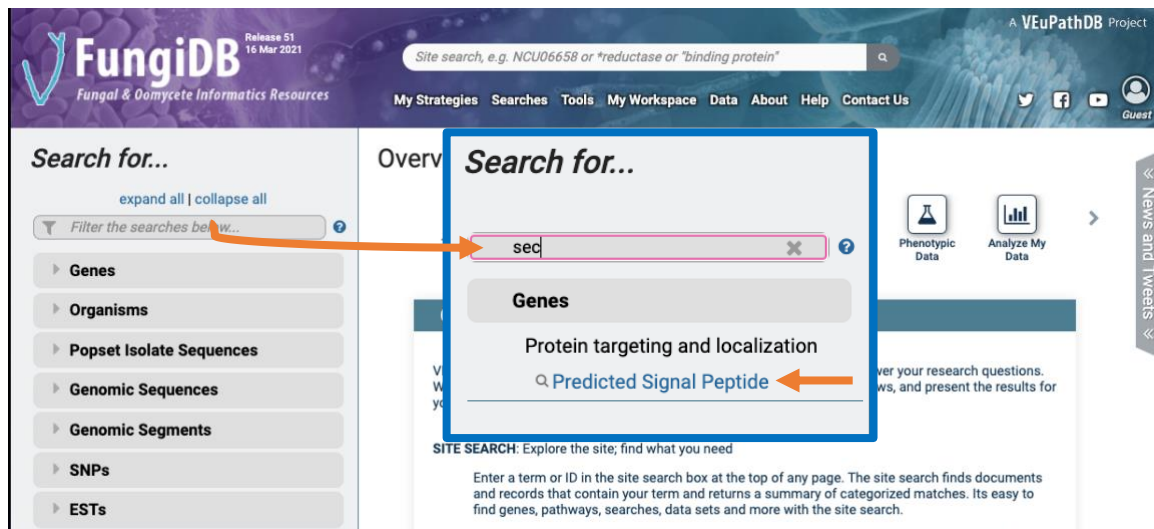
Creating Advanced Search Strategies in FungiDB

Learning objectives:

- Integrate diverse datatypes in a search strategy
- Leverage orthology and phylogenetic profile searches
- Create a nested strategy

This exercise walks you through the process of building a multi-step strategy, integrating different datatypes. The final search strategy identifies *Aspergillus* genes that are likely secreted, or membrane-bound, highly polymorphic, “essential” for parasite survival, not conserved in mammals and expressed in liver stages of the Plasmodium life cycle. There are many ways to build these strategies and order the steps to reach a similar answer.

1. Identify all genes in FungiDB that are predicted to have a secretory signal peptide as defined by SignalP. An easy way to identify a search type is to filter the searches on the left of the home page. Start typing a word to identify the search type. For example, start typing the word "secreted", you should see the searches being filtered even before you finish typing the complete word.



- Click on the search for genes by *Predicted Signal Peptide*. On the next page select all fungal organisms and click on the get answer button at the bottom of the page.

Identify Genes based on Predicted Signal Peptide

[Reset values](#)

Organism

Note: You must select at least 1 values for this parameter.
142 selected, out of 163

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

- ☒ Fungi
- ☐ Oomycota

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

[Get Answer](#)

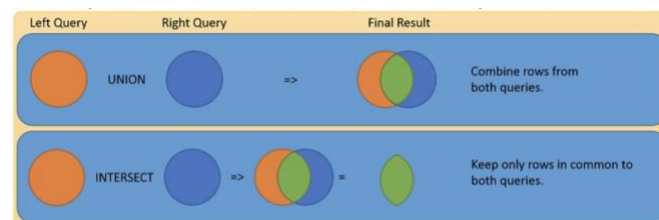
- The next step is to combine the signal peptide results with results of genes that are predicted to have at least one transmembrane domain (TM). Click on the add step button in the search strategy panel.

Unnamed Search Strategy [✎](#)

Signal Pep
112,170 Genes
Step 1

[+](#) Add a step

The popup window offers you the option to add additional steps and ways to combine the searches (intersect, union, minus). For this exercise, we are interested in finding genes that a signal peptide or a TM domain or both. To achieve this, select the Union option.



Once you select the option for combining the searches, find the search for transmembrane domain count. Notice that you can use the same query filtering mechanism as before. Start typing transmembrane to find this search. Once you find it click on it to open the search parameters.

