VEuPathDB
Eukaryotic Pathogen, Vector & Host Informatics Resources

OrthoMCL DB
Ortholog Groups of Protein Sequences

OrthoMCL 7

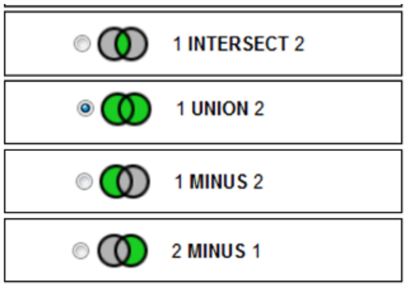
Search Strategies Tutorial[1]

OrthoMCL 7 is a **genome-scale database and website** (orthomcl.org) that uses protein sequence similarity and phylogenetic relationships among proteins to create groups of orthologous protein sequences called orthogroups. OrthoMCL offers a number of tools for exploratory data analysis. Its records can be mined using **search strategies** that take advantage of the ability of OrthoMCL to group both known and unknown proteins into multi-species orthogroups that share protein function. Proteins with known function can be identified in a model species, and then orthologs with analogous functions can be found in less studied species. Text searches can bring in useful annotation from any organism and find related proteins in an organism of interest.

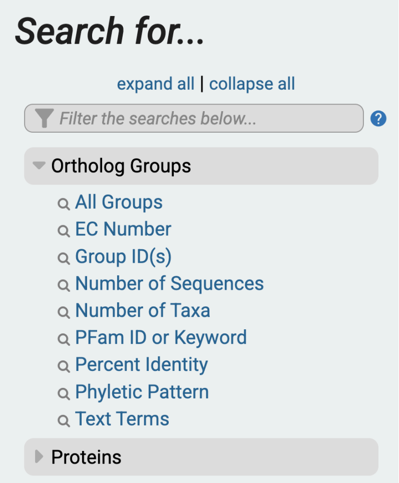
Basic Principles of Search Strategies

OrthoMCL has two types of ***Searches*** – for ***Ortholog Groups*** and for ***Proteins***. Several searches can be combined sequentially into a search strategy with Boolean operators to

- Narrow the results by the **Intersection** of two searches
- Add results of two searches (**Union**)
- **Subtract** one set of results from another


○ 1 INTERSECT 2
● 1 UNION 2
○ 1 MINUS 2
○ 2 MINUS 1

[1] Updated on April 7, 2025



Search for...

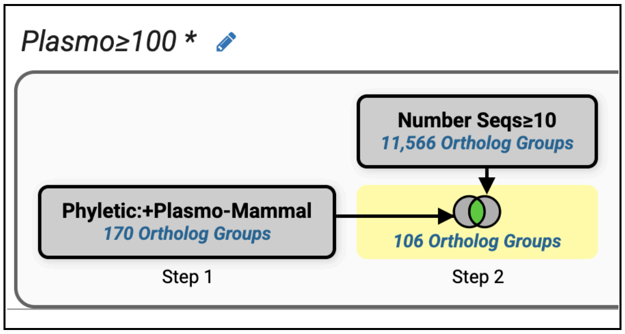
expand all | collapse all

Filter the searches below...

Ortholog Groups
- All Groups
- EC Number
- Group ID(s)
- Number of Sequences
- Number of Taxa
- PFam ID or Keyword
- Percent Identity
- Phyletic Pattern
- Text Terms

Proteins

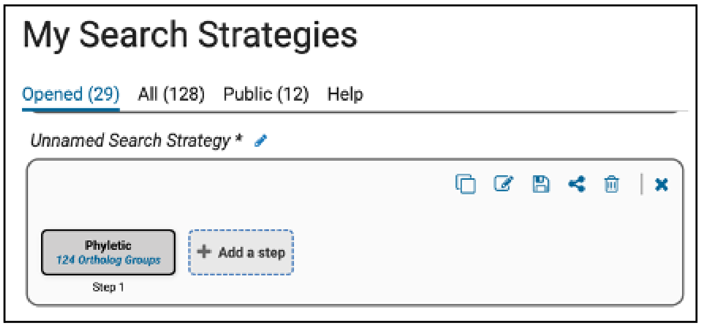
For example, combine by intersection an ***Ortholog Group Phyletic Pattern*** search and an ***Ortholog Group Number of Sequences*** search to find orthogroups with proteins present in *Plasmodium* but absent from mammals that contain at least 100 proteins.



