Orthology and Phyletic Patterns
Exercise 8

1. **Getting to OrthoMCL from EuPathDB databases**
   Note: For this exercise use http://cryptodb.org and http://orthomcl.org

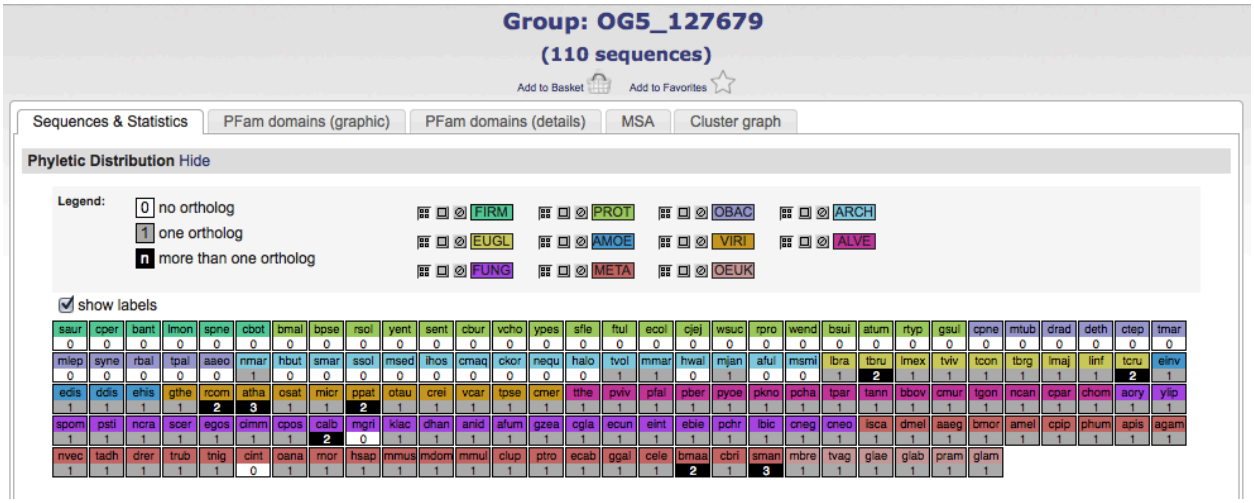
a. Go to the gene page for the *Cryptosporidium parvum* gene with the ID: cgd7_2290.
b. What does this gene do? It is annotated as a hypothetical protein!
c. Scroll down to the table labeled "Orthologs and Paralogs within CryptoDB". Does this gene have orthologs in other *Cryptosporidium* species? What about other organisms? (hint: click on the link below the table that takes you to OrthoMCL).

Orthologs and Paralogs within CryptoDB Hide

Gene	Organism	Product	is syntenic	has comments
Chro.70261	Cryptosporidium hominis TU502	hypothetical protein	yes	no
CMU_034340	Cryptosporidium muris RN66	hypothetical protein, conserved	yes	no

View the group (OG5_127679) containing this gene (cgd7_2290) in the OrthoMCL database

