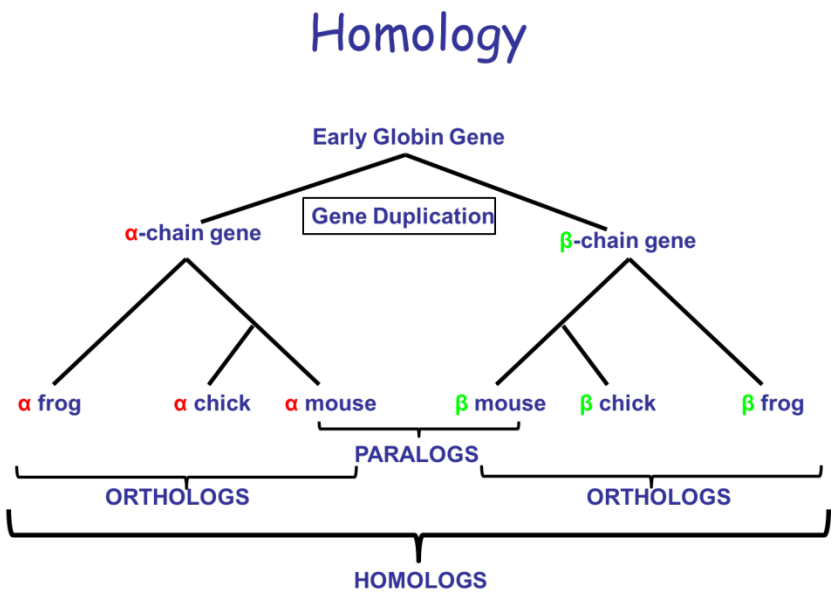
# Orthology and Phyletic Patterns



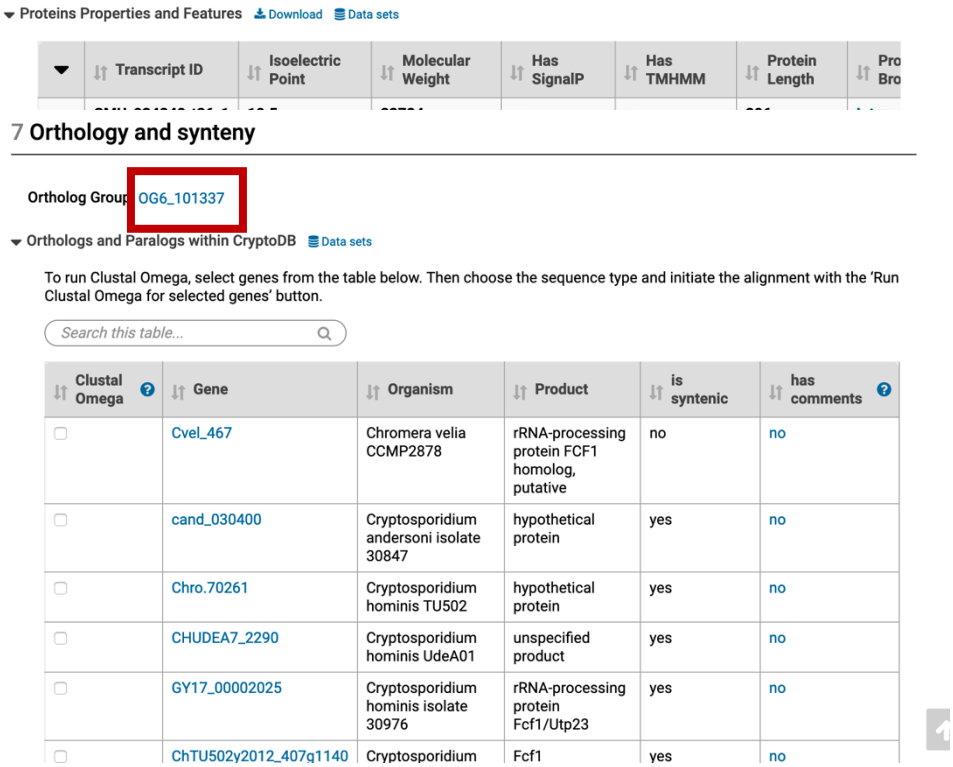

**Learning objectives:**

- Explore the orthology table on VEuPathDB gene pages
- Getting to OrthoMCL from VEuPathDB gene pages
- Run searches in OrthoMCL
- Explore the cluster graphs in OrthoMCL
- Leverage the phyletic pattern search
- Leverage the orthology transform tool