*Plasmo≥100 **

To create a search strategy, make any search from the Search menu on OrthoMCL.org, then on the search results page, hit the ***+Add a step*** button.

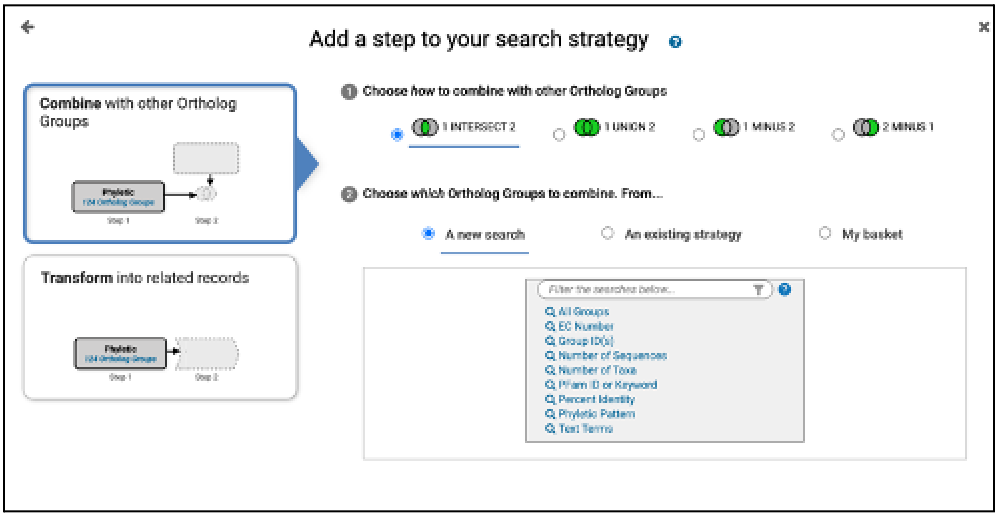
This brings up the Search Strategy window to configure the next step of the search.



My Search Strategies

Opened (29) All (128) Public (12) Help

*Unnamed Search Strategy **



Add a step to your search strategy

Combine with other Ortholog Groups

Transform into related records

1 Choose how to combine with other Ortholog Groups

● 1 INTERSECT 2 ○ 1 UNION 2 ○ 1 MINUS 2 ○ 2 MINUS 1

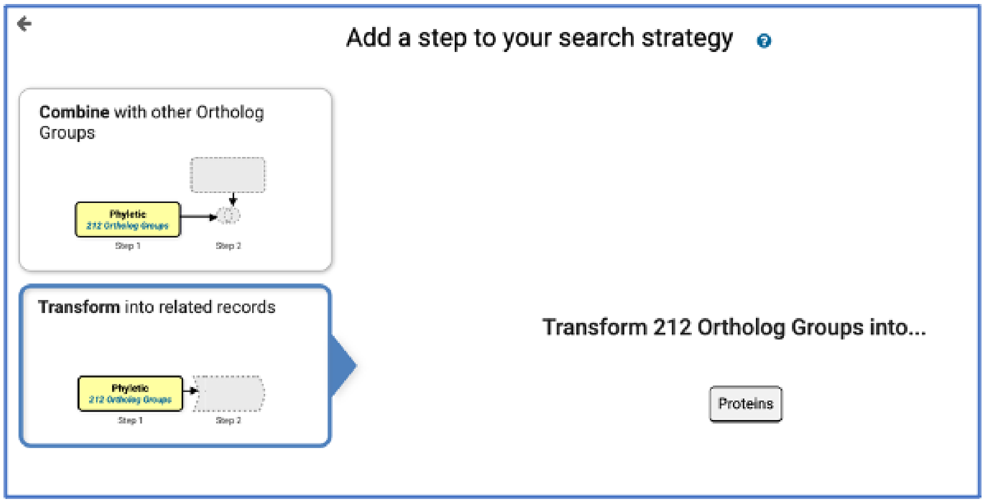
2 Choose which Ortholog Groups to combine. From...