← Add a step to your search strategy ? ✕

Combine with other Genes

Step 1 Step 2

Transform into related records

Step 1 Step 2

Use Genomic Colocation to combine with other features

Step 1 Step 2

1 Choose how to combine with other Genes

☒ 1 INTERSECT 2
 ☐ 1 UNION 2
 ☐ 1 MINUS 2
 ☐ 2 MINUS 1

2 Choose which Genes to combine. From...

☒ A new search
 ☐ An existing strategy
 ☐ My basket

- Protein targeting and localization
 - Predicted Signal Peptide
 - Transmembrane Domain Count
- Proteomics
 - Post-Translational Modification
- Transcriptomics
 - Microarray Evidence
 - RNA-Seq Evidence

4. For the TM search, again select all fungal organisms, use the default parameters and click on the Run Step button.

Search for Genes by Transmembrane Domain Count

The results will be ☐ unioned with ☐ the results of Step 1.

Organism

Note: You must select at least 1 values for this parameter.
142 selected, out of 163

select all | clear all | expand all | collapse all

Filter list below... ?

- ☒ Fungi
☐ Oomycota

select all | clear all | expand all | collapse all

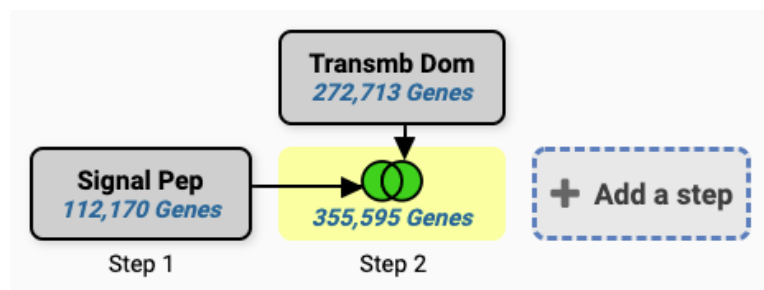
Minimum Number of Transmembrane Domains

1

Maximum Number of Transmembrane Domains

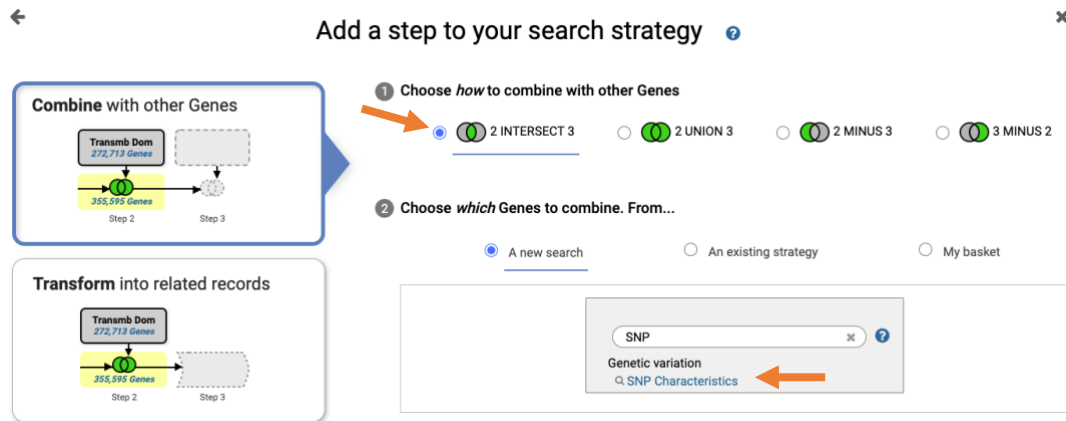
99

5. How many genes did you get? Since you used a union the number of results should be more than each of the individual steps that were combined.