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

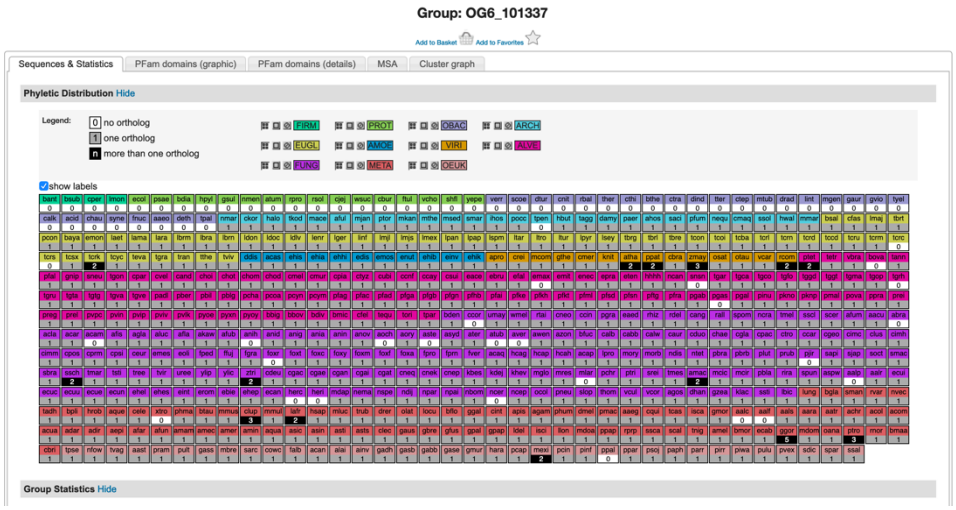
a. Go to the gene page for the *Cryptosporidium muris* gene with the ID: CMU_034340

b. What information on the gene page can you use to guess a function for this gene? It is annotated as a hypothetical protein! Hint: look at the orthologs table and the domains in the protein features graph. You may also want to visit some of the external links or take a look at InterPro domains.




c. Go to the Orthology and Synteny section and look at 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 Ortholog Group link above the table).

d. What about orthologs in organisms not in VEuPathDB? (hint: click on the Ortholog Group link above the table). Does it have any orthologs in bacteria or archaea? (Hint: mouse over the colorful boxes in the table to reveal the full species and phylum names).