● A new search ○ An existing strategy ○ My basket

Filter the searches below...

- All Groups
- EC Number
- Group ID(s)
- Number of Sequences
- Number of Taxa
- PFam ID or Keyword
- Percent Identity
- Phyletic Pattern
- Text Terms

It is also possible to ***Transform*** a set of orthogroups found with any search (search results) into the complete set of proteins in those groups and then use Protein search functions on the set of proteins. Similarly, a set of proteins in a search result can be transformed into their respective orthogroups and then interrogated with Ortholog Group searches.


Add a step to your search strategy
Combine with other Ortholog Groups
Phyletic
212 Ortholog Groups
Step 1
Step 2
Transform into related records
Phyletic
212 Ortholog Groups
Step 1
Step 2
Transform 212 Ortholog Groups into...
Proteins

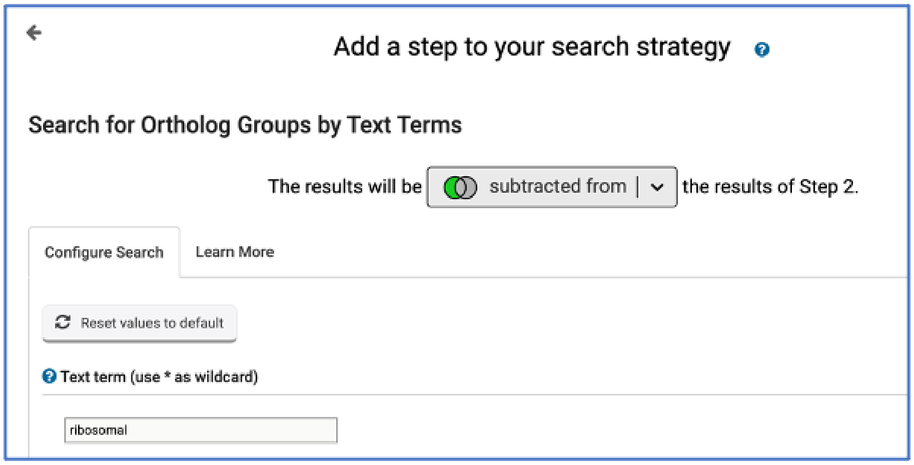
Example 1

The search question: *Plasmodium ovale* was the last of the exclusively human malaria parasites to be described and has remained the least well studied, with many poorly described proteins. Plasmodium species contain a non-photosynthetic plastid organelle called the apicoplast that is crucial to the malaria parasite's survival. The apicoplast is an organelle unique to organisms in the Apicomplexan clade. Due to the algal origin of the apicoplast (which contains its own DNA), many proteins and pathways are not shared by the human host, making it an attractive target for antimalarial drugs.

To characterize apicoplast genes in *Plasmodium ovale*, we could ask the question: **What proteins in *Plasmodium ovale* belong to ortholog groups with apicoplast specific functions?**

1. Search for the ***Proteins -> Text Terms ->*** "apicoplast" across all of OrthoMCL.
2. ***Add a Step*** to this search to ***Transform*** these proteins into orthogroups. The keywords for these orthogroups show that many contain ribosomal proteins, which are broadly conserved across all organisms. This can be confirmed by clicking on one of these ribosomal orthogroups (such as OG7_0002866) to access the orthogroup page and looking at the phyletic distribution of proteins.

3. ***Add a Step*** to this search strategy to ***Subtract*** the text term “ribosomal” as an Ortholog Group text term. The remaining groups are more likely to contain proteins with apicoplast-specific functions.



