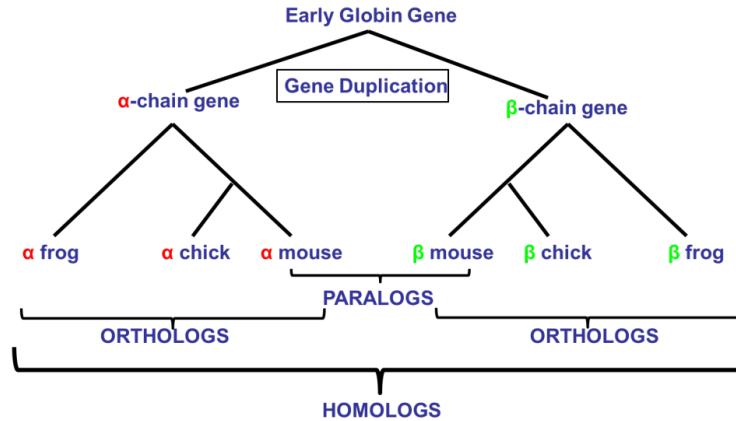


# FungiDB: Orthology and Phyletic Patterns

## Homology



### 1. Getting to OrthoMCL from EuPathDB databases

- Go to the gene record page for the *Cryptococcus gattii* gene CGB\_L0350W
- What is the function of this gene? How can you infer its function?
- Click on the “Orthology and Synteny” link on the left. Does this gene have orthologs in other *Cryptococcus* species? What about other organisms outside of fungi? (hint: click on the *Ortholog Group* OG5\_127157; this link will take you to OrthoMCL).

7 Orthology and synteny

Ortholog Group **OG5\_127157**

▼ Orthologs and Paralogs within EuPathDB [Data sets](#)

To run Clustal Omega, select genes from the table below. Then choose the sequence type and initiate the alignment with the 'Run Clustal Omega for selected genes' button.

Search this table... Showing 142 rows

<input type="checkbox"/>	Clustal Omega	Gene	Organism	Product	Is syntenic	has comments
<input type="checkbox"/>		AGR57_3474	Phanerochaete chrysosporium RP-78	Cation efflux protein	yes	no
<input type="checkbox"/>		CC1Q_14848	Coprinopsis cinerea okayama74130	hypothetical protein	yes	no
<input type="checkbox"/>		CKF44_05394	Cryptococcus neoformans var. grubii K999	cation:cation antiporter	yes	no

- Mouse over the colorful boxes in the table to reveal the full species and phylum names.

Group: **OG5\_127157**  
(139 sequences)

[Add to Basket](#) [Add to Favorites](#)

Sequences & Statistics | PFam domains (graphic) | PFam domains (details) | MSA | Cluster graph

Phylogenetic Distribution [Hide](#)

Legend: ☐ no ortholog ☒ one ortholog ☐ more than one ortholog

[Show labels](#)

Group	# Sequences	Average % Identity	Average % Connectivity	EC Numbers
OG5_127157	139	34.7	38.9	

EC Number none

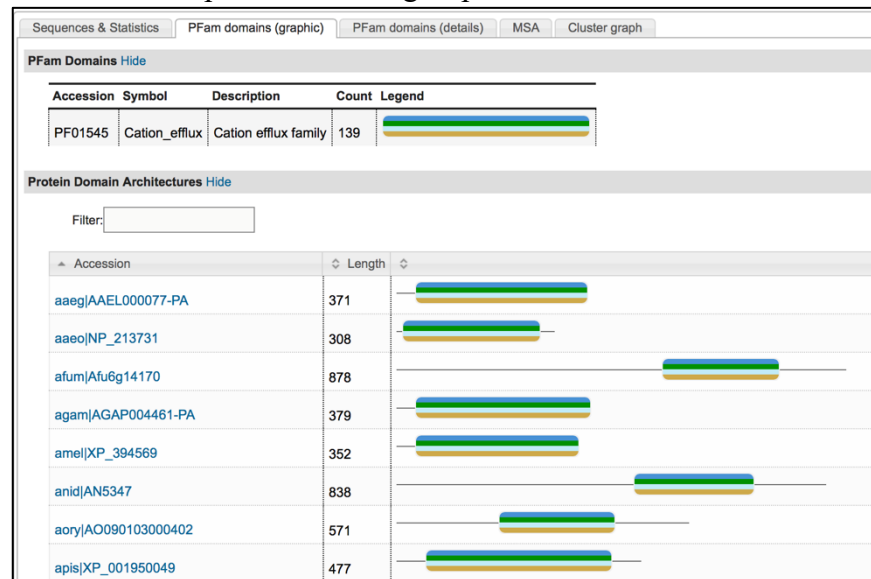
List of Sequences [Hide](#)