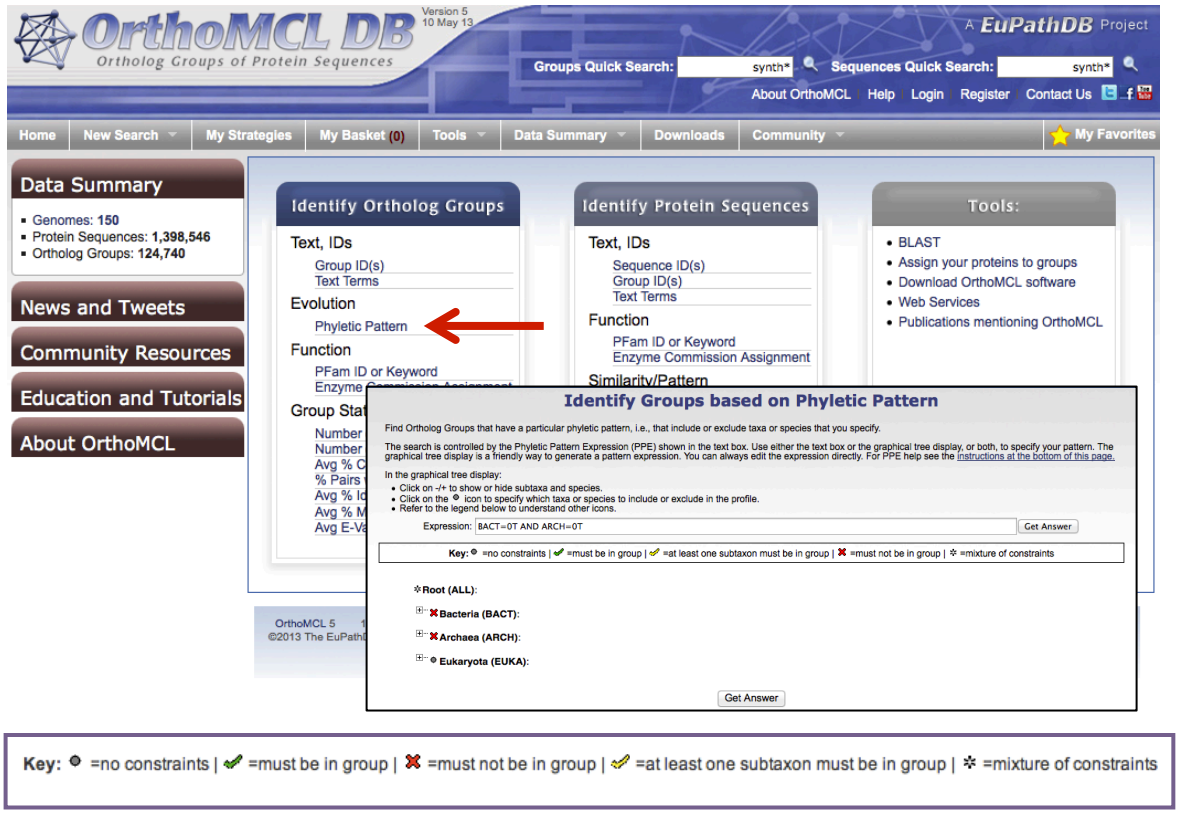
d. Does this protein have orthologs in other organisms? Does it have any orthologs in bacteria or archaea?
(hint: mouse over the colorful boxes in the table to reveal the full species and pylum names – see image below).



e. 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?
f. 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**
   Note: For this exercise use http://orthomcl.org/

   a. 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.



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

   b. How many protein groups do not contain orthologs from eukaryotes?

   c. Find all groups that contain orthologs from at least one species of *Cryptosporidium* and *Giardia* but not from bacteria or archaea.

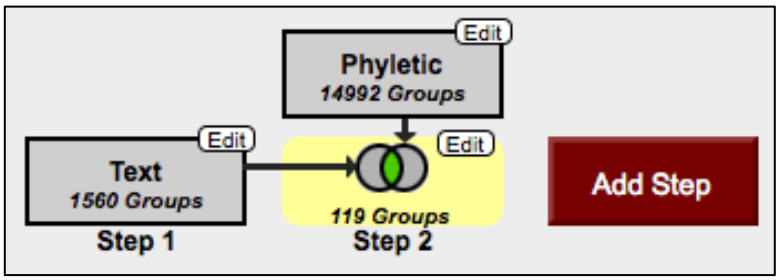
NOTE: All EuPathDB sites also have a phyletic pattern search that uses OrthoMCL data under Genes -> Evolution -> Orthology Phylogenetic Profile.