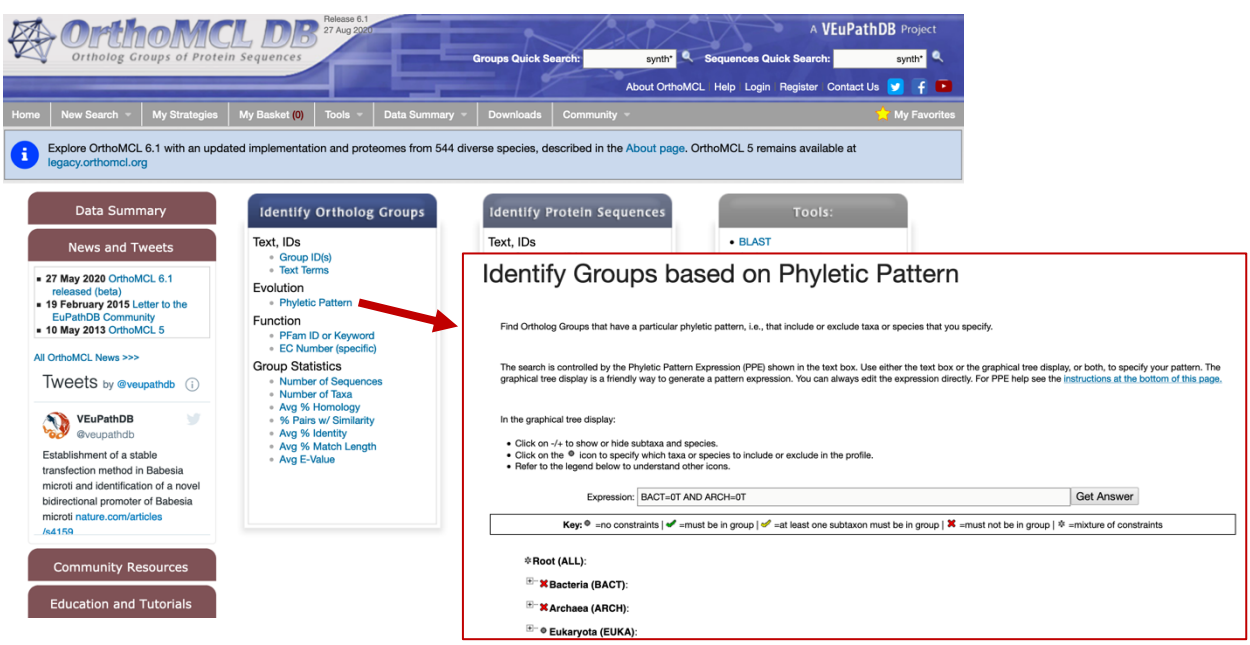

e. Look at the PFAM domain architectures found under the PFam domains (graphic) tab. Do all the proteins in this group have similar domain architecture?

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/](http://orthomcl.org/)**

How many protein groups in OrthoMCL <u>do not</u> have any orthologs in bacteria or archaea? (Hint: go to the "Phyletic Pattern" search in the Evolution section of the "Identify Ortholog Groups"

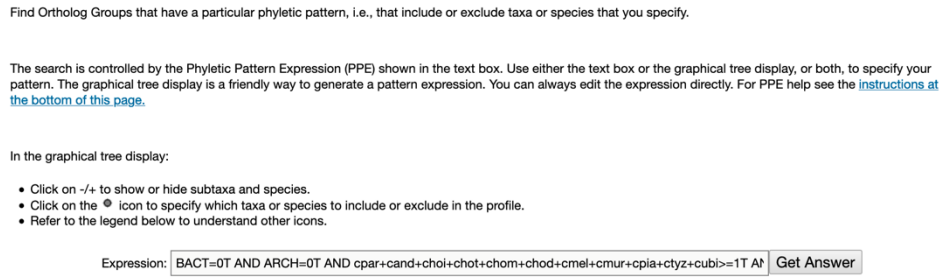


category). To specify a phyletic pattern click on the icon next to the taxonomic group or species to include or exclude it.

a. How many protein groups <u>do not</u> 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 (we will discuss why). 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 page to find additional information about expression parameters.

Before looking at the answer below, try this on your own or with the people sitting next to you.