Get Sequences: [As Fasta file](#) [As new strategy](#)

Filter:

Accession	Taxon	Length	Previous Groups	EC Numbers	Description
tbu Tb11.42.0002	Trypanosoma brucei	322	OG2_70893, OG3_10134, OG4_10082	N/A	cation transporter protein, putative
tbu Tb427mp.42.0002	Trypanosoma brucei	322	N/A	N/A	cation transporter protein, putative

- e. Is this gene widely conserved? Does it have protein homologs in *Trypanosoma brucei*, plants, or bacteria?
- f. Take a look at the PFAM domain architectures found under the *PFam domains (graphic)* tab. Do all the proteins in this group have similar domain architecture?



- g. Based on the orthologs, what do you think this protein might be doing? If you had to give this gene a name, what would you call it?

## 2. Using the phyletic pattern tool in OrthoMCL

All EuPathDB sites have a phyletic pattern search that uses OrthoMCL data under *Genes* -> *Evolution* -> *Orthology Phylogenetic Profile*. This search is very useful to identify genes in your organism of interest that are restricted in their phylogenetic profile. For example, you may be interested in identifying genes that are conserved among organisms in your genus but not present in the host as these genes may make good drug targets or vaccine candidates.

- How many protein groups in OrthoMCL **do not** have any orthologs in bacteria or archaea? *Hint: go to the Phyletic Pattern search in the Evolution section of the Identify Ortholog Groups category. To specify a phyletic pattern, click on the icon next to the taxonomic group or species to include or exclude it.*

**OrthoMCL DB** Version 5.10 May 13  
A EuPathDB Project

Groups Quick Search:  synch\* Sequences Quick Search:  synch\*

About OrthoMCL Help Login Register Contact Us

Home New Search My Strategies My Basket (0) Tools Data Summary Downloads Community My Favorites

**Data Summary**

- Genomes: 150
- Protein Sequences: 1,398,546
- Ortholog Groups: 124,740

**News and Tweets**

**Community Resources**

**Identify Ortholog Groups**

- Text, IDs
- Group ID(s)
- Text Terms
- Evolution
  - Phyletic Pattern
  - Function

**Identify Protein Sequences**

- Text, IDs
- Sequence ID(s)
- Group ID(s)
- Text Terms
- Function
  - PFam ID or Keyword
  - Sequence Conservation
  - Annotations

**Tools:**

- BLAST
- Assign your proteins to groups
- Download OrthoMCL software
- Web Services
- Publications mentioning OrthoMCL

### Identify Groups based on Phyletic Pattern

Find Ortholog Groups that have a particular phyletic pattern, i.e., that include or exclude taxa or species that you specify.

The search is controlled by the Phyletic Pattern Expression (PPE) shown in the text box. Use either the text box or the graphical tree display, or both, to specify your pattern. The graphical tree display is a friendly way to generate a pattern expression. You can always edit the expression directly. For PPE help see the [instructions at the bottom of this page](#).

In the graphical tree display:

- Click on +/- to show or hide subtaxa and species.
- Click on the icon to specify which taxa or species to include or exclude in the profile.
- Refer to the legend below to understand other icons.

Expression:

**Key:** =no constraints | =must be in group | =at least one subtaxon must be in group | =must not be in group | =mixture of constraints

**Root (ALL):**

- Bacteria (BACT):**
  - Firmicutes (FIRM):** bant cbot cper lmon saur spne
  - Proteobacteria (PROT):**
    - alpha-Proteobacteria (PROA):** atum bsui tpro rtyp wend
    - beta-Proteobacteria (PROB):** bmal bpse rsol
    - delta-Proteobacteria (PROD):** gsul
    - gamma-Proteobacteria (PROG):** cbur ecol ftul sent stle vcho yent ypes
    - epsilon-Proteobacteria (PROE):** cjei wsuc
  - Other Bacteria (OBAC):** aaeo cpne ctep deth drad mtub mlep rbal syne tmar tpal
- Archaea (ARCH):**
  - Euryarchaeota (EURY):** aful halo hwal msmi mjan mmar tvol
  - Crenarchaeota (CREN):** cmaq hbut ihos msed nmar smar ssol
  - Nanoarchaeota (NANO):** nequ
  - Korarchaeota (KORA):** ckor
- Eukaryota (EUKA):**
  - Alveolates (ALVE):**
    - Ciliates (CIL):** the
    - Apicomplexa (APIC):**
      - Coccidia (COCC):** chom cmur cpar ncan tgon