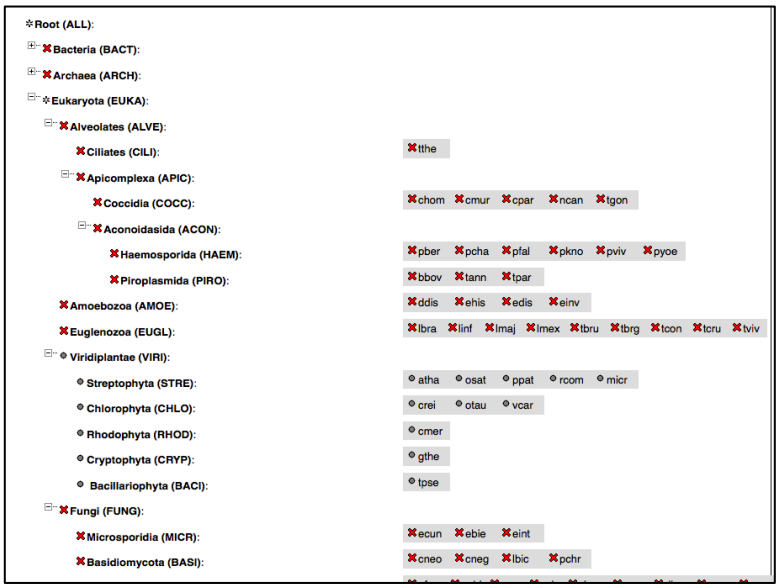
3. Combining searches in OrthoMCL

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

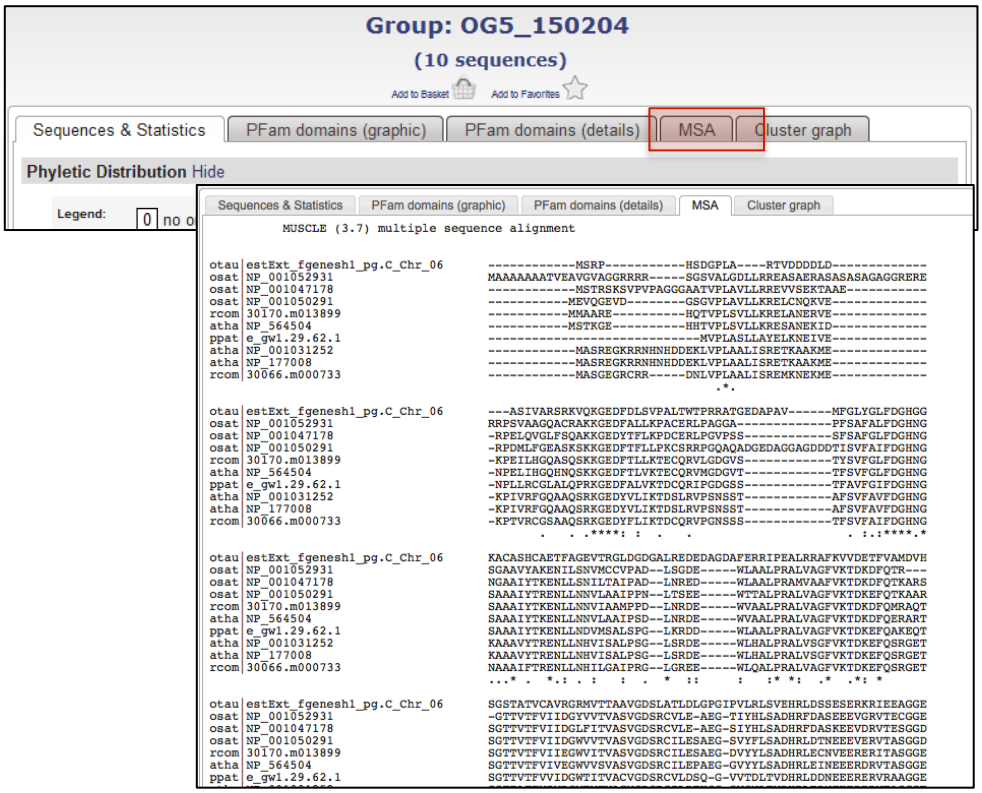
a. Use the text search to find groups that contain the word "*phosphatase*".

b. 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.).