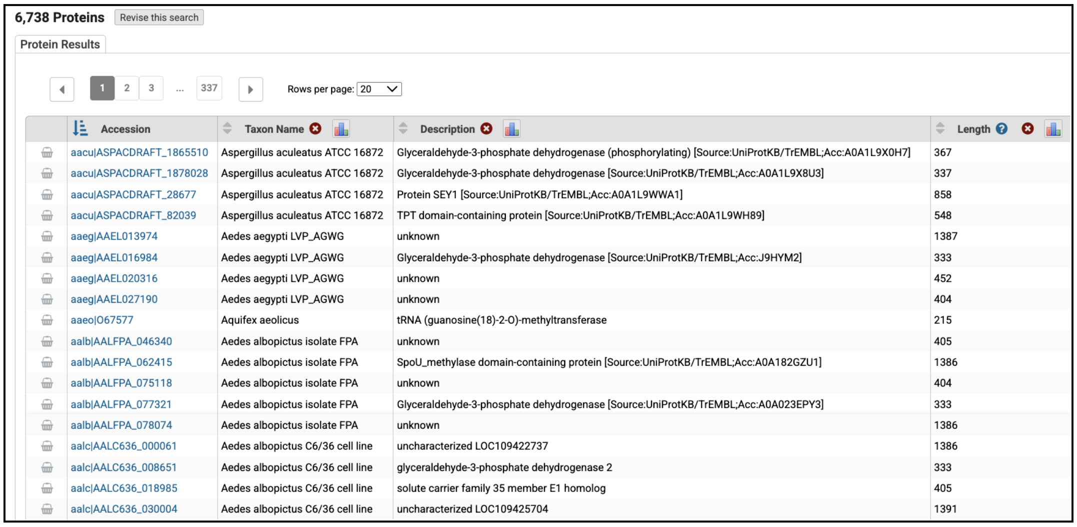
Add a step to your search strategy

Search for Ortholog Groups by Text Terms

The results will be subtracted from the results of Step 2.

Text term (use * as wildcard): ribosomal

4. ***Transform*** the orthogroups to proteins and look at the protein descriptions. The search results contain many proteins that are similar to apicoplast proteins, but are not annotated with this term in their protein description.

