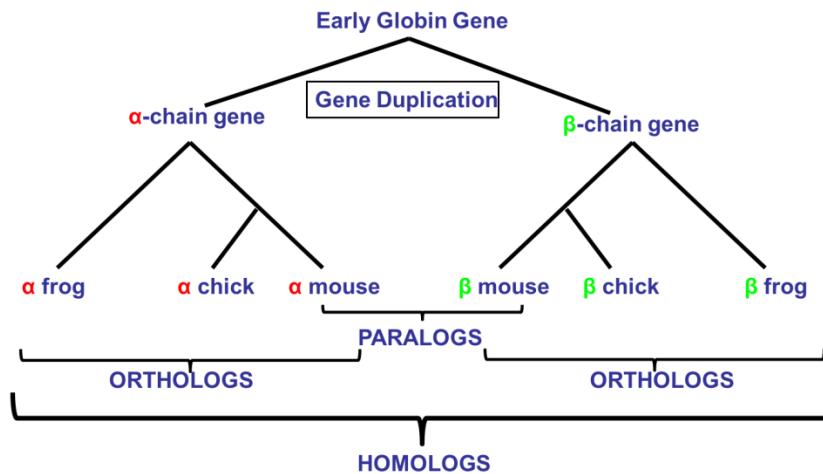


## Homology gene relationships via OrthoMCL DB

### Homology



#### **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

OrthoMCL is a genome-scale algorithm for grouping homologous protein sequences. Such homologous sequences share evolutionary history, and might also share function. Thus, homology predictions are important in predicting the function of newly identified genes. Indeed, detection of homologs has become more widespread with the rapid progress in genome sequencing and the discovery of protein sequences. Importantly, proteins in OrthoMCL groups have been shown to display a high degree of functional conservation (e.g., a group's proteins have consistent EC numbers), highlighting that OrthoMCL is useful for functional annotation of newly sequenced genomes.

OrthoMCL not only identifies groups shared by proteins from two or more species, but also groups representing species-specific gene expansion families. To achieve this, the OrthoMCL algorithm starts with reciprocal best BLAST hits within each proteome as potential in-paralog/recent paralog pairs and reciprocal best hits across any two proteomes as potential ortholog pairs. Related proteins are interlinked in a similarity graph. Then, MCL (Markov Clustering algorithm; [www.micans.org/mcl](http://www.micans.org/mcl)) is invoked to split mega-clusters. This process is analogous to the manual review in COG construction. MCL clustering is based on weights between each pair of proteins. Thus, to account for differences in evolutionary distance between any two organisms, the weights are normalized before running MCL.

The organism specific orthology information garnered from our OrthoMCL analysis of VEuPathDB organisms is presented on gene pages and integrated into an Orthology Phylogenetic Profile search. The OrthoMCL.org site offers a deep look into all data associated with the OrthoMCL results for orthology groups and proteins.

## 1. Getting to OrthoMCL from VEuPathDB databases

**Note:** For this exercise use <https://vectorbase.org> and <http://orthomcl.org>

- Use the VectorBase [Site Search](#) to visit AAEL007697 gene page for *Aedes aegypti* LVP\_AGWG.
- What information on the gene page can you use to guess a function for this gene? It is annotated as a unspecified product! 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.

AAEL007697

expand all | collapse all

Search section names...

- 1 Gene models
- 2 Annotation, curation and identifiers
- 3 Link outs**
- 4 Genomic Location
- 5 Literature
- 6 Taxonomy
- 7 Orthology and synteny
- 8 Phenotype
- 9 Transcriptomics
- 10 Sequence analysis
- 11 Sequences
- 12 Structure analysis
- 13 Protein features and properties**
- 14 Function prediction
- 15 Pathways and interactions
- 16 Immunology

expand all | collapse all

### 13 Protein features and properties

▼ Attributes and Protein Browser [Download](#) [Data sets](#)

This section contains features of each protein isoform: isoelectric point, molecular weight, sequence, transmembrane domains (TMHMM), InterPro domains, predicted export domain, PSIPRED Helix, Strand and Coil predictions.

