# Basic OrthoMCL Functionality

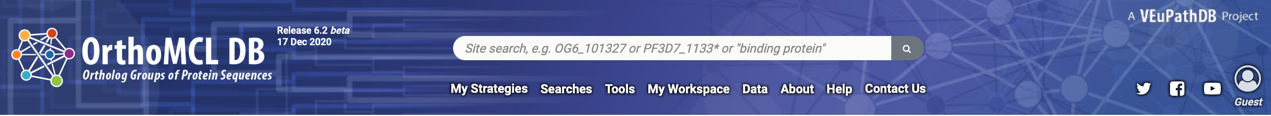
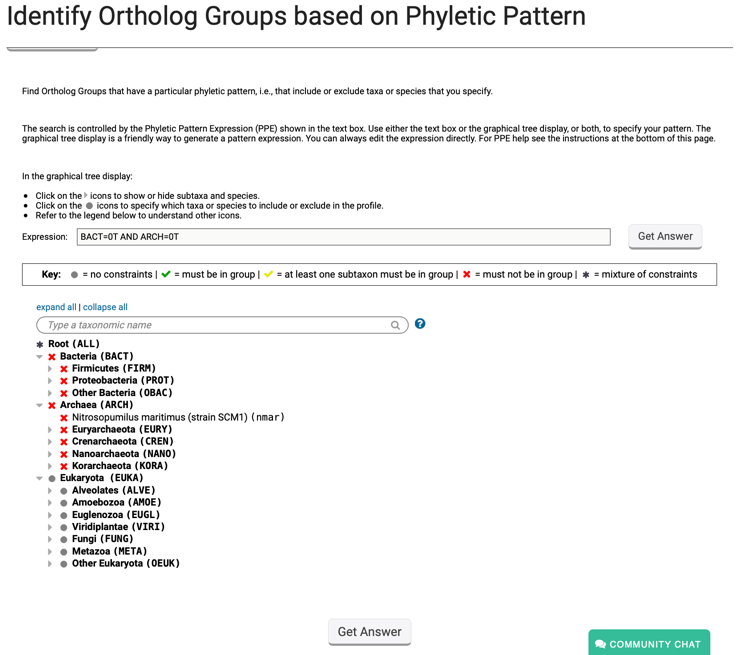
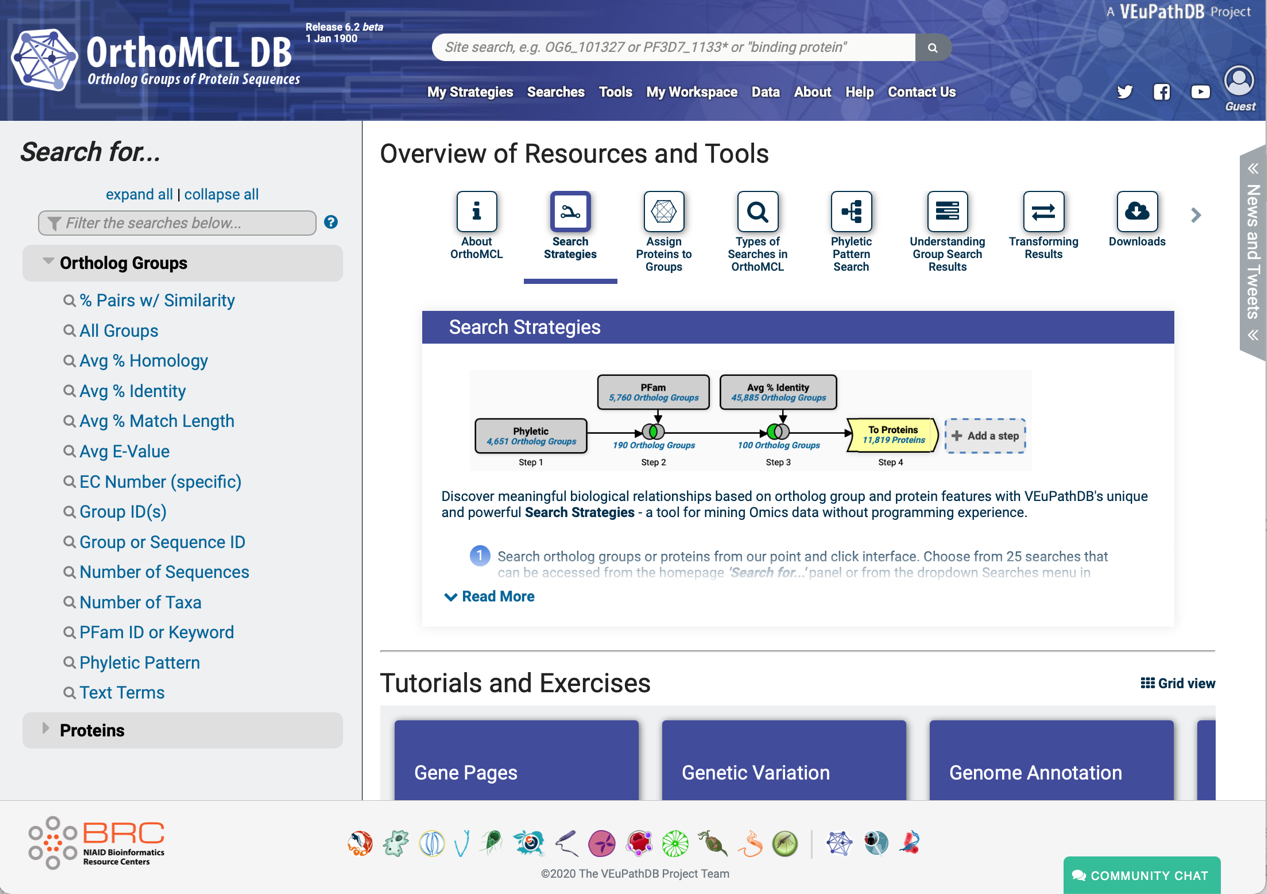
# <http://orthomcl.org/>