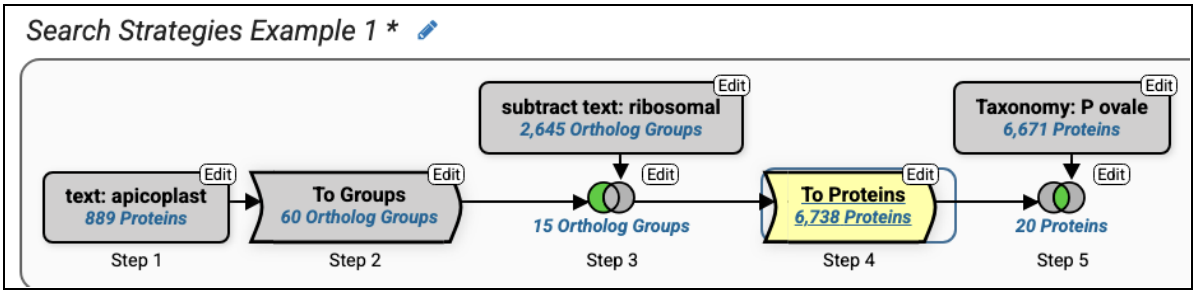


6,738 Proteins

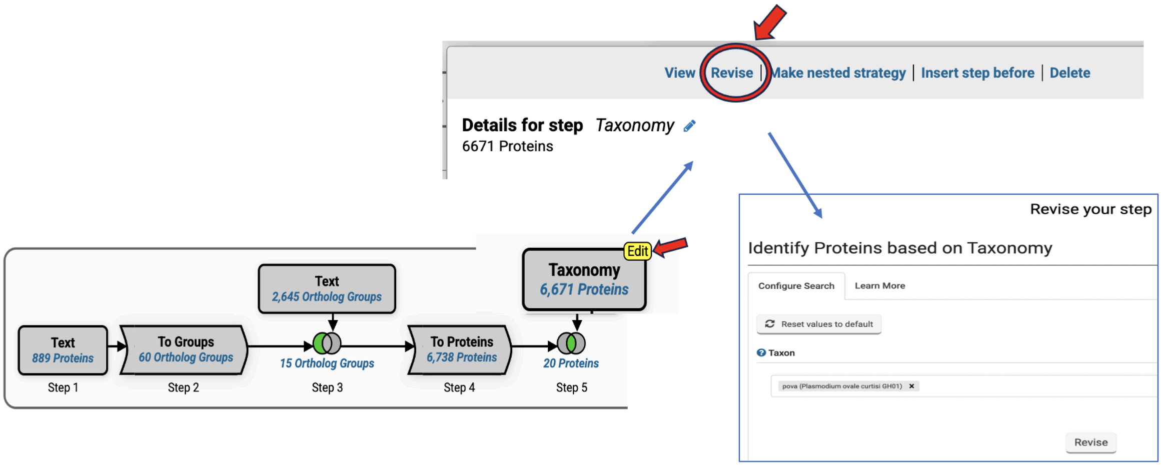
Rows per page: 20

Accession	Taxon Name	Description	Length
aacu|ASPACDRAFT_1865510	Aspergillus aculeatus ATCC 16872	Glyceraldehyde-3-phosphate dehydrogenase (phosphorylating) [Source:UniProtKB/TrEMBL;Acc:A0A1L9X0H7]	367
aacu|ASPACDRAFT_1878028	Aspergillus aculeatus ATCC 16872	Glyceraldehyde-3-phosphate dehydrogenase [Source:UniProtKB/TrEMBL;Acc:A0A1L9X8U3]	337
aacu|ASPACDRAFT_28677	Aspergillus aculeatus ATCC 16872	Protein SEY1 [Source:UniProtKB/TrEMBL;Acc:A0A1L9WWA1]	858
aacu|ASPACDRAFT_82039	Aspergillus aculeatus ATCC 16872	TPT domain-containing protein [Source:UniProtKB/TrEMBL;Acc:A0A1L9WH89]	548
aaeg|AAEL013974	Aedes aegypti LVP_AGWG	unknown	1387
aaeg|AAEL016984	Aedes aegypti LVP_AGWG	Glyceraldehyde-3-phosphate dehydrogenase [Source:UniProtKB/TrEMBL;Acc:J9HYM2]	333
aaeg|AAEL020316	Aedes aegypti LVP_AGWG	unknown	452
aaeg|AAEL027190	Aedes aegypti LVP_AGWG	unknown	404
aaeo|O67577	Aquifex aeolicus	tRNA (guanosine(18)-2-O)-methyltransferase	215
aalb|AALFPA_046340	Aedes albopictus isolate FPA	unknown	405
aalb|AALFPA_062415	Aedes albopictus isolate FPA	SpoU_methylase domain-containing protein [Source:UniProtKB/TrEMBL;Acc:A0A182GZU1]	1386
aalb|AALFPA_075118	Aedes albopictus isolate FPA	unknown	404
aalb|AALFPA_077321	Aedes albopictus isolate FPA	Glyceraldehyde-3-phosphate dehydrogenase [Source:UniProtKB/TrEMBL;Acc:A0A023EPY3]	333
aalb|AALFPA_078074	Aedes albopictus isolate FPA	unknown	1386
aalc|AALC636_000061	Aedes albopictus C6/36 cell line	uncharacterized LOC109422737	1386
aalc|AALC636_008651	Aedes albopictus C6/36 cell line	glyceraldehyde-3-phosphate dehydrogenase 2	333
aalc|AALC636_018985	Aedes albopictus C6/36 cell line	solute carrier family 35 member E1 homolog	405
aalc|AALC636_030004	Aedes albopictus C6/36 cell line	uncharacterized LOC109425704	1391

5. ***Add a Step*** to filter the proteins by ***Taxonomy*** for *Plasmodium ovale* curtisi GH01to find 20 proteins, many of which have no description or functional information, but are similar to apicoplast proteins annotated in other organisms.