	Transcript ID	Isoelectric Point	Molecular Weight	Has SignalP
▼	AAEL007697-RA	10.78	28290	no

[View in protein browser](#)

Reference Sequence 50 100

InterPro Domains

HAMAP SFID SUPERFAMILY TIGRFAM PIRSF PRINTS PRODOM PROSITE patterns PFAM InterProScan

- c. Go to the Orthology and Synteny section and look at the table labeled “Orthologs and Paralogs within VectorBase”. Does this gene have orthologs in other mosquitoes? What about other organisms? (hint: scan the organism column in the table)

## 7 Orthology and synteny

---

Ortholog Group [OG6\\_101337](#)

▼ Orthologs and Paralogs within VectorBase [Data sets](#)

To run Clustal Omega, select genes from the table below. Then choose the sequence type and initiate the alignment with the 'Run Clustal Omega for selected genes' button.

Search this table...

Clustal Omega	Gene	Product	Organism	Reference Strain?	Syntenic
<input type="checkbox"/>	AALC636_012600	rRNA-processing protein FCF1 homolog	Aedes albopictus C6/36 cell line	no	no
<input type="checkbox"/>	AALC636_027391	rRNA-processing protein FCF1 homolog	Aedes albopictus C6/36 cell line	no	no
<input type="checkbox"/>	AALFPA_064762	rRNA-processing protein FCF1 homolog	Aedes albopictus Foshan FPA	no	no
<input type="checkbox"/>	AALB001602	U3 small nucleolar RNA-associated protein [Source:Projected from Anopheles gambiae (AGAP009504) VB Community Annotation]	Anopheles albimanus STECLA	yes	no

- d. What about orthologs in organisms not in VEuPathDB? (hint: click on the Ortholog Group link above the table to examine the orthology information for the group at OrthoMCL.org). Does it have any orthologs in bacteria or archaea? (hint: click on Hide zero counts)

## 1 Phyletic distribution

---

▼ Phyletic Distribution of Proteins [?](#) [Download](#)

Numbers refer to the number of proteins in that organism or taxonomic group.

[expand all](#) | [collapse all](#)   
 Hide zero counts

Type a taxonomic name

▼ Eukaryota (EUKA)	633
► Alveolates (ALVE)	119
► Amoebozoa (AMOE)	14
► Euglenozoa (EUGL)	69
► Fungi (FUNG)	232
► Metazoa (META)	132
► Other Eukaryota (OEUK)	48
► Viridiplantae (VIRI)	19
▼ Archaea (ARCH)	26
Nitrosopumilus maritimus (strain SCM1) (nmar)	1
► Crenarchaeota (CREN)	13
► Euryarchaeota (EURY)	10
► Korarchaeota (KORA)	1
► Nanoarchaeota (NANO)	1

- e. Scroll down to the PFam domains section. Domain architectures are found under the PFam Architecture of Each Protein table and are described in the PFam Legend table. Do all the proteins in this group have similar domain architecture? What is the distribution of the PF04900 domain across the 659 proteins in this ortholog group? PF00149?

[Add to basket](#) [Add to favorites](#) [Download Ortholog Group](#)

## Ortholog Group: OG6\_101337

**Group Type:** Core

**Total Number of Proteins:** 659

**Keywords:** domain containing protein; pinc domain containing protein; source; uniProtKB/TrEMBL;Acc

**EC Numbers:** 3.1.-.- (2)

**Top PFam Domains:** PF04900 (630), PF01850 (3), PF00149 (1), PF00160 (1), PF00227 (1), PF05811 (1), PF13638 (1)

### 4 PFam domains

[▼ PFam Legend](#)