**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:

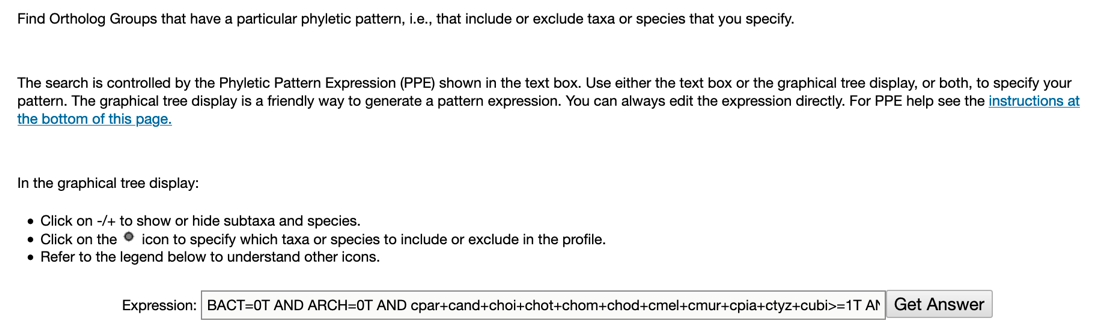


1. Using the expression box. Type the expression using hints available at the bottom of the search page.
2. Using the selectable tree menu. Click on the circle next to the taxon you want to include or exclude.
3. How many protein groups do not contain orthologs from bacteria and archaea?
4. Find all 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.