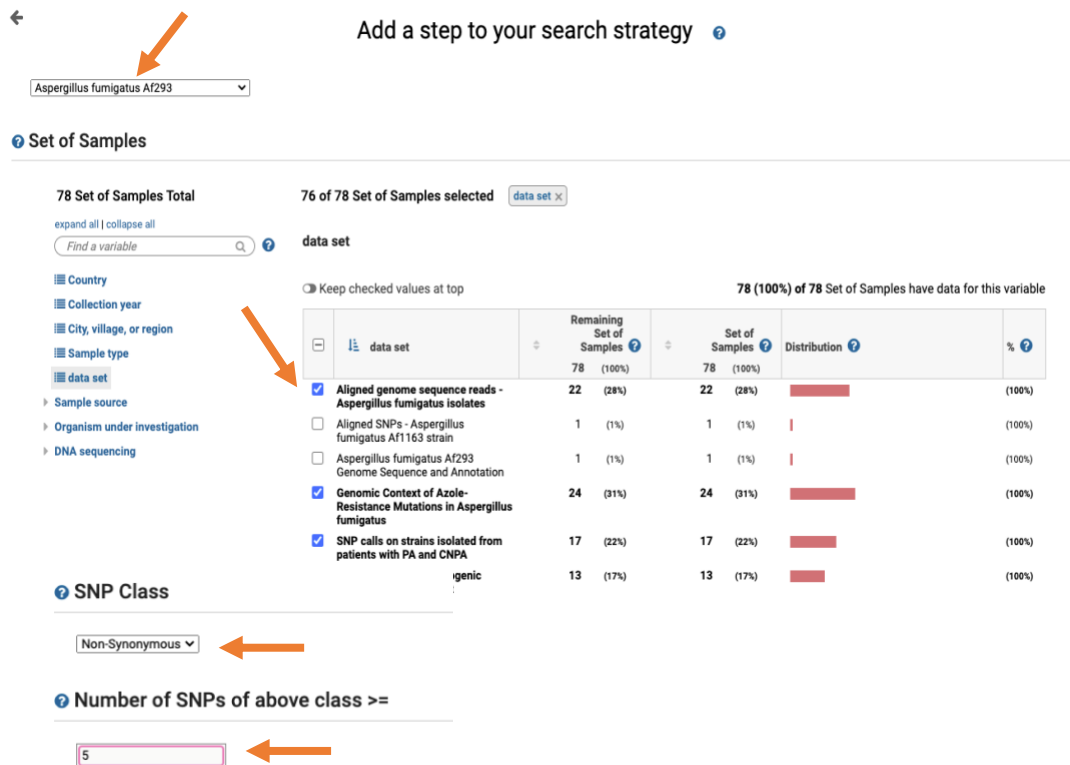


6. Next, identify genes from step 2 that contain at least 5 non-synonymous SNPs (non-synonymous SNPs are single nucleotide polymorphisms that result in an amino acid change).

Start by clicking on the *Add Step* button and then use the filter box to find the *SNP characteristics* search.

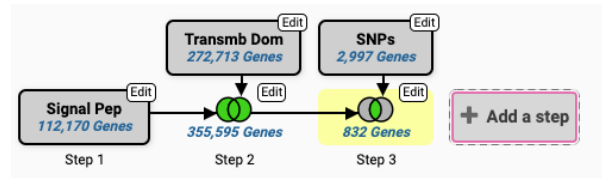


7. On the Genes by SNP characteristics search popup, select *Aspergillus fumigatus* Af293 from the drop-down menu, select datasets as shown below, and then scroll down to



specify SNP class and number of SNPs per gene. Set the number of *SNPs of above class* ≥ 5 . After you select these parameters, scroll down to click on Run Step.

What do the results look like? What species are represented in the results? Is this surprising? Remember that your last search only queried *A. fumigatus* Af293 data.



- Determine how many of these genes are also differentially expressed during macrophage infection with fungal pathogens. Click on add step then search for the RNA-seq Evidence search. Type “trans” for transcriptomics or “RNA” to filter available searches.

← Add a step to your search strategy ⓘ

Combine with other Genes

Transform into related records

Use Genomic Colocation to combine with other features

1 Choose *how* to combine with other Genes

☒ 3 INTERSECT 4
 ☐ 3 UNION 4
 ☐ 3 MINUS 4
 ☐ 4 MINUS 3

2 Choose *which* Genes to combine. From...

☒ A new search
 ☐ An existing strategy
 ☐ My basket

- Protein targeting and localization
 - Predicted Signal Peptide
 - Transmembrane Domain Count
- Proteomics
 - Post-Translational Modification
- Transcriptomics
 - Microarray Evidence
 - RNA-Seq Evidence

- On the next page find data that queries transcriptomics dataset for *C. albicans* where cells infected macrophages

← Add a step to your search strategy ⓘ

Search for Genes by RNA-Seq Evidence

The results will be intersected with the results of Step 3.