*Search this table...*



Accession	Symbol	Description	Count	Legend
PF04900	Fcf1	Fcf1	630	
PF01850	PIN	PIN domain	3	
PF13638	PIN_4	PIN domain	1	
PF00227	Proteasome	Proteasome subunit	1	
PF00160	Pro_isomerase	Cyclophilin type peptidyl-prolyl cis-trans isomerase/CLD	1	
PF05811	DUF842	Eukaryotic protein of unknown function (DUF842)	1	
PF00149	Metallophos	Calcineurin-like phosphoesterase	1	

[▼ PFam Architecture of Each Protein](#)

*Search this table...*



Accession	Taxon	Core/Peripheral	Protein Length	
aacu ASPACDRAFT_77294	Aspergillus aculeatus ATCC 16872	Peripheral	189	
aaeg-old AAEL007697	Aedes aegypti LVP_AGWG (old build 2019-12-20)	Core	241	
aaeg AAEL007697	Aedes aegypti LVP_AGWG	Peripheral	241	

- f. Based on the orthologs and the PFam domains shared by the group, 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/>

- How many orthology groups OrthoMCL do not have any orthologs in bacteria or archaea?  
How many protein groups do not contain orthologs from bacteria and archaea?

The screenshot shows the OrthoMCL DB website interface. At the top, there is a navigation bar with links for "My Strategies", "Searches", "Tools", "My Workspace", "Data", "About", "Help", and "Contact Us". Below the navigation bar, there is a search bar with the placeholder "Site search, e.g. OG6\_106861 or PF3D7\_1133\* or 'binding protein'" and a search icon. On the left side, there is a sidebar titled "Search for..." with a "collapse all" link and a "Filter the searches below..." button. The sidebar contains a section titled "Ortholog Groups" with the following options: "% Pairs w/ Similarity", "All Groups", "Avg % Homology", "Avg % Identity", "Avg % Match Length", "Avg E-Value", "EC Number", "Group ID(s)", "Group or Sequence ID", "Number of Sequences", "Number of Taxa", "PFam ID or Keyword", "Phyletic Pattern" (which has a pink arrow pointing to it), and "Text Terms". In the main content area, there is a large box titled "Identify Ortholog Groups based on Phyletic Pattern". Inside this box, there is a key: "Key: ● = no constraints | ✓ = must be in group | ⚡ = at least one subtaxon must be in group". Below the key, there is a search bar with the placeholder "Type a taxonomic name" and a search icon. A list of taxonomic names is shown, each preceded by a small icon indicating constraints: 

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

Ortholog Group Results							
	Ortholog Group	Total Number Proteins	Keywords	Top PFam Domains	EC Numbers	Archaea	Bacteria
1	OG6_100001	14736	unknown; hypothetical protein; conserved hypothetical protein	PF13388 (4233), PF04665 (3687), PF04851 (212)	N/A	0 / 27 (0%)	0 / 47 (0%)
2	OG6_100002	6864	unknown; conserved hypothetical protein	PF12943 (5254), PF10544 (1424), PF04383 (2), PF12789 (2)	N/A	0 / 27 (0%)	0 / 47 (0%)
3	OG6_100003	6578	hypothetical protein; conserved hypothetical protein; unknown	PF12789 (2592), PF06022 (45), PF02349 (1), PF03770 (1), PF07679 (1), PF12295 (1)	1.4.1.2 (2)	0 / 27 (0%)	0 / 47 (0%)

- b. Find all groups that contain orthologs from at least one species of Arthropoda but not from bacteria or archaea.

Before looking at the answer below, try this on your own or with your neighbour classmate.