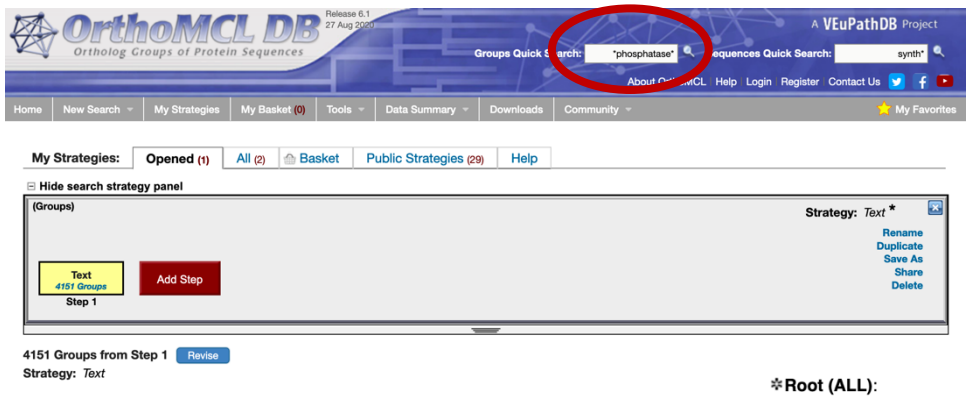

BACT=0T AND ARCH=0T AND cpar+cand+choi+chot+chom+chod+cmel+cmur+cpia+ctyz+cubi>=1T AND gass+gadh+gasb+gabb+gase+gmur>=1T

All VEuPathDB sites also have a phyletic pattern search that uses OrthoMCL data under Genes -> Evolution -> 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 but not present in the host as these genes may make good drug targets or vaccine candidates. Optional: go to your favorite VEuPathDB site and run this search to identify all genes that are not present in human or mouse.

3. **Combining searches in OrthoMCL** (Use http://orthomcl.org for this exercise).

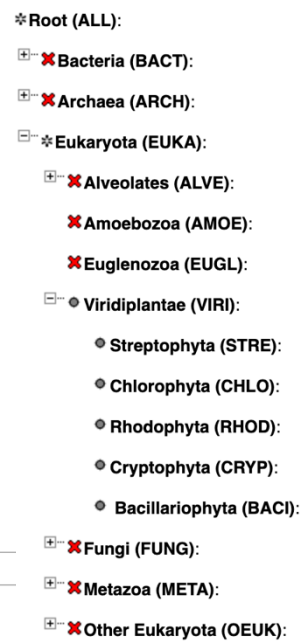
Find all plant proteins that are likely phosphatases that do not have orthologs outside of plants.

a. Use the text search **to find OrthoMCL groups** that contain the word "*phosphatase*" (note that the search should be run without the quotation marks but with the asterisks).




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).

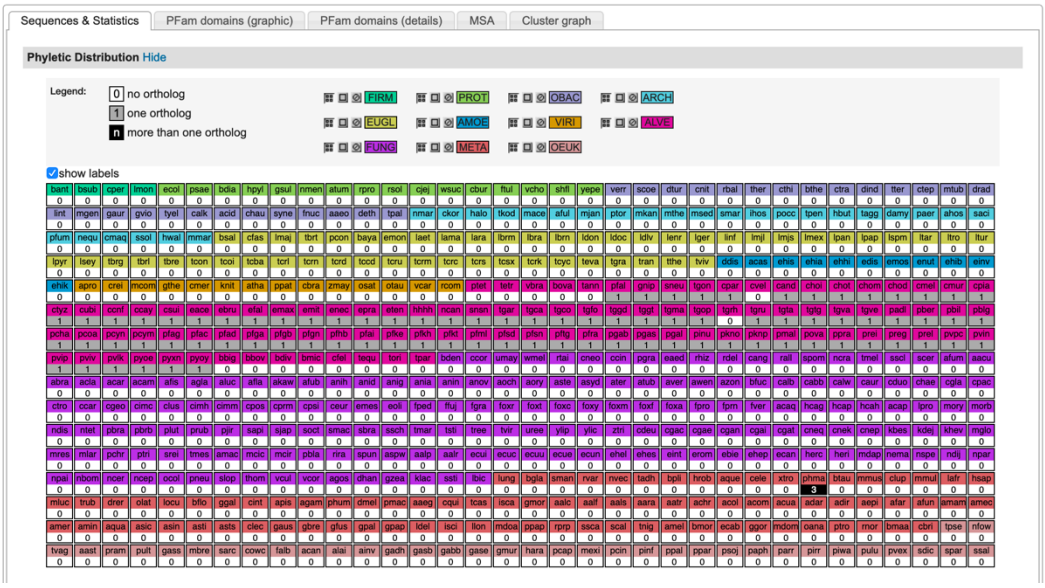
c. How many groups did you return? Explore the multiple sequence alignments from some of these groups. (Hint: click on a group ID and open the MSA tab).