Filter Data Sets:

Legend: Similarity Differential Expression Fold Change Percentile SenseAntisense

Organism	Data Set	Choose a Search
<i>Aspergillus fumigatus</i> Af293	Adaptation to different growth environments (blood) (Irmer et al. 2015)	
<i>Candida albicans</i> SCS314	Fresh human whole blood infected with <i>Candida albicans</i> (Sieber et al. 2018)	
<i>Candida glabrata</i> CBS 138	Fresh human whole blood infected with <i>Candida glabrata</i> (Sieber et al. 2018)	
<i>Candida parapsilosis</i> CDC317	Fresh whole blood infected with <i>Candida parapsilosis</i> (Sieber et al. 2018)	
<i>Candida tropicalis</i> MYA-3404	Fresh whole blood infected with <i>Candida tropicalis</i> (Sieber et al. 2018)	

Configure the RNA-Seq search to look for genes that are up-regulated in cells infecting live macrophages (Ca only 4h – reference; infected MO live Ca 4h – comparison).

← Add a step to your search strategy ?

with a fold change ≥ 2

between each gene's minimum expression value
(or a Floor of 10 reads)

in the following Reference Samples

- ☐ Ca only 1h
- ☐ Ca only 2h
- ☒ Ca only 4h
- ☐ Ca unexposed
- ☐ Ca exposed 0h
- ☐ Ca exposed 1h

select all | clear all

and its maximum expression value
(or the Floor selected above)

in the following Comparison Samples

- ☐ Infected MO live Ca 1h
- ☐ Infected MO live Ca 2h
- ☒ Infected MO live Ca 4h
- ☐ Infected MO dead Ca 2h
- ☐ Infected MO dead Ca 4h
- ☐ Ca only 0h

select all | clear all

Run Step

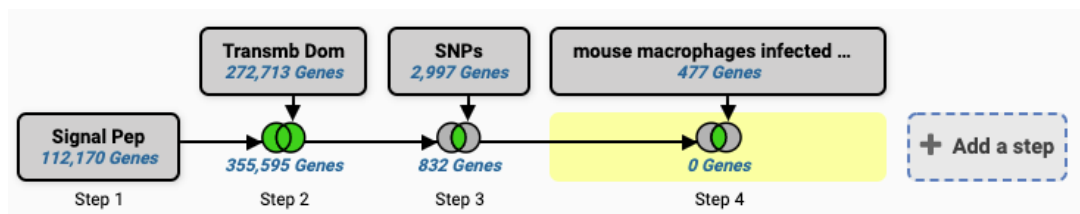
For each gene, the search calculates:

$$\text{fold change} = \frac{\text{comparison expression value}}{\text{reference expression value}}$$

and returns genes when fold change ≥ 2 .

You are searching for genes that are up-regulated between one reference sample and one comparison sample.

10. How many results did you get?



Why did you get 0 results?

Remember that the previous search was a list of *A. fumigatus* genes and this RNA-Seq was from *Candida albicans*.

How can you change this? You can convert the list of *C. albicans* genes into *A. fumigatus* genes. To do this follow these steps:

- Hover your mouse of the RNA-seq step then click on the edit option on that step.
- In the popup window, click on the **orthologs** link.

Details for step mouse macrophages infected with *C. albicans* (fc)

477 Genes

Experiment	Transcriptional analysis of sorted subpopulations of macrophages infected with <i>C. albicans</i> unstranded
Direction	up-regulated
Reference Samples	Ca only 4h
Operation Applied to Reference Samples	minimum
Comparison Samples	Infected MO live Ca 4h
Operation Applied to Comparison Samples	maximum
fold difference \geq	2
Floor	10 reads
Protein Coding Only	protein coding

- c. In the next window select which organism(s) you would like to transform to. For this exercise select *A. fumigatus* Af293 and click on the Run step button.

Organism

Note: You must select at least 1 values for this parameter.
0 selected, out of 163

add these | clear these | select only these
select all | clear all

Af2

☐ Fungi

☐ Ascomycota

☐ Eurotiomycetes

☐ Eurotiales

☐ Aspergillaceae

☐ Aspergillus

☐ Aspergillus fumigatus

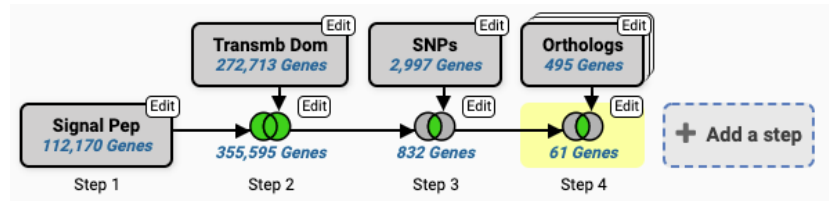
☐ Aspergillus fumigatus Af293

add these | clear these | select only these
select all | clear all

Syntenic Orthologs Only?

no

- d. Did you get results now?