Expression: ARTH>=1T AND ARCH=0T AND BACT=0T Get Answer

**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)
- \* Eukaryota (EUKA)
  - ▶ ● Alveolates (ALVE)
  - ▶ ● Amoebozoa (AMOE)
  - ▶ ● Euglenozoa (EUGL)
  - ▶ ● Fungi (FUNG)
  - \* Metazoa (META)
    - ▶ ✅ Arthropoda (ARTH)
    - ▶ ● Chordata (CHOR)
    - ▶ ● Nematodes (NEMA)
    - ▶ ● Other Metazoa (OMET)
    - ▶ ● Other Eukaryota (OEUK)
    - ▶ ● Viridiplantae (VIRI)
- ✗ Archaea (ARCH)
  - ▶ ✗ Nitrosopumilus maritimus (strain SCM1) (nmar)
  - ▶ ✗ Crenarchaeota (CREN)
  - ▶ ✗ Euryarchaeota (EURY)
  - ▶ ✗ Korarchaeota (KORA)
  - ▶ ✗ Nanoarchaeota (NANO)
- ✗ Bacteria (BACT)
  - ▶ ✗ Firmicutes (FIRM)
  - ▶ ✗ Other Bacteria (OBAC)
  - ▶ ✗ Proteobacteria (PROT)

↑

Before you click on Get Answer, scroll down to the bottom of the page to find additional information about expression parameters.

The expression parameter of the query we just did is:

Expression ARTH>=1T AND ARCH=0T AND BACT=0T

If just with clicks you can not construct a query, try typing a expression.

- c. All VEuPathDB sites also have a phyletic pattern search that uses OrthoMCL data under Genes -> Orthology and synteny -> Orthology Phylogenetic Profile.

**VectorBase**  
Bioinformatics Resources for  
Invertebrate Vectors of Human Pathogens

Release 57  
21 Apr 2022

**Search for...**

expand all | collapse all

Filter the searches below...

▼ Genes

- ▶ Annotation, curation and identifiers
- ▶ Function prediction
- ▶ Gene models
- ▶ Genomic Location
- ▶ Immunology
- ▼ Orthology and synteny
- ➡ Orthology Phylogenetic Profile
- 🔍 Paralog Count

**PlasmoDB**  
Plasmodium Informatics Resources

Release 57  
21 Apr 2022

**Search for...**

expand all | collapse all

Filter the searches below...

▼ Genes

- ▶ Annotation, curation and identifiers
- ▶ Epigenomics
- ▶ Function prediction
- ▶ Gene models
- ▶ Genetic variation
- ▶ Genomic Location
- ▶ Immunology
- ▼ Orthology and synteny
- ➡ Orthology Phylogenetic Profile
- 🔍 Paralog Count

This search is very useful to identify genes in your organism of interest that are restricted in their profile. For example, if you are working with a parasite species, you might want to identify genes that are conserved among organisms in your genus (e.g., *Plasmodium*) but not present in the host (e.g., human) 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.

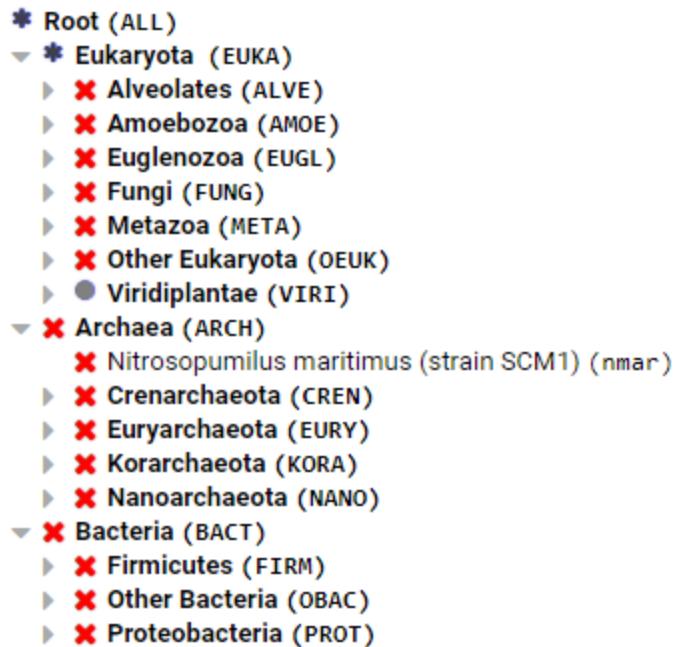
### 3. Combining searches in OrthoMCL (Use <http://orthomcl.org> for this exercise).

