

Advanced Search Strategies

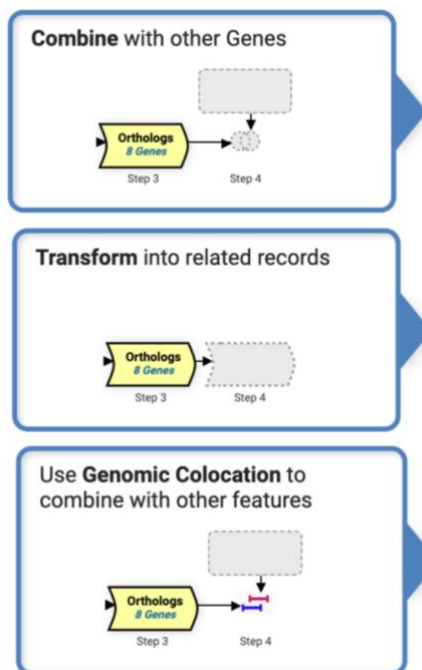
Learning objectives

- Deploy search for different types of data and create advanced search strategy in FungiDB.

The strategy system offers a unique system of structured searches that can be combined to create multi-step *in-silico* experiments. As seen above, searches can be deployed from the site search, or the ‘Search For...’ menu on the home page, and from the ‘Searches’ dropdown menu in the header of every page.

Searches listed under the “Genes” category will return a list of gene IDs, while searches listed under the ‘SNPs’ or ‘Metabolic Pathways’ will return record relevant to SNPs data (e.g., sequences) and metabolic pathways, respectively.

When creating multi-step search strategy, the search strategy steps can be combined via three methods:



Combine with other Genes: compares results that are gene lists.

Transform into related records: transforms results into orthologs (e.g. *Aspergillus* > *Candida*), metabolic pathways or compounds.

Use Genomic Colocation to combine with other features: cross-references different types of data – e.g., gene lists with metabolic pathways.

Within the search strategy, each step is connected via the system of Boolean operators that can intersect, unite, or subtract similar records (e.g., gene lists) and cross-references different types of data via the genomic colocation option. Additionally, steps can be masked off from the strategy with the help of “ignore step” Boolean operators that allow quick strategy modification without the need of step deletion.

Revise as a boolean operation

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

Revise as a span operation

☐ 1 RELATIVE TO 2, using genomic colocation

Ignore one of the inputs

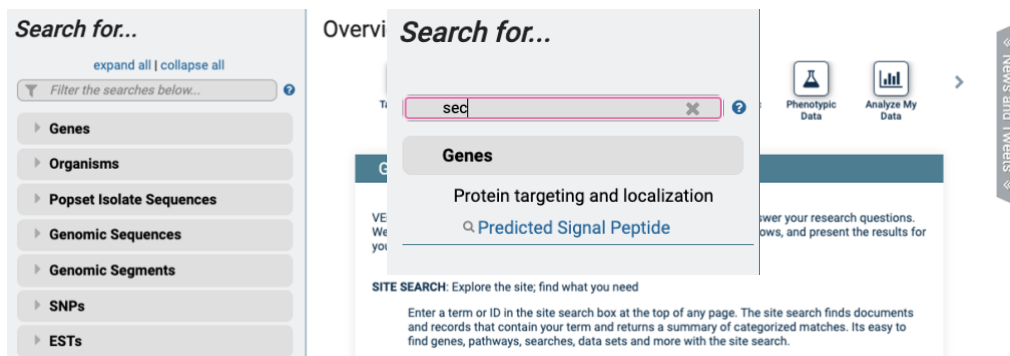
☐ IGNORE 2 ☐ IGNORE 1

Revise

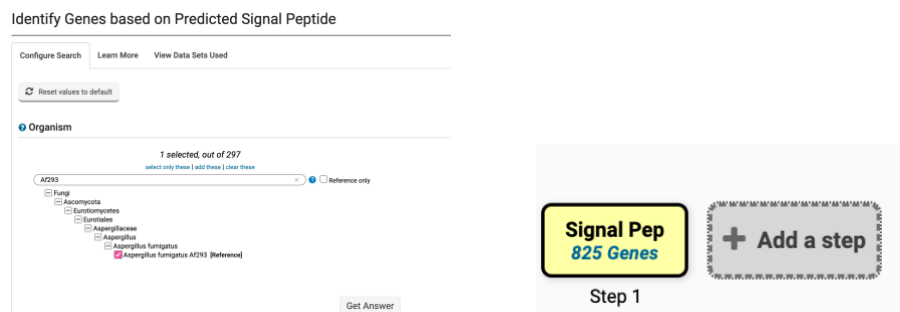
Creating advanced search strategies in FungiDB.