11. Next, identify how many of these genes do not have orthologs in mammals.

To do this add a step for genes based on *Orthology Phylogenetic Profile*. Again you can filter the searches by typing the word “phylogenetic” or “ortho”.



Add a step to your search strategy

Combine with other Genes

Orthologs 495 Genes

61 Genes

Step 4

Step 5

Transform into related records

Orthologs 495 Genes

61 Genes

Step 4

Step 5

1 Choose how to combine with other Genes

☒ 4 INTERSECT 5 ☐ 4 UNION 5 ☐ 4 MINUS 5 ☐ 5 MINUS 4

2 Choose which Genes to combine. From...

☒ A new search ☐ An existing strategy ☐ My basket

ortho

Orthology and syteny


Orthology Phylogenetic Profile

Paralog Count

On the next page select *A. fumigatus* Af293 and then configure the phylogenetic profile by finding Mammalia under Chordata which are under Metazoa. Click twice on the circle next to Mammalia – it should become a red X

← Add a step to your search strategy ?

Search for Genes by Orthology Phylogenetic Profile

The results will be  intersected with | v the results of Step 4.

? Find genes in these organisms

Note: You must select at least 1 values for this parameter.
1 selected, out of 163

add these | clear these | select only these
select all | clear all

Af293

☐ Fungi

☐ Ascomycota

☐ Eurotiomycetes

☐ Eurotiales

☐ Aspergillaceae






☐ Aspergillus


















☐ Aspergillus fumigatus

☒ Aspergillus fumigatus Af293

add these | clear these | select only these
select all | clear all

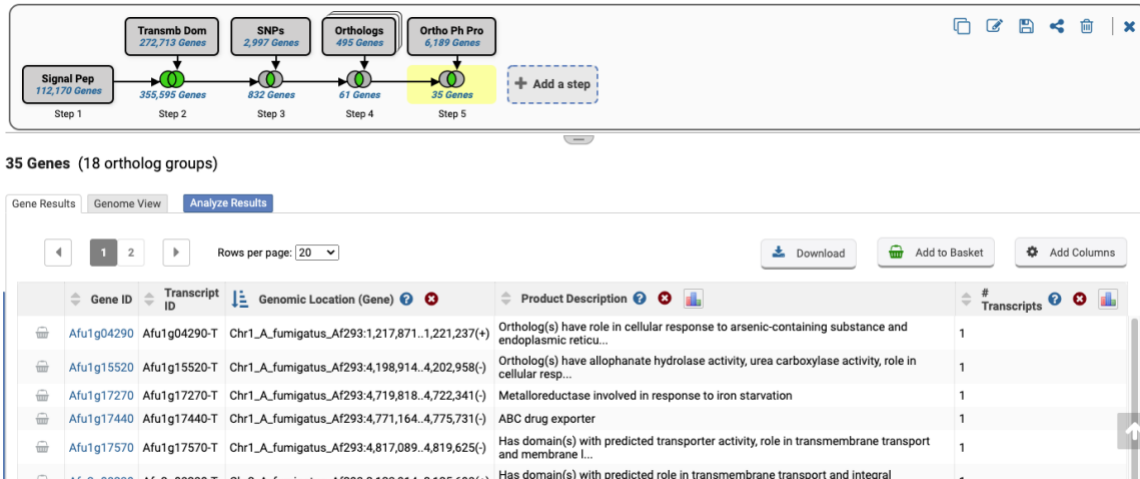
? Select orthology profile

Click on  to determine which organisms to include or exclude in the orthology profile.
( = no constraints /  = must be in group /  = must not be in group /  = mixture of constraints)