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

- Use the text search **to find OrthoMCL groups** that contain the word \*phosphatase\* (note that the search should be run with the asterisks).

The screenshot shows the OrthoMCL search interface. On the left, there is a sidebar with various search filters like 'Search for...', 'Ortholog Groups', and 'Text Terms'. A pink arrow points from the 'Text Terms' link to the main search area. The main area has a title 'Identify Ortholog Groups based on Text Terms' and a text input field containing '\*phosphatase\*'. Below it, there's a section for 'Fields' with many checkboxes checked, including 'EC Number', 'Keywords', 'Ortholog group', etc. A pink circle highlights the 'Get Answer' button. To the right, a box labeled 'Step 1' contains the text 'Text 4,981 Ortholog Groups'. A pink arrow points from the 'Get Answer' button towards this box.

- Add a step and run a phylogenetic 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. Examine your results. How many groups were returned by the search? What is the distribution of plant proteins in each orthology group?

Ortholog Group	Total Number Proteins	Keywords	Top PFam Domains	EC Numbers	Viridiplantae	Archaea	Alveolata
OG6_134309	58	containing protein; domain containing protein; leucine rich repeat containing protein; nb-arc dom...	PF00931 (47), PF13855 (9), PF07985 (1)	3.1.3.16 (31)	5 / 14 (36%)	0 / 27 (0%)	0 / 125 (0%)
OG6_108065	37	ppm-type phosphatase domain containing protein; uncharacterized protein	PF00481 (25), PF00227 (5)	N/A	1 / 14 (7%)	0 / 27 (0%)	0 / 125 (0%)
OG6_112109	26	phosphatase; ppm-type phosphatase domain containing protein	PF00481 (26), PF02148 (1), PF07576 (1), PF13639 (1)	3.1.3.16 (6)	10 / 14 (71%)	0 / 27 (0%)	0 / 125 (0%)
OG6_112423	24	ippc domain containing protein	PF03372 (22)	3.1.3.36 (2), 3.1.3.56 (2), 3.1.3.86 (2), 3.1.3.- (1)	7 / 14 (50%)	0 / 27 (0%)	0 / 125 (0%)

- d. Run a multiple sequence alignment for OG6\_112109. Click on the group ID in your result table and navigate to the List of Proteins section of the group page. The Clustal Omega tool is integrated into the table. There are several formats available for the Clustal output, making it easy to take these results to other visualization programs.

**OrthoMCL DB** Ortholog Groups of Protein Sequences Release 6.0 21 Apr 2022 Site search, e.g. OG6\_106861 or PF3D7\_1133\* or "binding protein" My Strategies Searches Tools My Workspace Data About Help Contact Us

**OG6\_112109** < expand all | collapse all Search section names... ▾

1 Phyletic distribution 2 Group summary 3 List of proteins 4 Pfam domains 5 Cluster graph expand all | collapse all

**3 List of proteins**

▼ List of All Proteins [Download](#) To align sequences, select proteins from the table below. Then choose the 'Output format' and click the 'Run Clustal Omega for selected' button.

Search this table...

