

## Basic OrthoMCL Functionality

<http://orthomcl.org/>

### Learning objectives:

- Run searches in OrthoMCL
- Run phyletic pattern searches using check boxes or an expression
- Combine searches using the strategy system
- Explore individual ortholog group pages
- Explore the group cluster graphs

### 1. Using the Phyletic Pattern search in OrthoMCL

The “Phyletic Pattern” search is an ortholog group search – look under the ortholog groups category and explore the available searches. Can you find the one called “Phyletic Pattern”? There are two ways to specify a phyletic pattern:

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

The screenshot shows the OrthoMCL DB website. The top navigation bar includes the site logo, release information (6.2 beta, 17 Dec 2020), a search bar, and links to My Strategies, Searches, Tools, My Workspace, Data, About, Help, and Contact Us. The sidebar on the left has a 'Search for...' section with 'expand all | collapse all' and a filter dropdown. Under 'Ortholog Groups', several search options are listed, including 'Phyletic Pattern'. The main panel is titled 'Overview of Resources and Tools' and features a section for 'Identify Ortholog Groups based on Phyletic Pattern'. This section includes instructions on how to use the Phyletic Pattern Expression (PPE) and a graphical tree display. The graphical tree display shows a taxonomic tree with checkboxes for various taxa. The expression box contains the text 'BACT-QT AND ARCH-QT'. The bottom of the page features logos for BRC, NIAID Bioinformatics Resource Centers, and the VEuPathDB Project Team, along with a 'COMMUNITY CHAT' button.

**Search for...**  
expand all | collapse all  
Filter the searches below...

**Ortholog Groups**

- Q % Pairs w/ Similarity
- Q All Groups
- Q Avg % Homology
- Q Avg % Identity
- Q Avg % Match Length
- Q Avg E-Value
- Q EC Number (specific)
- Q Group ID(s)
- Q Group or Sequence ID
- Q Number of Sequences
- Q PFam ID or Keyword
- Q Phyletic Pattern
- Q Text Terms

**Proteins**

**Overview of Resources and Tools**

**Identify Ortholog 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 the icons to show or hide subtaxa and species.
- Click on the icons 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

expand all | collapse all  
Type a taxonomic name

- Root (ALL)
- ✗ Bacteria (BACT)
- ✗ Firmicutes (FIRM)
- ✗ Proteobacteria (PROT)
- ✗ Other Bacteria (OBAC)
- ✗ Archaea (ARCH)
- ✗ Nitrososphaera maritima (strain SCM1) (nmar)
- ✗ Euryarchaeota (EURY)
- ✗ Crenarchaeota (CREN)
- ✗ Nanoarchaeota (NANO)
- ✗ Korarchaeota (KORA)
- ✗ Eukaryota (EUKA)
- ✗ Alveolates (ALVE)
- ✗ Amoebozoa (AMOE)
- ✗ Euglenozoa (EUGL)
- ✗ Viridiplantae (VIRI)
- ✗ Fungi (FUNG)
- ✗ Metazoa (META)
- ✗ Other Eukaryota (OEUK)

©2020 The VEuPathDB Project Team

1. Using the expression box. Type the expression using hints available at the bottom of the search page.

2. Using the selectable tree menu. Click on the circle next to the taxon you want to include or exclude.


- a. How many protein groups do not contain orthologs from bacteria and archaea?
- b. Find all groups that contain orthologs from at least one species of *Cryptosporidium* and *Giardia* but not from bacteria or archaea. If you are getting frustrated trying to figure this one out, you have a right to be! You cannot answer this question by using the check boxes. However, OrthoMCL has an added feature that allows you to enter an expression to define the phyletic pattern. This option provides additional flexibility. Can you figure out what expression to use to answer this question? (hint: scroll down to the bottom of the search page to find additional information about expression parameters).

Before looking at the answer below, try this on your own to see if you can figure it out.