This strategy can be **Revised** to find orthologs of apicoplast-related proteins in any VEuPathDB organism.

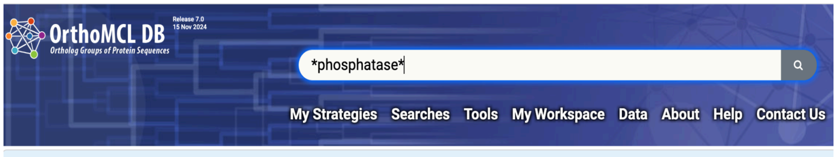


Example 2

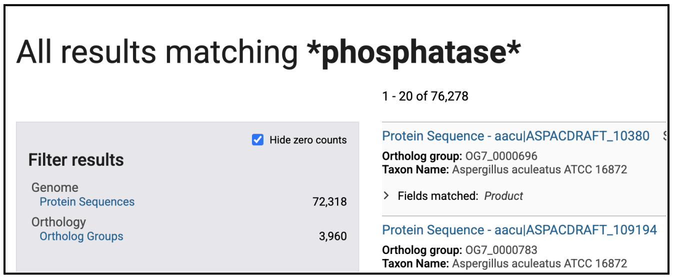
The search question: Pathogenic fungi are fungi that cause disease in humans, animals, or plants. Notable examples include human and animal pathogens such as *Candida albicans*, *Cryptococcus neoformans* and *Aspergillus fumigatus*, and plant pathogens such as *Magnaporthe oryzae* and *Ustilago maydis*. Phosphatases play key roles in fungal pathogenesis, particularly nutrient acquisition, immune evasion and host interaction, signal transduction, and tissue invasion. The importance of phosphatases in fungal biology has led to their consideration as potential targets for antifungal drug development.

To identify phosphatases that are potential antifungal drug targets, we could ask the question: **What fungal proteins are likely to be phosphatases and do not have orthologs in any organism outside of the fungal kingdom?**

1. Use the ***site search*** to look for *phosphatase* (use asterisks- wildcard character- to find any combination of the word “phosphatase”).



2. Choose ortholog groups in the ***Filter Results*** panel at the left and click the “Export as a Search Strategy” button.



All results matching ***phosphatase***

1 - 20 of 76,278

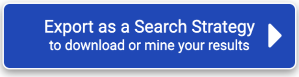
☑ Hide zero counts

Filter results

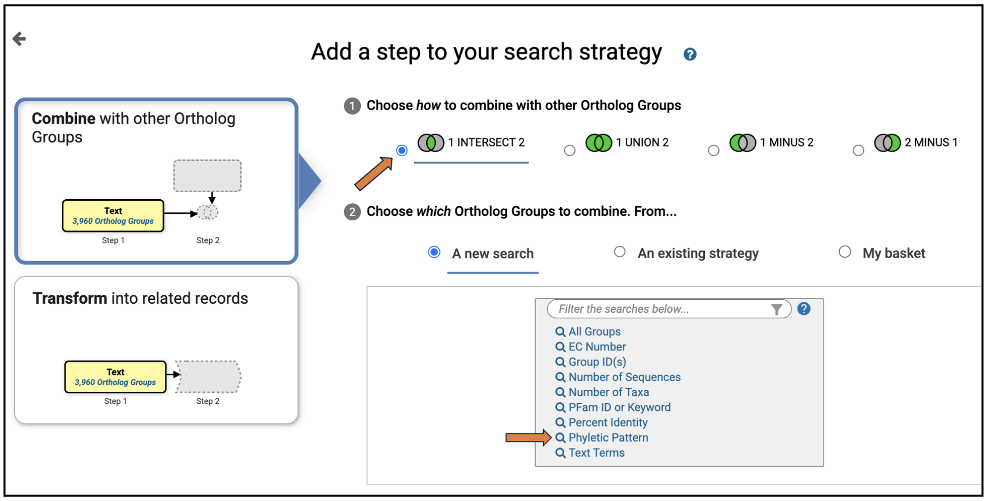
Genome	
Protein Sequences	72,318
Orthology	
Ortholog Groups	3,960