4. **Exploring a specific OrthoMCL group - examining the cluster graph.** (Use http://orthomcl.org for this exercise).

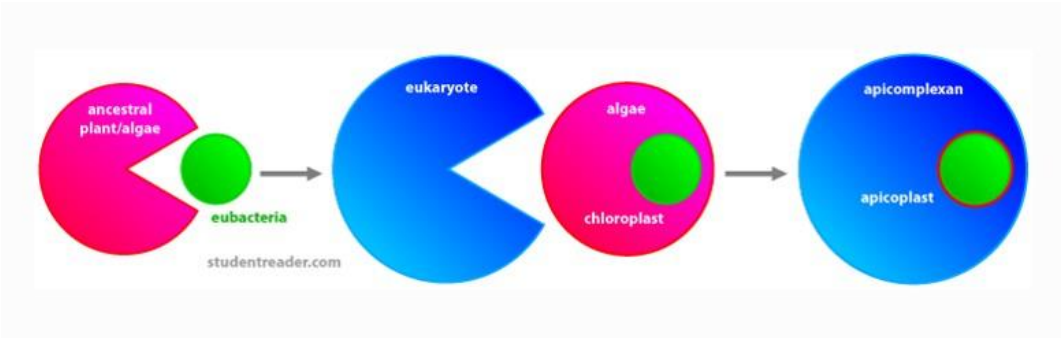
   a. Visit the orthomcl group OG6_131670. You can either type the ID in the group quick search option at the top of the page of follow this link: http://orthomcl.org/group/OG6_131670

   b. *Examine the "Sequences & Statistics" tab:* Based on the product descriptions of the members of this group, what kind of a proteins are in this group? What is the phylogenetic distribution of the members of this group?