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:

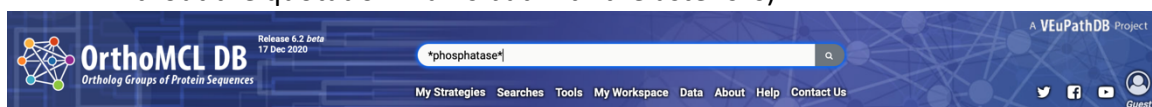
BACT=0T AND ARCH=0T AND cpar+cand+choi+chot+chom+chod+cmel+cmur+cpia+ctyz+cubi>=1T AND gass+gadh+gasb+gabb+gase+gmur>=1T

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

## 2. Combining searches in OrthoMCL

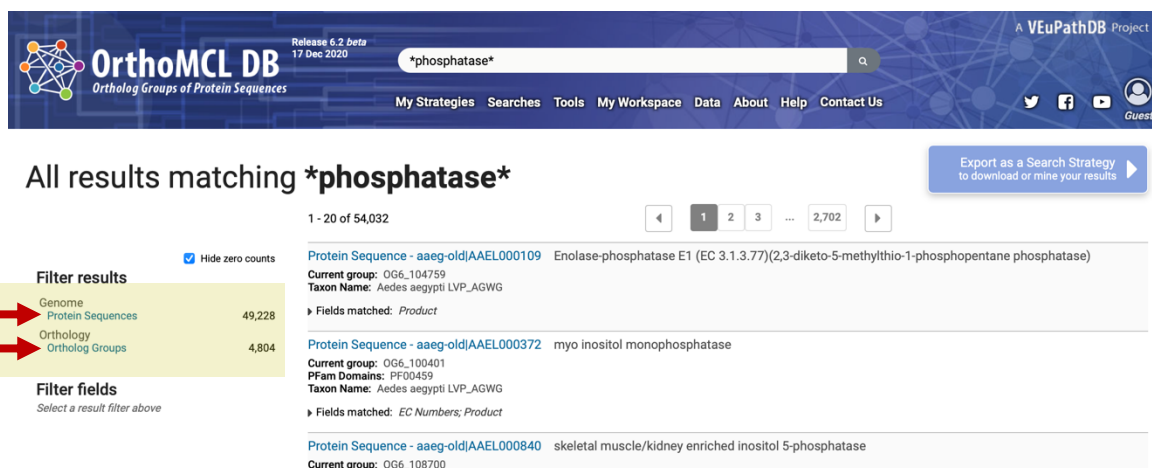
Find all plant proteins that are likely phosphatases that do not have orthologs outside of plants.

- a. Use the site search box at the top of the page in the header to **find OrthoMCL groups** that contain the word “\*phosphatase\*” (note that the search should be run without the quotation marks but with the asterisks).



- How many proteins did you identify?
- How many groups did you identify?

**Note: that numbers in screen shots will likely be different on the site due to frequent updates.**



OrthoMCL DB  
Orthology Groups of Protein Sequences

Release 6.2 beta  
17 Dec 2020

\*phosphatase\*

My Strategies Searches Tools My Workspace Data About Help Contact Us

A VEuPathDB Project

Export as a Search Strategy  
to download or mine your results

All results matching **\*phosphatase\***

1 - 20 of 54,032

Filter results

☒ Hide zero counts

Genome 49,228

Orthology 4,804

Filter fields

Select a result filter above

Protein Sequence - aaeg-old|AAEL000109 Enolase-phosphatase E1 (EC 3.1.3.77)(2,3-diketo-5-methylthio-1-phosphopentane phosphatase)