Protein Sequence - aacu|ASPACDRAFT_10380
Ortholog group: OG7_0000696
Taxon Name: Aspergillus aculeatus ATCC 16872
> Fields matched: *Product*

Protein Sequence - aacu|ASPACDRAFT_109194
Ortholog group: OG7_0000783
Taxon Name: Aspergillus aculeatus ATCC 16872



3. Click ***Add a Step*** in the Search Strategy and choose Intersect with a Phyletic Pattern search.



Add a step to your search strategy

Combine with other Ortholog Groups

Transform into related records

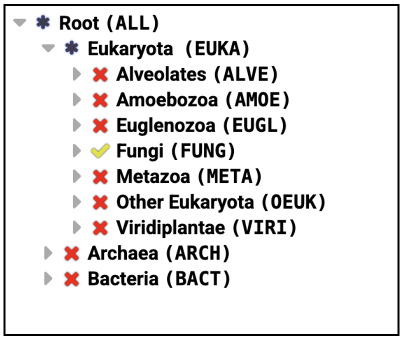
1 Choose *how* to combine with other Ortholog Groups

● 1 INTERSECT 2 ○ 1 UNION 2 ○ 1 MINUS 2 ○ 2 MINUS 1

2 Choose *which* Ortholog Groups to combine. From...

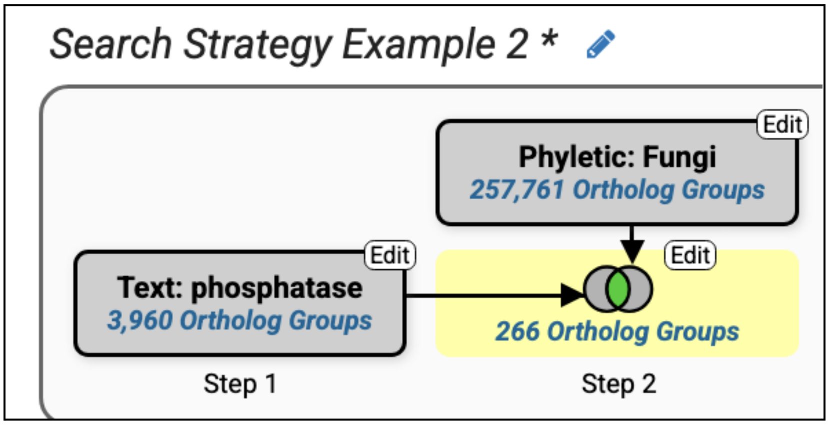
● A new search ○ An existing strategy ○ My basket

4. The ***Phyletic Pattern*** search is easy to set up with the clickable interface on the taxonomic tree menu (any fungi, not in any other clade). After setting up the search, click “***run step***”



- ✱ Root (ALL)
  - ✱ Eukaryota (EUKA)
    - ✗ Alveolates (ALVE)
    - ✗ Amoebozoa (AMOE)
    - ✗ Euglenozoa (EUGL)
    - ✓ Fungi (FUNG)
    - ✗ Metazoa (META)
    - ✗ Other Eukaryota (OEUK)
    - ✗ Viridiplantae (VIRI)
  - ✗ Archaea (ARCH)
  - ✗ Bacteria (BACT)

5. How many fungi-specific orthogroups of phosphatase proteins are found?



Search Strategy Example 2 *

Phyletic: Fungi
257,761 Ortholog Groups

Text: phosphatase
3,960 Ortholog Groups

266 Ortholog Groups

Questions? Comments?

Contact us- help@veupathdb.org