- Modify the above search to identify protein groups that do not contain orthologs from eukaryotes? How many did you find?
- Find all groups that contain orthologs from at least one species of *Cryptococcus* (cneo, cneg) and *Aspergillus* (anid, afum) but not from any bacteria or archaea. To run this search, you will have to use the expression panel at the top of the search. (Remember this question is requiring that the groups must contain at least one *Cryptococcus* and one *Aspergillus* species). **See the description and help at the bottom of the search page.**

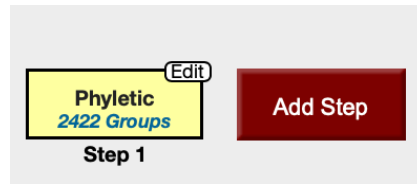
In the graphical tree display:

- Click on +/- to show or hide subtaxa and species.
- Click on the icon to specify which taxa or species to include or exclude in the profile.
- Refer to the legend below to understand other icons.

Expression:

**Key:** =no constraints | =must be in group | =at least one subtaxon must be in group | =must not be in group | =mixture of constraints

The correct expression to use is: BACT=0T AND ARCH=0T AND cneo+cneg>=1T AND afum+anid>=1T



- What happens if you revise the strategy to increase the stringency where *Cryptococcus bacillisporus* (cneo), *Cryptococcus neoformans var grubii* H99 (cneg) *Aspergillus nidulans* (anid), and *Aspergillus fumigatus* (afum) must be in the group?
- Find all groups that contain at least one subtaxon from *Basidiomycota* and *Ascomycota* but not from bacteria or archaea, or microsporidia

Key: =no constraints | =must be in group | =at least one subtaxon must be in group | =must not be in group | \* =mixture of constraints

\*Root (ALL):

Bacteria (BACT):

Firmicutes (FIRM):

Proteobacteria (PROT):

alpha-Proteobacteria (PROA):

beta-Proteobacteria (PROB):

delta-Proteobacteria (PROD):

gama-Proteobacteria (PROG):

epsilon-Proteobacteria (PROE):

Other Bacteria (OBAC):

Archaea (ARCH):

Euryarchaeota (EURY):

Crenarchaeota (CREN):

Nanoarchaeota (NANO):

Korarchaeota (KORA):

\*Fungi (FUNG):

Microsporidia (MICR):

Basidiomycota (BASI):

Ascomycota (ASCO):

bant cbot cper lmon saur spne

atum bsui rpro rtyp wend

bmal bpse rsol

gsul

cbur ecol ftul sent sfle vcho yent ypes

cjej wsuc

aaeo cpne ctep deth drad mtub mlep rbal syne tmar tpal

aful halo hwal msmi mjan mmr tvol

cmaq hbut ihos msed nmar smar ssol

nequ

ckor

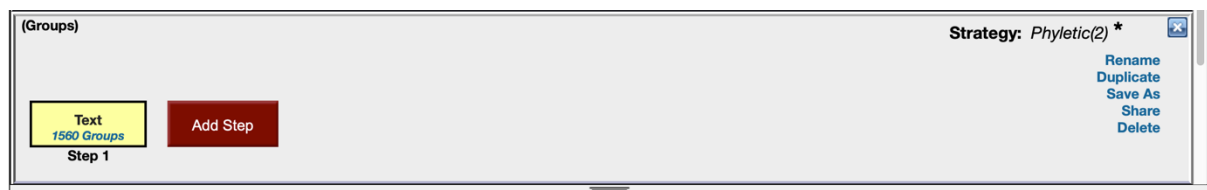
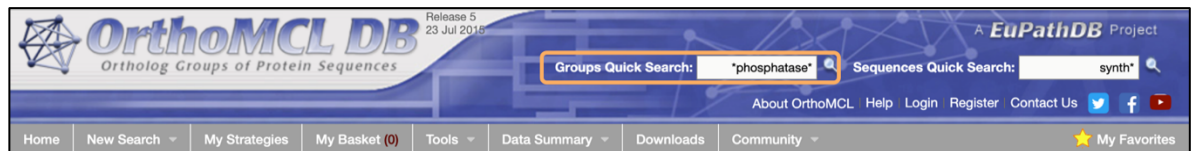
ecun ebie eint

