

Orthology and Phyletic Patterns Exercise 9

9.1 Getting to OrthoMCL from EuPathDB databases

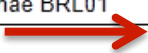
Note: For this exercise use <http://www.microsporidiadb.org>

- Go to the gene page for the *Encephalitozoon cuniculi* gene with the ID: ECU07_0290.
- What does this gene do? It is annotated as hypothetical!
- Scroll down to the table labeled “Orthologs and Paralogs within MicrosporidiaDB”. Does this gene have orthologs in other *Encephalitozoon* species? What about other organisms?

Hint: click on the link below the table that takes you to OrthoMCL.

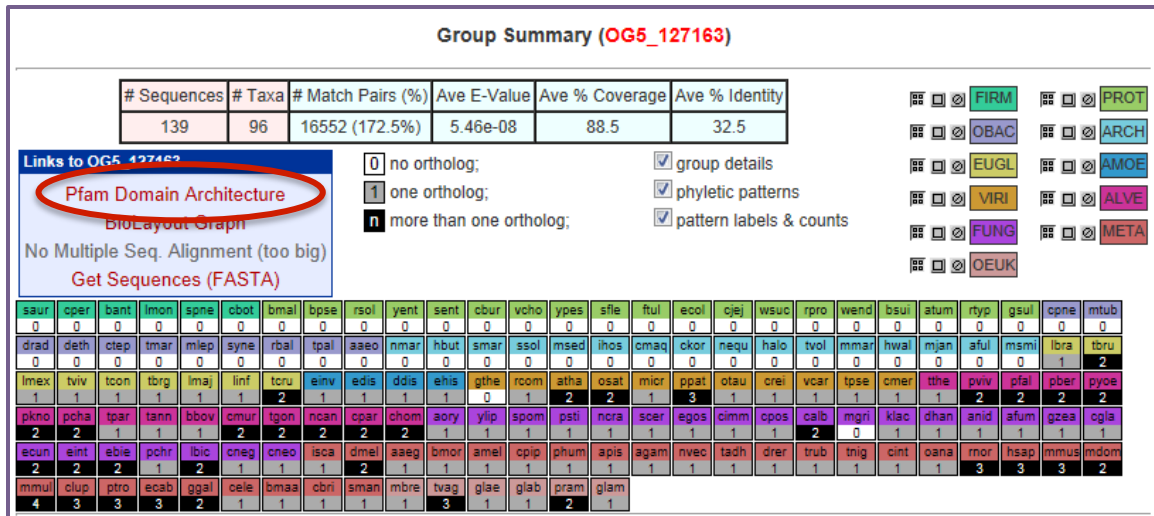
Orthologs and Paralogs within MicrosporidiaDB [Hide](#)

Gene	Organism	Product	is syntenic	has comments
ECU10_1730	Encephalitozoon cuniculi GB-M1	hypothetical protein	no	no
EHEL_070250	Encephalitozoon hellem ATCC 50504	hypothetical protein	yes	no
EHEL_101770	Encephalitozoon hellem ATCC 50504	hypothetical protein	no	no
Ein07_0230	Encephalitozoon intestinalis	hypothetical protein	yes	no
Ein10_1700	Encephalitozoon intestinalis	hypothetical protein	no	no
EBI_25559	Enterocytozoon bieneusi H348	hypothetical protein	yes	no
EBI_27325	Enterocytozoon bieneusi H348	hypothetical protein	no	no
NCER_100684	Nosema ceranae BRL01	hypothetical protein	no	no
NCER_101866	Nosema ceranae BRL01	hypothetical protein	no	no

 [Find the group containing ECU07_0290 in the OrthoMCL database](#)

- 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 tables to reveal the full species andylum names – see image below.



- e. Take a look at the PFAM domain architectures. 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?

9.2 Using the phyletic pattern tool in OrthoMCL

Note: For this exercise use <http://www.beta.orthomcl.org/>

- a. How many protein groups in OrthoMCL do not have any orthologs in bacteria or archaea?
Hint: go to “Search for Groups by Evolution...Phyletic Pattern”.