Current group: OG6\_104759

Taxon Name: Aedes aegypti LVP\_AGWG

Fields matched: Product

Protein Sequence - aaeg-old|AAEL000372 myo inositol monophosphatase

Current group: OG6\_100401

Pfam Domains: PF00459

Taxon Name: Aedes aegypti LVP\_AGWG

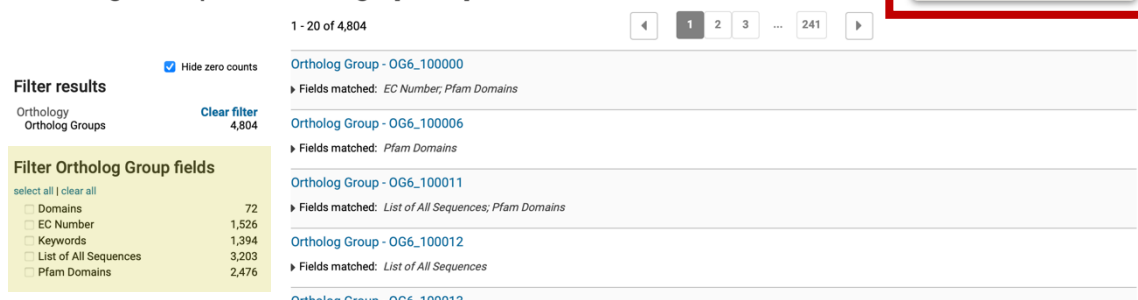
Fields matched: EC Numbers; Product

Protein Sequence - aaeg-old|AAEL000840 skeletal muscle/kidney enriched inositol 5-phosphatase

Current group: OG6\_108700

- Click on ortholog groups to filter the results to only show the ortholog groups containing the word phosphatase. Notice that you can filter the group results even further by the group fields. Also, notice that the export as a Search Strategy button is now active. (Note: this is because all the results are now only one type of record: groups.)
- Export the ortholog group results as a search strategy by clicking on the blue “Export as a Search Strategy” button at the top right of the page.

## Ortholog Groups matching **\*phosphatase\***



Ortholog Groups matching **\*phosphatase\***

1 - 20 of 4,804

Filter results

☒ Hide zero counts

Orthology 4,804

Ortholog Groups 4,804

Filter Ortholog Group fields

select all | clear all

Domains 72

EC Number 1,526

Keywords 1,394

List of All Sequences 3,203

Pfam Domains 2,476

Ortholog Group - OG6\_100000

Fields matched: EC Number; Pfam Domains

Ortholog Group - OG6\_100006

Fields matched: Pfam Domains

Ortholog Group - OG6\_100011

Fields matched: List of All Sequences; Pfam Domains

Ortholog Group - OG6\_100012

Fields matched: List of All Sequences

Ortholog Group - OG6\_100013

Export as a Search Strategy  
to download or mine your results

## My Search Strategies

Opened (1) All (25) Public (7) Help

Unnamed Search Strategy \*

4,804 Ortholog Groups

Revise this search

Ortholog Group Results

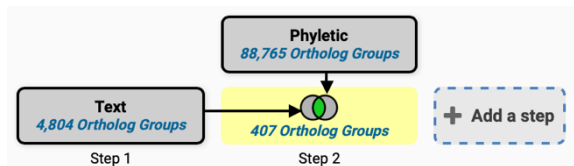
Rows per page: 50

Download Add to Basket Add Columns

Ortholog Group	Score	Total Number Proteins	Keywords	PFam Domains	Archaea	Bacteria	Alveolata	Amoeba	Euglenozoa	Fungi
OG6_100000	1	12129	N/A	N/A	0 / 27 (0%)	12 / 47 (26%)	62 / 108 (57%)	4 / 13 (31%)	22 / 52 (42%)	133 / 171 (78%)

- d. Add a step and run a phyletic pattern search for groups that contain any plant protein but do not contain any other organism outside plants. (hint: make sure everything has a red x on it except for plants (Viridiplantae (VIRI)), which should be a grey circle).

- e. How many groups did you return?