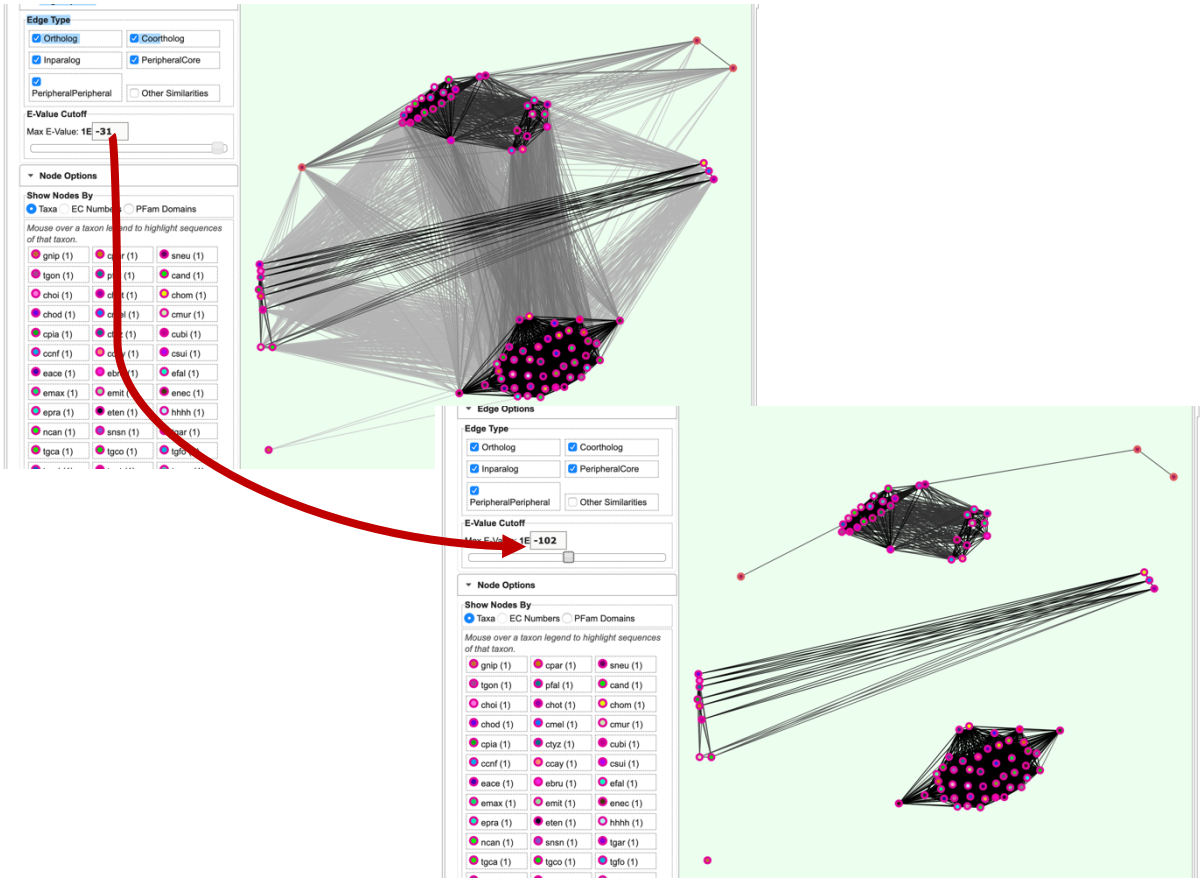


c. *Examine the "Cluster Graph" tab:* Modify the E-value cutoff slider. What happens when you increase or decrease the E-value? Can you identify subclusters?

5. **Using the orthology transform tool to identify apicoplast targeted genes in *Toxoplasma* and *Neospora*.**
   **Note: For this exercise use [http://veupathdb.org](http://veupathdb.org)**



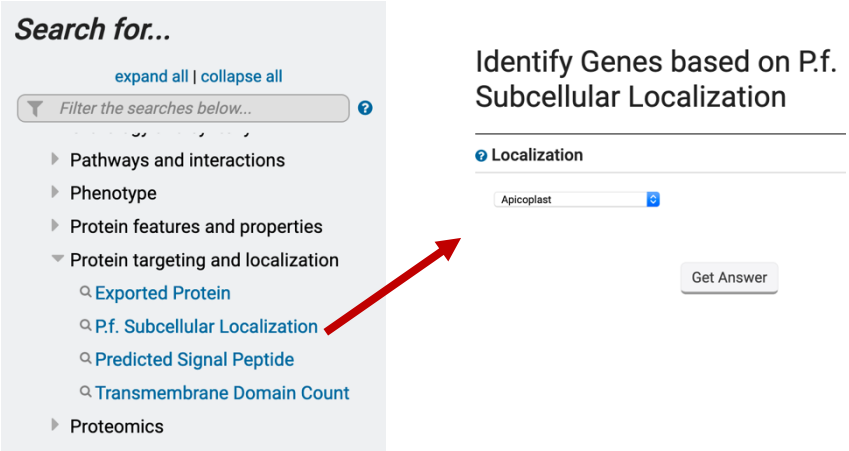

The apicoplast likely became encased in four membranes via a double endosymbiotic event.



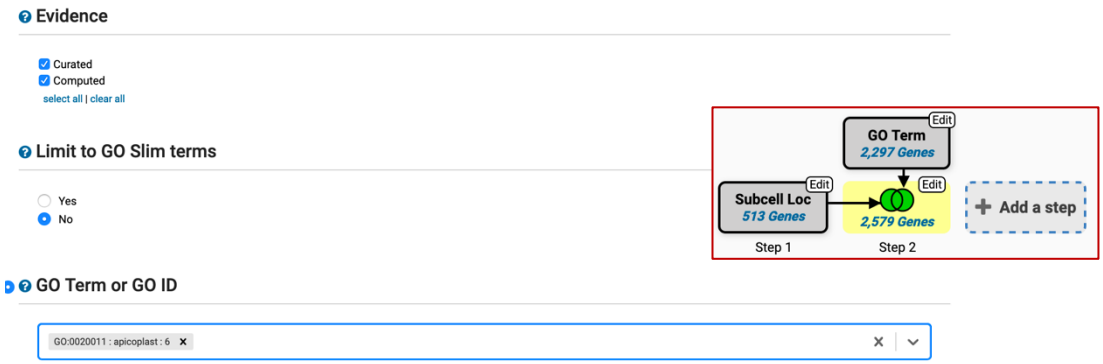

The chloroplast arose by engulfment of a cyanobacteria by a plant/algae ancestor. An algae was