In silico experiment: Identify *Aspergillus fumigatus* Af293 genes that have a signal peptide and non-synonymous mutations identified by whole genome sequencing (WGS) of clinical isolates. Next, determine which genes may be putative vaccine targets (are known epitopes that do not have orthologs in humans).

- **Deploy the “Predicted Signal Peptide” search**
 1. Select the search from the “Search for...” panel (shown below) or the “Searches” menu from the top menu.
Hint: use the filter box to quickly bring up relevant search.



2. Set the organism to “Aspergillus fumigatus Af293 [Reference]” and click on the “Get Answer” button.



- **Find genes with non-synonymous SNPs using WGS data of clinical isolates aligned to the reference genome *Aspergillus fumigatus* Af293.**
 1. Click on the “Add a step” button with the search strategy.
 2. Within the “Combine with other Genes”, use the “1 INTERSECT 2” Boolean operator and filter available searches to identify and deploy the “SNP Characteristics” search.
 3. Click on the “SNP Characteristics” link in blue to deploy the search.

← 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

Genetic variation

SNP Characteristics

Find genes which have SNPs generated by HTS sequencing meeting specific criteria (Non-synonymous / synonymous SNP ratio, SNP density, etc).

Next, specify parameters of the SNP characteristics search:

1. Select *Aspergillus fumigatus* Af293 as the genome of interest.

Search for Genes by SNP Characteristics

The results will be ☒ Intersected with ☐ the results of Step 1.

Configure Search Learn More View Data Sets Used

Organism

1 selected

☒ AC293
 ☐ Reference only

☐ Ascomycota
 ☐ Eurotiomycetes
 ☐ Eurotiomycetes
 ☐ Aspergillaceae
 ☐ Aspergillus
 ☒ Aspergillus fumigatus AC293 (Reference)

2. Select datasets.

Within the Set of Samples section, click on the Data Set category and select two datasets:

- SNP call on WGS of *Aspergillus fumigatus* drug-resistant clinical isolates
- SNP calls on *A. fumigatus* strains isolated from patients with PA and CNPA

Set of Samples

1,118 Set of Samples Total 39 of 1,118 Set of Samples selected

expand all | collapse all

Find a variable

☒ culture medium
 ☒ Fungal organism
 ☒ Data Set
 ☐ Sample
 ☐ Sample collection
 ☐ Geographic location
 ☐ Sample source
 ☐ Organism under investigation

Data Set

☐ Keep checked values at top

1,118 (100%) of 1,118 Set of Samples have data for this variable

<input type="checkbox"/> Data Set	Remaining Set of Samples	Set of Samples	Distribution	%
Find items	1,118 (100%)	1,118 (100%)		
<input type="checkbox"/> Aligned SNPs - <i>Aspergillus fumigatus</i> AfT163 strain	1 (< 1%)	1 (< 1%)		(100%)
<input type="checkbox"/> <i>Aspergillus fumigatus</i> Af293 Genome Sequence and Annotation	1 (< 1%)	1 (< 1%)		(100%)
<input type="checkbox"/> <i>Aspergillus fumigatus</i> LH-EVOL strains	4 (< 1%)	4 (< 1%)		(100%)
<input type="checkbox"/> Genomic Context of Azole-Resistance Mutations in <i>Aspergillus fumigatus</i>	24 (2%)	24 (2%)		(100%)
<input checked="" type="checkbox"/> SNP call on WGS of <i>Aspergillus fumigatus</i> drug-resistant clinical isolates	22 (2%)	22 (2%)		(100%)
<input checked="" type="checkbox"/> SNP calls on <i>A. fumigatus</i> strains isolated from patients with PA and CNPA	17 (2%)	17 (2%)		(100%)

- Indicate specific SNP characteristics.
Scroll down the parameter selection window and choose to deploy the search using the following:

SNP Class = Non-Synonymous

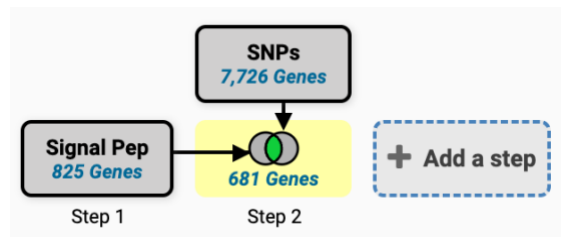
Number of SNPs of above class >=1

SNP Class

Non-Synonymous ▾

Number of SNPs of above class >=

1



- Identify known *Aspergillus* epitopes.**

Epitopes are recognized by immune system and can be used for vaccine development.