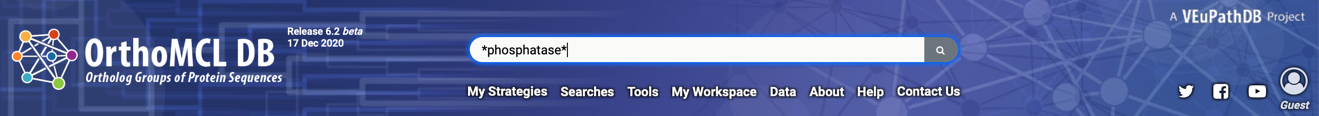
All VEuPathDB sites also have a phyletic pattern search that uses OrthoMCL data under *Genes -> Orthology and Synteny -> Orthology Phylogenetic Profile*. This search is very useful to identify genes in your organism of interest that are restricted in their profile. For example, you frequently want to 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.

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

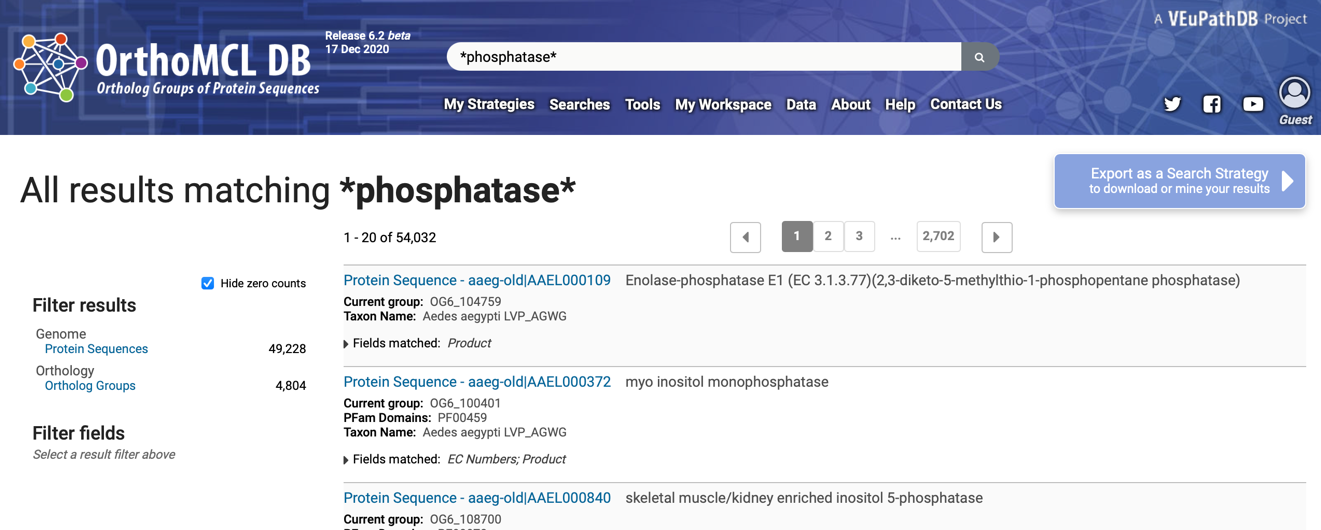


1. **Combining searches in OrthoMCL**

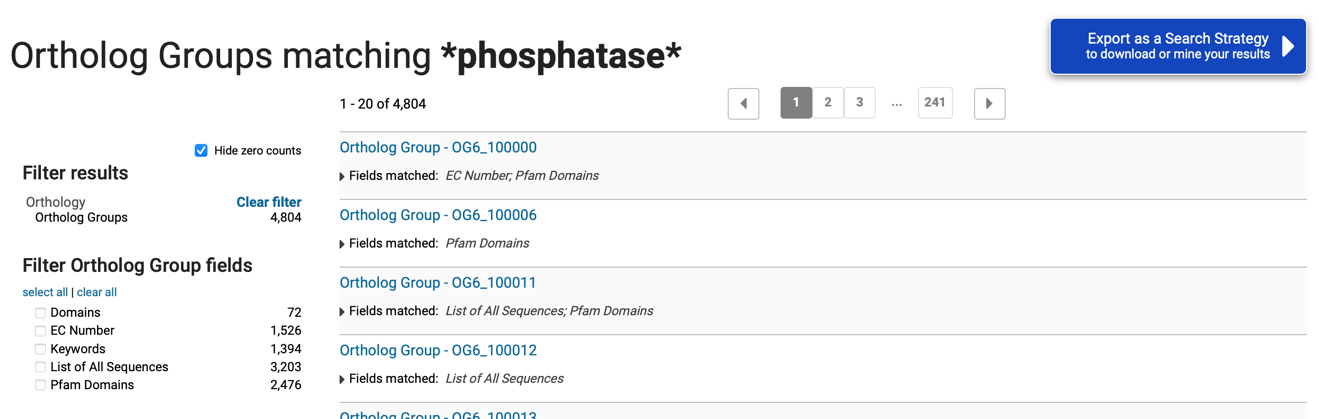
Find all plant proteins that are likely phosphatases that do not have orthologs outside of plants.