then engulfed by the ancestor of all apicomplexans. Thus, an apicoplast organelle arose with four membranes.

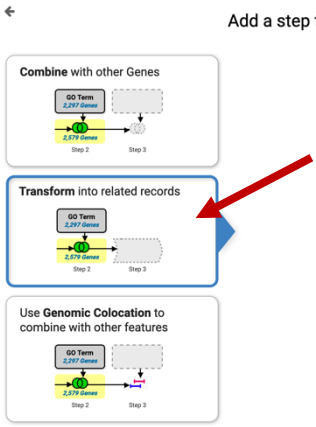
a. Start by finding genes in *Plasmodium* that are predicted to target to the apicoplast. Hint: click on “Protein targeting and localization” then on “P.f. Subcellular Localization”. You can further expand your list of potentially Apicoplast targeted proteins by running a GO terms search for the term “apicoplast” or the GO ID: GO:0020011 in *P falciparum* 3D7 (hint, click



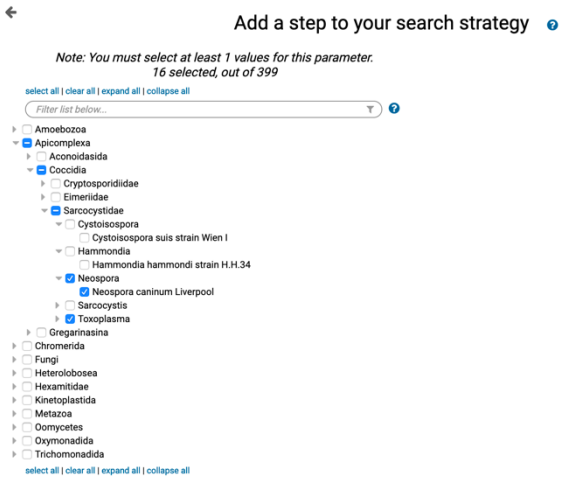

on add step the go to the function prediction category and select the GO term search). Which Boolean operation did you use? Union or intersect?




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








Although *Cryptosporidium* is an apicomplexan parasite it has 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 and use the ortholog transform back to Toxoplasma and Neospora genes for the subtraction to complete.

## Add a step to your search strategy

The results will be subtracted from the results of Step 3.

### Organism

*Note: You must select at least 1 values for this parameter.*
*11 selected, out of 399*

add these | clear these | select only these
select all | clear all

crypto

- Apicomplexa
  - Coccidia
    - Cryptosporidiidae
      - Cryptosporidium andersoni
        - Cryptosporidium andersoni isolate 30847
      - Cryptosporidium hominis
        - Cryptosporidium hominis TU502
        - Cryptosporidium hominis UdeA01
        - Cryptosporidium hominis isolate 30976
        - Cryptosporidium hominis isolate TU502_2012
      - Cryptosporidium meleagridis
        - Cryptosporidium meleagridis strain UKMEL1
      - Cryptosporidium muris
        - Cryptosporidium muris RN66
      - Cryptosporidium parvum
        - Cryptosporidium parvum IOWA-ATCC
        - Cryptosporidium parvum Iowa II
      - Cryptosporidium tyzzeri
        - Cryptosporidium tyzzeri isolate UGA55
      - Cryptosporidium ubiquitum
        - Cryptosporidium ubiquitum isolate 39726

Opened (1) All (1) Public (20) Help

*Unnamed Search Strategy **