- \* Root (ALL)
  - \* Bacteria (BACT)
    - \* Firmicutes (FIRM)
    - \* Proteobacteria (PROT)
    - \* Other Bacteria (OBAC)
  - \* Archaea (ARCH)
    - \* Nitrosopumilus maritimus (strain SCM1) (nmar)
    - \* Euryarchaeota (EURY)
    - \* Crenarchaeota (CREN)
    - \* Nanoarchaeota (NANO)
    - \* Korarchaeota (KORA)
  - \* Eukaryota (EUKA)
    - \* Alveolates (ALVE)
    - \* Amoebozoa (AMOE)
    - \* Euglenozoa (EUGL)
    - \* Viridiplantae (VIRI)
    - \* Fungi (FUNG)
    - \* Metazoa (META)
    - \* Other Eukaryota (OEUK)

### 3. Exploring a specific OrthoMCL group - examining the cluster graph.

- a. Visit the OrthoMCL group OG6\_131670. You can quickly get to this group by first running a site search with the group ID. Click on the group ID to get to its page.

- b. The group page is divided into 5 sections:

- Phyletic distribution
- Group summary
- List of proteins
- PFam domains
- Cluster graph

Examine each of the above sections - is it clear what each section contains?

c. Examine the phyletic distribution tree. What taxa does this group contain?

▼ Phyletic Distribution of Proteins [Download](#)

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

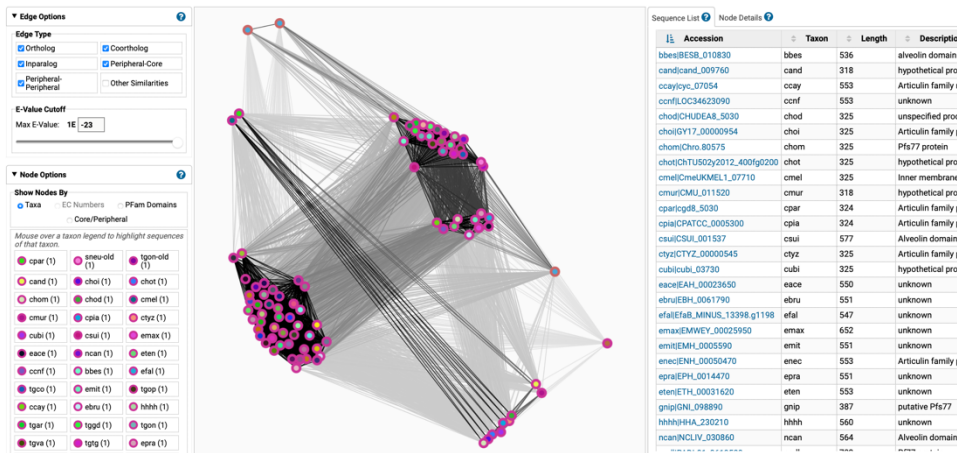
☒ Hide zero counts

Type a taxonomic name	?
▼ Eukaryota (EUKA)	93
▼ Alveolates (ALVE)	90
▼ Apicomplexa (APIC)	90
Gregarina niphandrodes Unknown strain (gnip)	1
▼ Coccidia (COCC)	44
▼ Aconoidasida (ACON)	45
▼ Haemosporida (HAEM)	45
▼ Metazoa (META)	3
▼ Chordata (CHOR)	3
▼ Mammalia (MAMM)	3
Physeter macrocephalus (Sperm whale) (Physeter catodon) (phma)	3

d. Examine the cluster graph for this group (hint: go to the cluster graph section of the page and then click on the “Click to open the Cluster graph in a new tab”)

Cluster Graph: OG6\_131670 (93 proteins) ?