- b. How many protein groups do not contain orthologs from eukaryotes?
 Hint: click on the icon to specify which taxa or species to include or exclude in the profile.

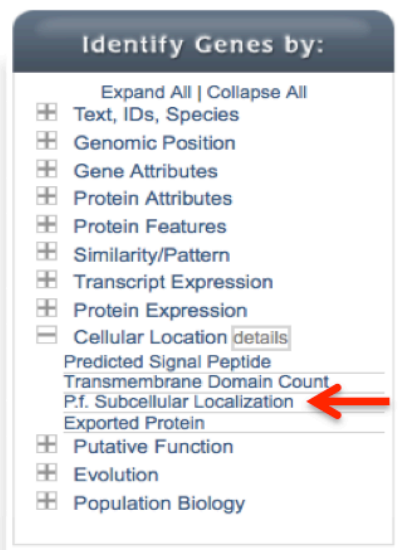
Key: ○ =no constraints | ✓ =must be in group | ✗ =must not be in group | ✓ =at least one subtaxon must be in group | ✳ =mixture of constraints

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

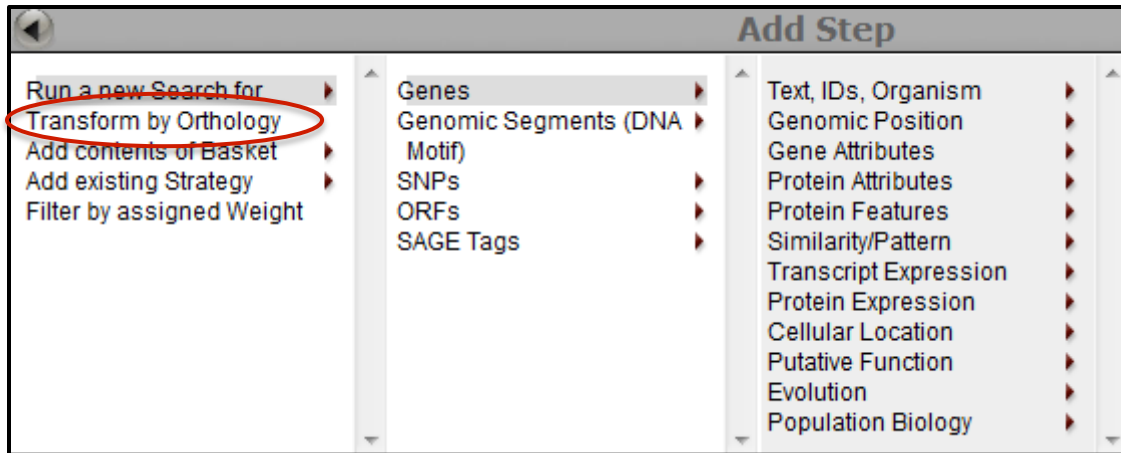
9.3 Using the orthology transform tool to identify apicoplast targeted genes in *Toxoplasma* and *Neospora*.

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.

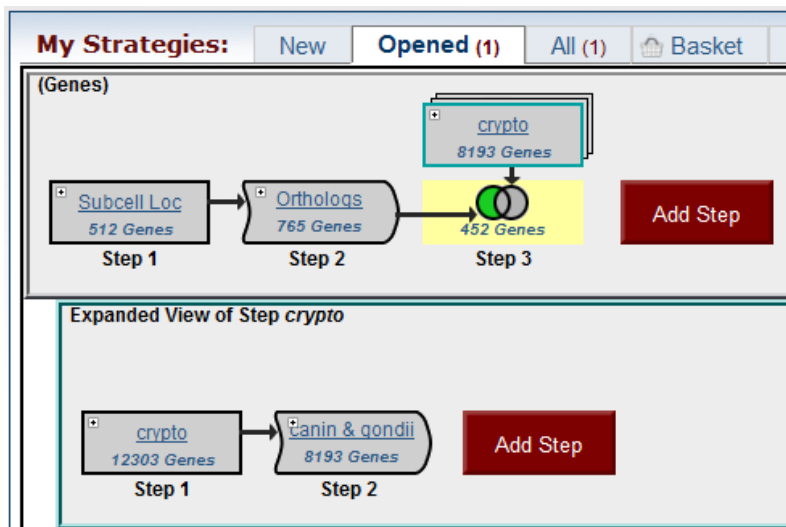


- b. Transform the results of the above search to their *Toxoplasma* orthologs.
 Hint: add a step, then select “Transform by Orthology”. On the search page, select all *Toxoplasma* and *Neospora*.