cneo cneg lbic pchr

afum anid aory cgla cimm cpos dhan egos gzea klac mgri ncra psti scer calb spom

### 3. Combining searches in OrthoMCL.

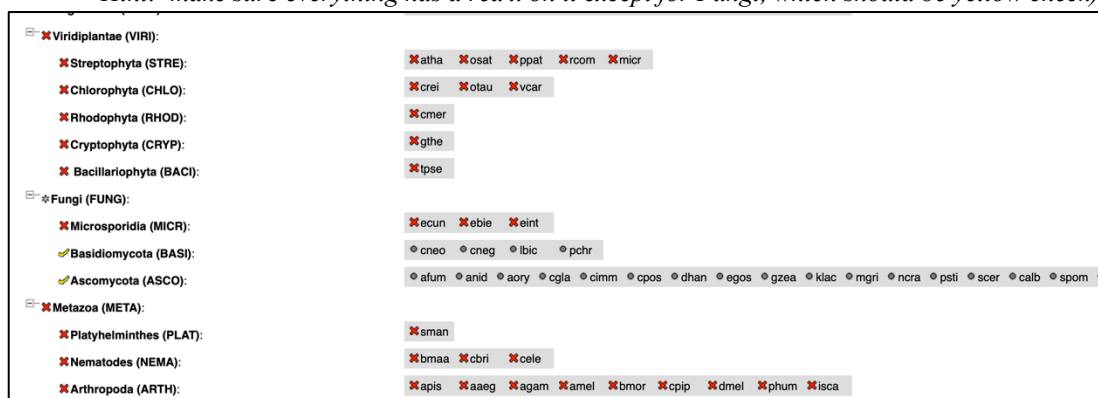
- Find all plant proteins that are likely phosphatases that do not have orthologs outside of plants.
- Use the text search to find OrthoMCL groups that contain the word \*phosphatase\*. Use wild cards to search for any word combinations.



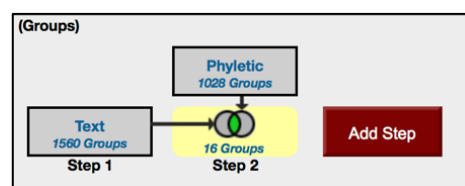
1560 Groups from Step 1 [Revise](#)  
Strategy: *Phyletic(2)*

Group Results		Phyletic pattern					
First 1 2 3 4 5 Next Last		Advanced Paging		Download	Add to Basket	Add Columns	
Group	# Sequences	Average % Identity	Average % Connectivity	EC Numbers	Score	Found in	
OG5_126538	1816	42.6	33.8	N/A	100	Sequences	
OG5_126657	239	33.1	24.6	3.1.3.16 (1)	100	EcNumber, PFams, Sequences	

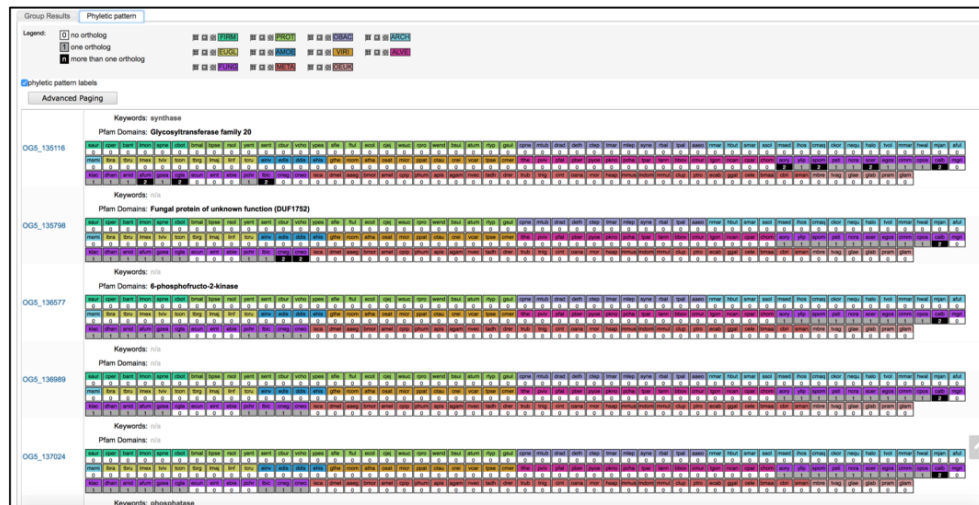
- Add a step and run a phyletic pattern search for groups that contain at least one subtaxon from Ascomycota and Basidiomycota.  
*Hint: make sure everything has a red x on it except for Fungi, which should be yellow check).*