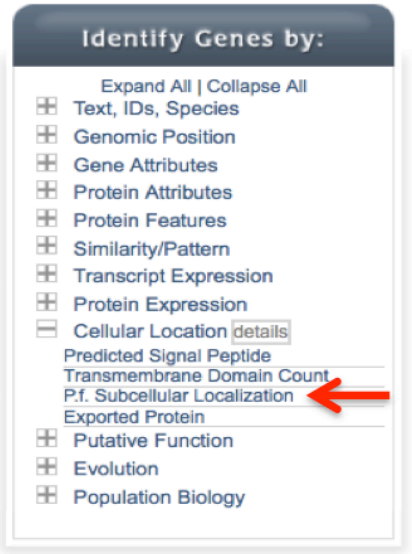


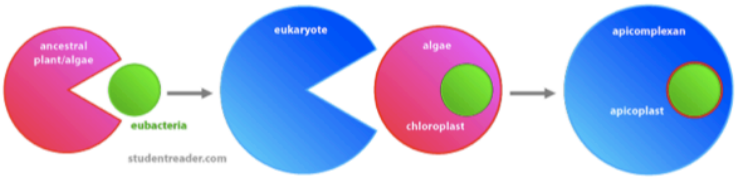
c. How many groups did you return? Explore the multiple sequence alignments from some of these groups. (Hint: click on a group ID and open the MSA tab).



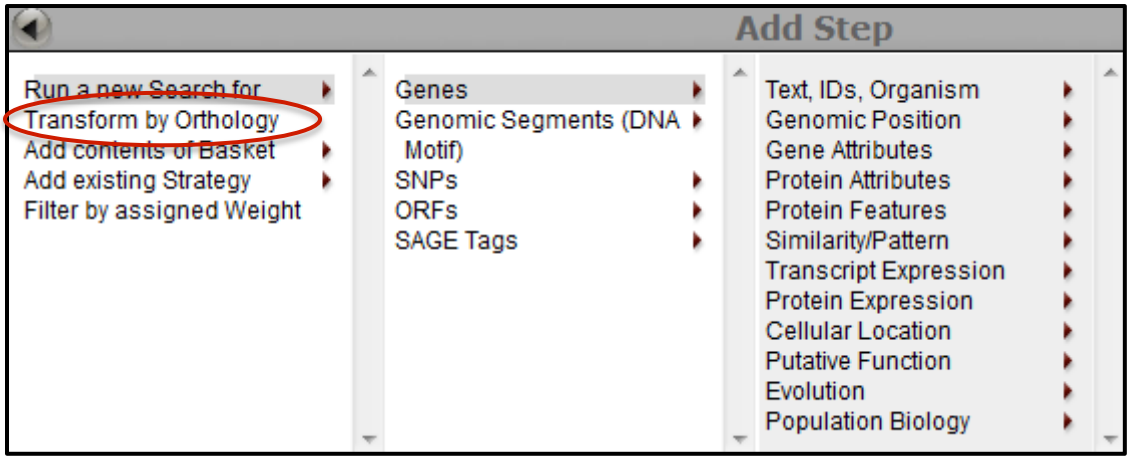
4. **(Optional) Using the orthology transform tool to identify apicoplast targeted genes in *Toxoplasma* and *Neospora*.**
   Note: For this exercise use http://eupathdb.org

a. Start by finding genes in *Plasmodium* that are predicted to target to the apicoplast. Hint: click on "Cellular Location" then on "P.f. Subcellular Localization"; see image below.





b. Transform the results of the above search to their *Toxoplasma* orthologs.
Hint: add a step, then select "Transform by Orthology". On the search page, select all *Toxoplasma* and *Neopspora*.



c. Although *Cryptosporidium* is an apicomplexan parasite it has actually lost its apicoplast! Can you use this fact to refine your results from the above search?

Hint: try subtracting out any orthologs present in *Cryptosporidium*. You will need to use a nested strategy.