Clustal Omega	Accession	Description	Organism	Taxon
<input checked="" type="checkbox"/>	vcar D8UBL1	PPM-type phosphatase domain-containing protein	Volvox carteri f. nagariensis	Viridiplantae
<input checked="" type="checkbox"/>	crei A0A2K3DZC7	PPM-type phosphatase domain-containing protein	Chlamydomonas reinhardtii (Chlamydomonas smithii)	Viridiplantae
<input checked="" type="checkbox"/>	vcar D8TYP9	Uncharacterized protein	Volvox carteri f. nagariensis	Viridiplantae
<input checked="" type="checkbox"/>	apro A0A087SRW5	PPM-type phosphatase domain-containing protein	Auxenochlorella protothecoides (Green microalgae) (Chlorella protothecoides)	Viridiplantae
<input checked="" type="checkbox"/>	cbra A0A388JMB4	PPM-type phosphatase domain-containing protein	Chara braunii (Braun's stonewort)	Viridiplantae
<input checked="" type="checkbox"/>	apro A0A087SJZ6	PPM-type phosphatase domain-containing protein	Auxenochlorella protothecoides (Green microalgae) (Chlorella protothecoides)	Viridiplantae
<input checked="" type="checkbox"/>	crei A0A2K3DBF3	PPM-type phosphatase domain-containing protein	Chlamydomonas reinhardtii (Chlamydomonas smithii)	Viridiplantae
<input checked="" type="checkbox"/>	osat Q0JMD4	Probable protein phosphatase 2C 3	Oryza sativa subsp. japonica (Rice)	Viridiplantae
<input checked="" type="checkbox"/>	zmay A0A1D6PCB8	PPM-type phosphatase domain-containing protein	Zea mays (Maize)	Viridiplantae
<input checked="" type="checkbox"/>	ppat A0A2K1L7H1	PPM-type phosphatase domain-containing protein	Physcomitrium patens (Spreading-leaved earth moss) (Physcomitrella patens)	Viridiplantae
<input checked="" type="checkbox"/>	knit A0A1Y1LB8	PPM-type phosphatase domain-containing protein	Klebsormidium nitens (Green alga) (Ulothrix nitens)	Viridiplantae
<input checked="" type="checkbox"/>	zmay A0A1D6MTG2	PPM-type phosphatase domain-containing protein	Zea mays (Maize)	Viridiplantae
<input checked="" type="checkbox"/>	ppat A0A2K1K2S8	PPM-type phosphatase domain-containing protein	Physcomitrium patens (Spreading-leaved earth moss) (Physcomitrella patens)	Viridiplantae
<input checked="" type="checkbox"/>	ppat A0A2K1JUE3	PPM-type phosphatase	Physcomitrium patens (Spreading-	Viridiplantae
<a href="#">Check All</a> <a href="#">Uncheck All</a>				

Please note: selecting a large number of proteins will take several minutes to align.

Output format: [Mismatched highlighted](#) ↗

[Run Clustal Omega for selected proteins](#)

#### 4. Explore a specific OrthoMCL group - examining the cluster graph. Use <http://orthomcl.org>

- Visit the Orthomcl group OG6\_131670. Type the ID into the site search to navigate to it.
- Examine the Phyletic Distribution. What is the phylogenetic distribution of the members of this group? The distribution is presented as a tree. Expand the tree to view the distribution.