- ✱ All Organisms expand all | collapse all
- ▶  Bacteria (BACT)
 - ▶  Archaea (ARCH)
 - ▼ ✱ Eukaryota (EUKA)
 - ▶  Alveolates (ALVE)
 - ▶  Amoebozoa (AMOE)
 - ▶  Euglenozoa (EUGL)
 - ▶  Viridiplantae (VIRI)
 - ▶  Fungi (FUNG)
 - ▼ ✱ Metazoa (META)
 - ▶  Nematodes (NEMA)
 - ▶  Arthropoda (ARTH)
 - ▼ ✱ Chordata (CHOR)
 - ▶  Branchiostoma floridae (Florida lancelet) (bfl1o)
 - ▶  Xenopus tropicalis (Western clawed frog) (xtro)
 - ▶  Actinopterygii (ACTI)
 - ▶  Aves (AVES)
 - ▶  Mammalia (MAMM)
 - ▶  Tunicates (TUNI)
 - ▶  Other Metazoa (OMET)
 - ▶  Other Eukaryota (OEUK)

Run Step

12. Explore your final results. Do they make sense?

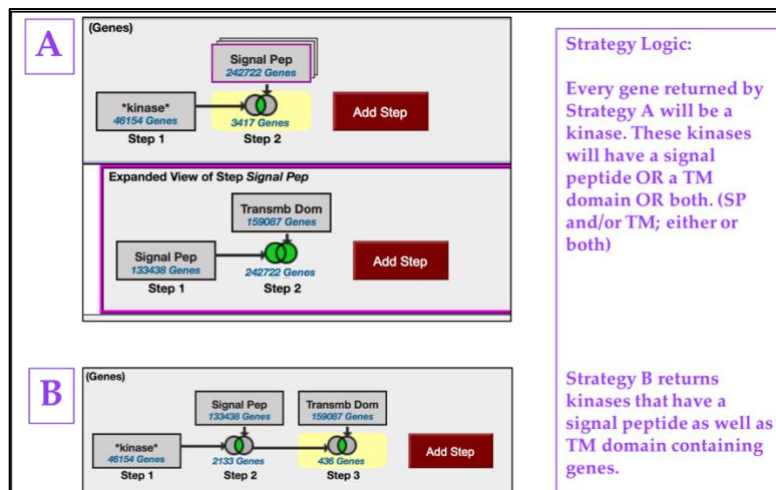


Here is a link to this search strategy:

<https://fungidb.org/fungidb/app/workspace/strategies/import/9cfb2758c0f0deda> Note that you can revise any of the steps in the strategy (e.g. insert steps) to explore the data further. You can also save your strategy and share it with others or make it public.

13. Create a nested strategy.

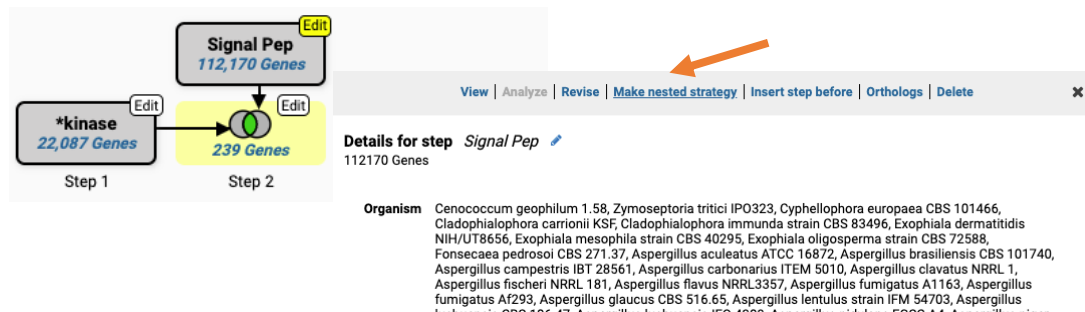
Nesting a strategy allows you to control the order in which your search result sets are combined. Consider the difference between A and B and how search conditions reflect on the results.