- Click on the “Add a step” button.
- Within the “Combine with other Genes”, use the “2 INTERSECT 3” Boolean operator and filter available searches for “epitope” to identify and deploy the “Epitope Presence (IEDB)” search.

Combine with other Genes

Transform into related records

Use Genomic Colocation to combine with other features

1 Choose how to combine with other Genes

☒ 2 INTERSECT 3
 ☐ 2 UNION 3
 ☐ 2 MINUS 3
 ☐ 3 MINUS 2

2 Choose which Genes to combine. From...

☒ A new search
 ☐ An existing strategy
 ☐ My basket

Immunology
 Q Epitope Presence (IEDB)

- Set organism to *Aspergillus fumigatus* Af293.
- Set Confidence to “high” and “Medium” and click on the Run Step button.

Organism

1 selected, out of 232

[select only these](#) | [add these](#) | [clear these](#)

Af293

☐ Reference only

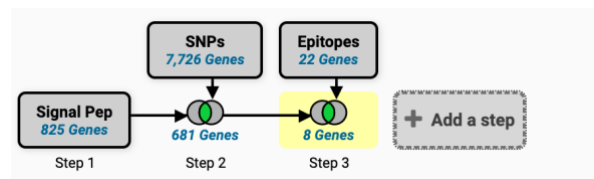
- ☐ Fungi
 - ☐ Ascomycota
 - ☐ Eurotiomycetes
 - ☐ Eurotiales
 - ☐ Aspergillaceae
 - ☐ Aspergillus
 - ☒ Aspergillus fumigatus Af293 [Reference]

Confidence

- ☒ High
- ☒ Medium
- ☐ Low

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

Run Step



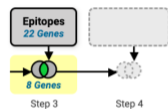
Identify Aspergillus epitopes that do not have orthologs in humans.

1. Click on the “Add a step” button.

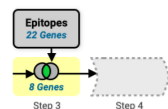
2. Select to deploy the “Orthology Phylogenetics Profile” search.

This search uses OrthoMCL algorithm to identify fungal orthologs across all species in VEuPathDB. Run this search if you want to explore species outside those supported in FunigDB.org.

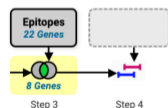
Combine with other Genes



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

orthology

Orthology and synteny
[Q Orthology Phylogenetic Profile](#)
[Q Paralog Count](#)

3. Set parameters for the “Search for Gene by Orthology Phylogenetic Profile”:
Find genes in these organisms: *Aspergillus fumigatus* Af293
Select orthology profile: *Homo sapiens* REF (hsap) must not be in group

Find genes in these organisms

1 selected, out of 297
[select only these](#) | [add these](#) | [clear these](#)

Af29

☐ Reference only

- ☐ Fungi
 - ☐ Ascomycota
 - ☐ Eurotiomycetes
 - ☐ Eurotiales
 - ☐ Aspergillaceae
 - ☐ Aspergillus
 - ☐ *Aspergillus fumigatus*
 - ☒ *Aspergillus fumigatus* Af293 [Reference]

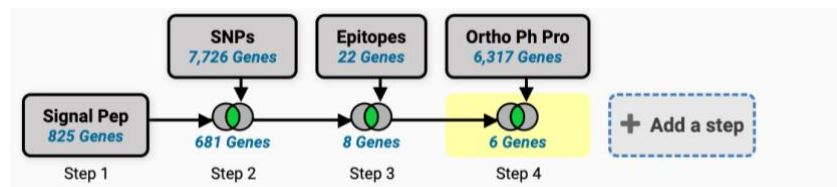
Select orthology profile

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)

sapiens

- All Organisms
 - Eukaryota (EUKA)
 - Metazoa (META)
 - Chordata (CHOR)
 - Mammalia (MAMM)
 - Homo sapiens* REF (hsap)

How many genes did you get?



Search strategy links:

FGC2024 advanced search strategy 1:

<https://fungidb.org/fungidb/app/workspace/strategies/import/741f30239b2cc68d>