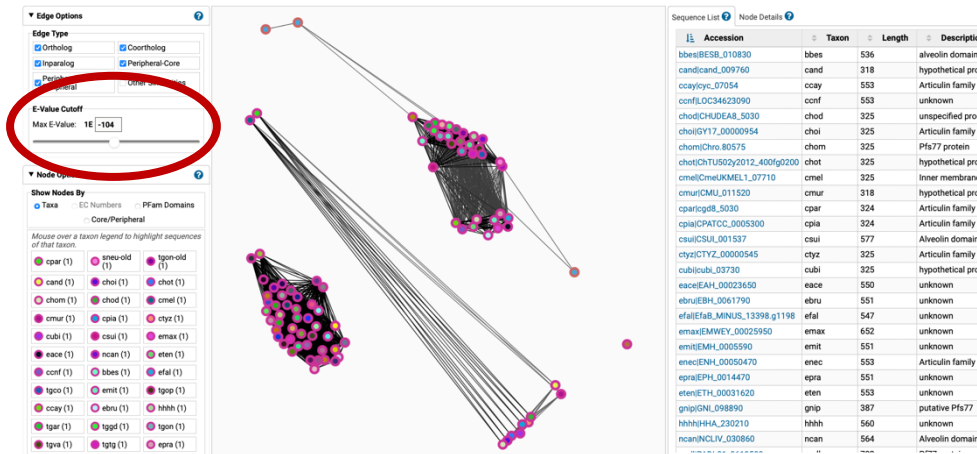
[Back to Group page](#)



You can interact with the cluster graph. For example, move the slide to increase the E-value cutoff stringency (e.g., to a more negative number). Can you identify subclusters? Click on the nodes in the graph – notice how the organism is updated on the right.

Cluster Graph: OG6\_131670 (93 proteins) ?

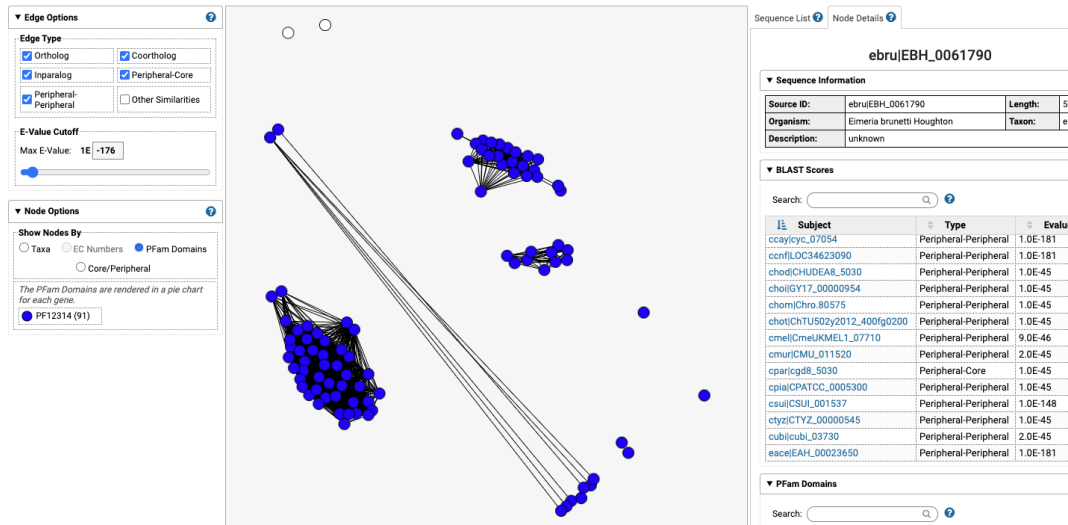
[Back to Group page](#)



On the left of the page in the Node Options panel, click on PFam Domains to see which proteins have the various PFam domains.

#### Cluster Graph: OG6\_131670 (93 proteins) ?

[Back to Group page](#)



And in the same panel, click on Core/Peripheral to observe which proteins were derived from Core species and which proteins were derived from Peripheral species. Proteins from Core species were used in the initial OrthoMCL algorithm to form Core ortholog groups. Proteins from Peripheral species were mapped into these Core groups by sequence similarity (determined by BLAST score).

#### Cluster Graph: OG6\_131670 (93 proteins) ?

[Back to Group page](#)