1. 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.***

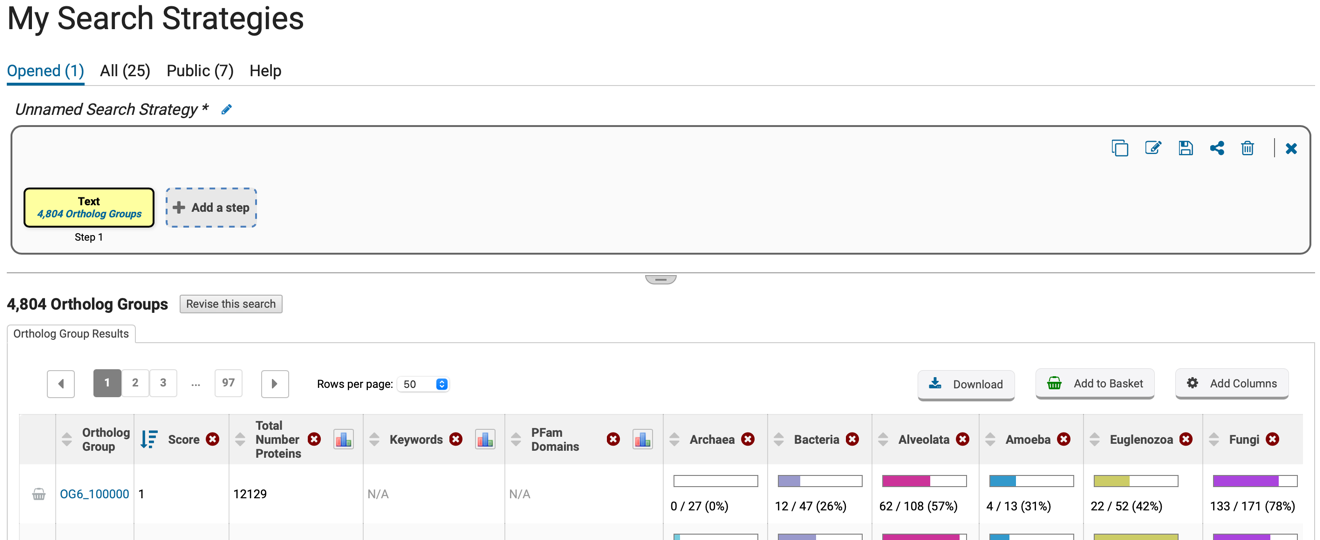


1. Click on ortholog groups to filter the results to only show the ortholog groups containing the word phosphatase. Notice that you can filter the group results even further by the group fields. Also, notice that the export as a Search Strategy button is now active. (Note: this is because all the results are now only one type of record: groups.)