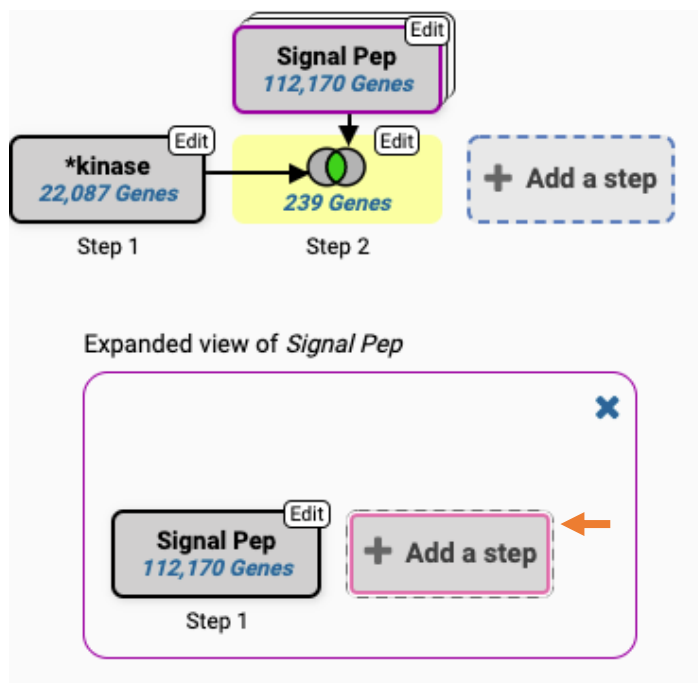
In this section let's go back to the site search strategy where we identified genes matching **kinase** in the *Product description*. We will use this strategy to create a nested strategy for signal peptides and transmembrane domains as shown in A. Here is the **kinase** strategy link:

<https://fungidb.org/fungidb/app/workspace/strategies/import/fla2d582b9e59f0d>

- Add a step to create a search for signal peptides
- Click on the *edit* option to select the *Make nested strategy* link:



c. Add a step to add the *Transmembrane Domain Count* search in *Fungi*:



Search for Genes by Transmembrane Domain Count

The results will be unioned with the results of Step 1.

Organism

Note: You must select at least 1 values for this parameter.
142 selected, out of 163

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

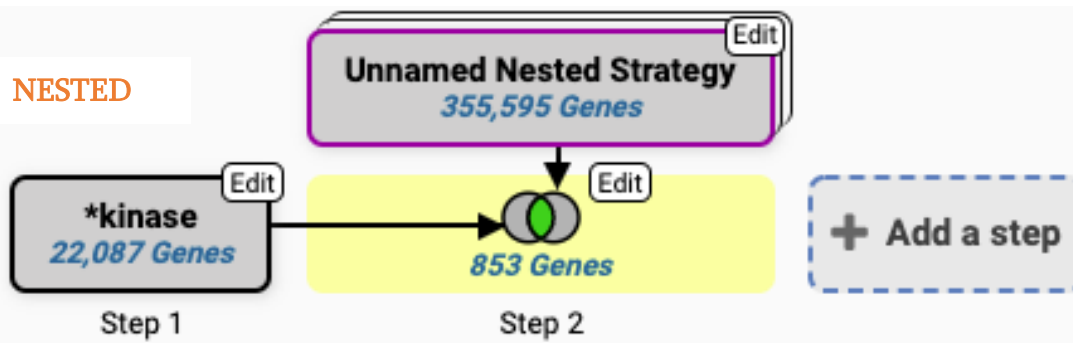
Filter list below...

- ☒ Fungi
- ☐ Oomycota

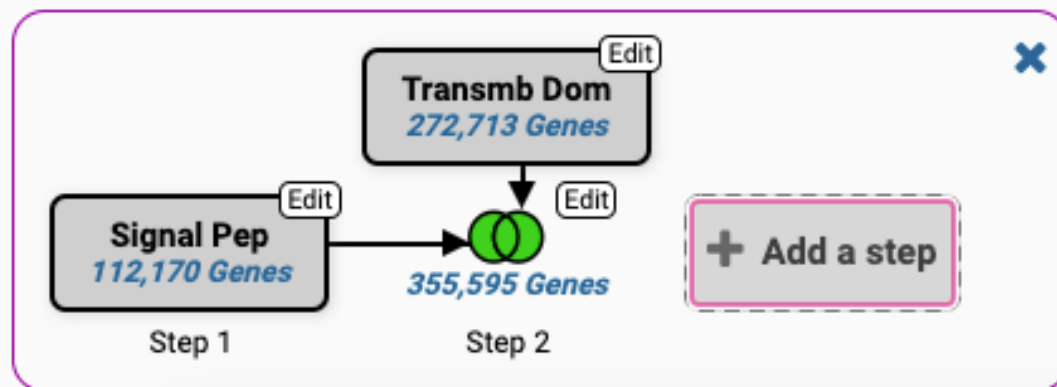
[select all](#) | [clear all](#) | [expand all](#) | [collapse all](#)

Compare the results of the nested strategy to a linear non-nested one:

NESTED



Expanded view of *Unnamed Nested Strategy*



NOT NESTED