- 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?

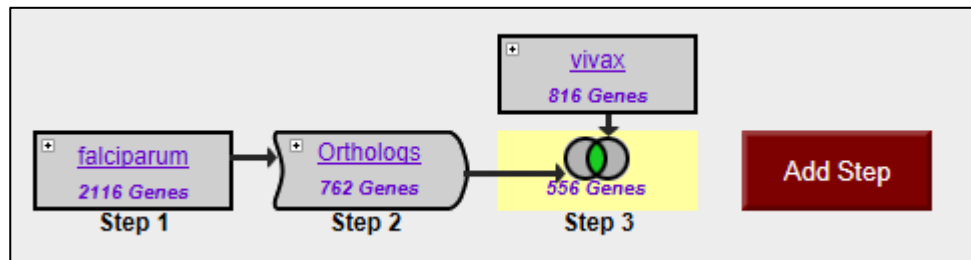
Hint: try subtracting out any orthologs present in *Cryptosporidium*. You will need to use a nested strategy.



9.4(Optional) Use the orthology transform tool to identify as many *P. falciparum* genes containing signal peptides as possible.

Note: For this exercise use <http://www.plasmodb.org>

- How many genes in *P.falciparum* are annotated with signal peptides (use default settings)?
- How many *P. vivax* genes are annotated with signal peptides (use default settings)?
- How many genes on these two lists are in common?
Hint, use the ortholog query to transform between organisms.

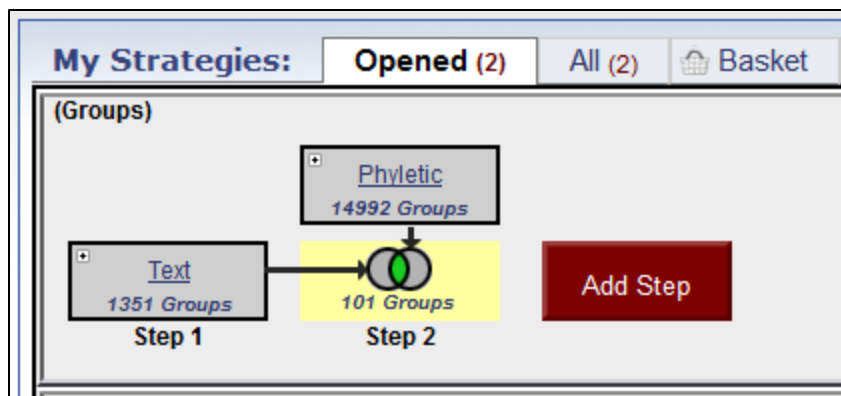


- How many *P. falciparum* orthologs of vivax genes with signal peptides do not themselves contain signal peptides? Why might this be the case? Look at a couple of these using the synteny viewer to generate some hypotheses.
- Using PlasmoDB, generate the most comprehensive list of *P. falciparum* genes that may contain signal peptides. How many did you find?

9.5 (optional) Integrated 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 groups that contain the word “phosphatase”.
- 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.



- 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.

The screenshot shows a detailed view of a group: "Group: OG5_150204" with "(10 sequences)". There are "Add to Basket" and "Add to Favorites" options. A navigation bar includes tabs for "Sequences & Statistics", "PFam domains (graphic)", "PFam domains (details)", "MSA" (highlighted with a red box), and "Cluster graph". Below the tabs, there is a "Phyletic Distribution Hide" section and a legend showing "0 no ortholog". At the bottom, there are colored boxes for "FIRM", "PROT", "OBAC", and "ARCH".