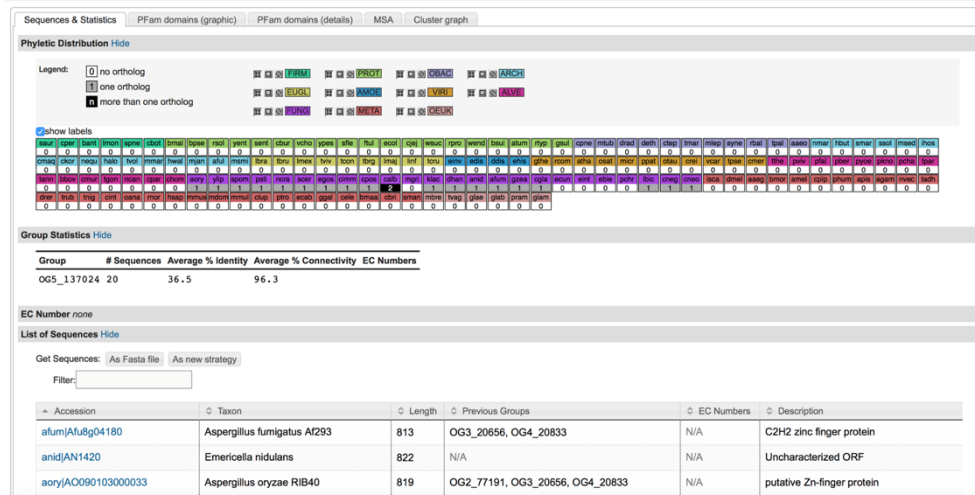
- How many groups did you return?



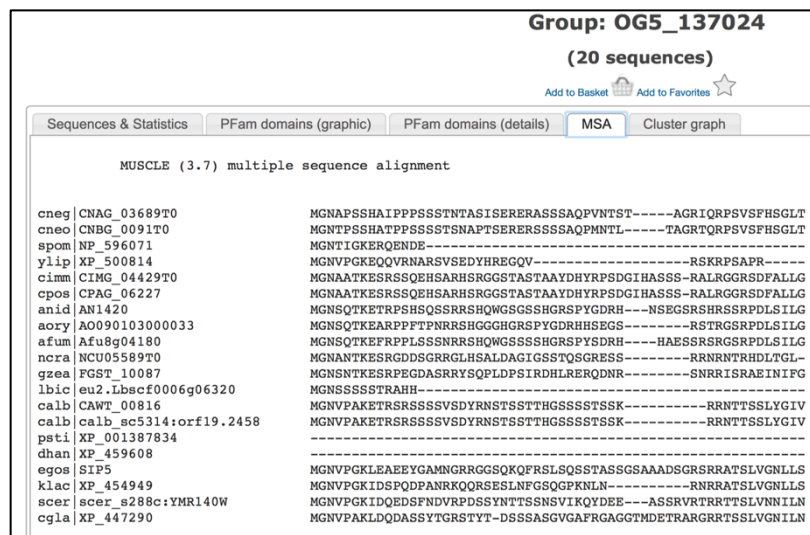
- Explore the phyletic pattern.



- Click on the group OG5\_137024. Explore the multiple sequence alignments and Cluster graphs.



- Explore the multiple sequence alignments from some of these groups and nodes for selected species. (Hint: click on a *group ID* and then open the *MSA* and *Cluster graph* tabs).



- *Examine the Cluster Graph tab:* Modify the E-value cutoff slider.
  - What happens when you increase the E-value?
  - What happens when you decrease the E-value?
  - Can you identify subclusters?