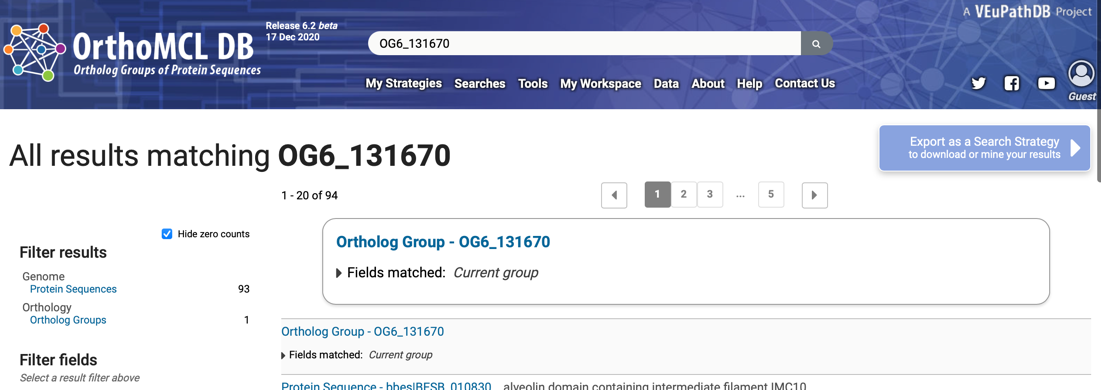
1. Export the ortholog group results as a search strategy by clicking on the blue “Export as a Search Strategy” button at the top right of the page.
2. A picture containing text

   Description automatically generatedAdd 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).



1. Diagram

   Description automatically generatedHow many groups did you return?
2. **Exploring a specific OrthoMCL group - examining the cluster graph.**
3. 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.



1. The group page is divided into 5 sections:
   1. Phyletic distribution
   2. Group summary
   3. List of proteins
   4. PFam domains
   5. Cluster graph

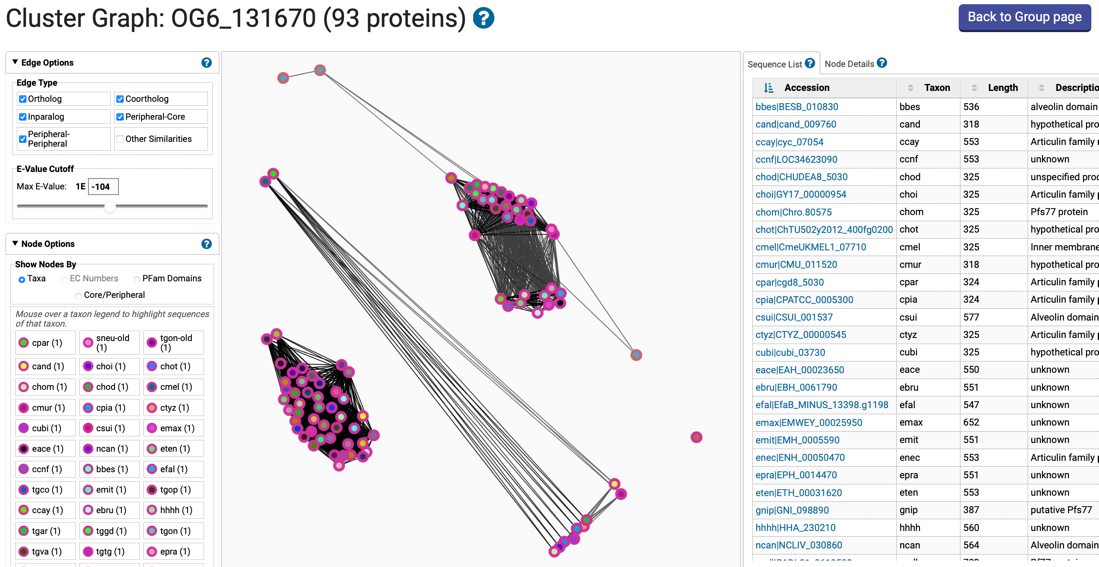
Examine each of the above sections – is it clear what each section contains?

1. Graphical user interface, text, application, email

   Description automatically generatedExamine the phyletic distribution tree. What taxa does this group contain?
2. Chart

   Description automatically generatedExamine 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”*)

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.



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

Chart, scatter chart

Description automatically generated

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).

Graphical user interface

Description automatically generated