## 1 Phyletic distribution

### ▼ Phyletic Distribution of Proteins

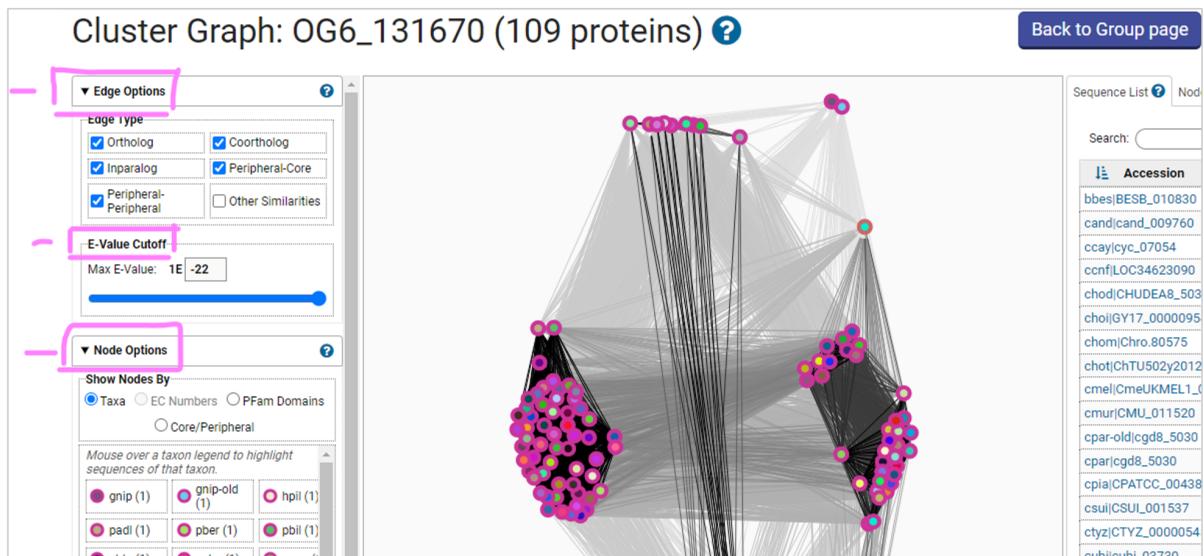
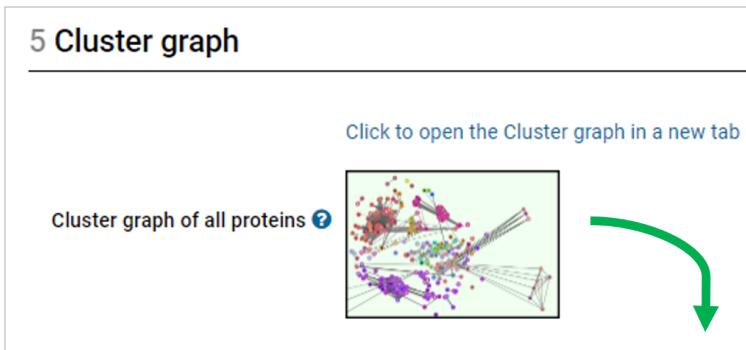
Numbers refer to the number of proteins in that organism or taxonomic group.

 expand all
 collapse all
 Hide zero counts

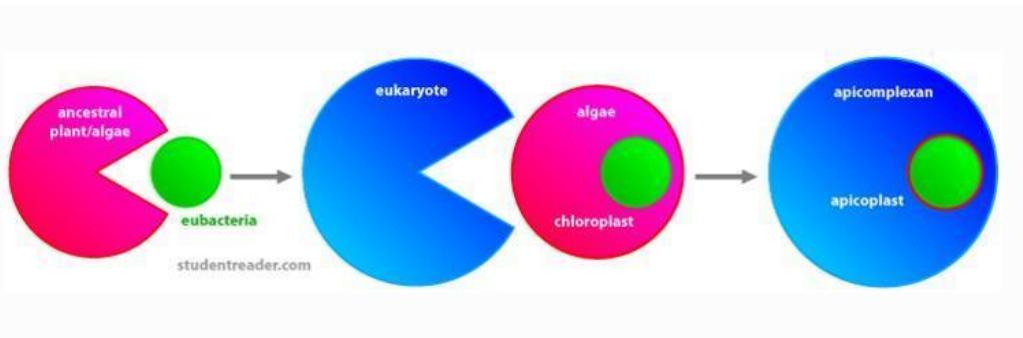
 

 <b>Eukaryota (EUKA)</b>	109
 <b>Alveolates (ALVE)</b>	106
 <b>Metazoa (META)</b>	3

- c. Navigate to the Cluster graph tab. Modify the E-value cutoff slider. What happens when you increase or decrease the E-value? Can you identify subclusters of orthologs? The view of the graph can be changed using the Edge type options and the Node options.



5. Using the orthology transform tool to identify apicoplast targeted genes in *Toxoplasma* and *Neospora*. Note: For this exercise use <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: Navigate to the [P.f. 3D7 Subcellular Localization](#) search.



**VEuPathDB**  
Eukaryotic Pathogen, Vector & Host Informatics Resource

Release 57  
21 Apr 2022

**Search for...**

X
?

**Genes**

- Protein targeting and localization
- Exported Protein
- Pf3D7 Subcellular Localization** →
- Predicted Signal Peptide
- Transmembrane Domain Count

### Identify Genes based on Pf3D7 Subcellular Localization

**Localization**

Apicoplast

**Get Answer**

Subcell Loc  
513 Genes

+ **Add a step**

Step 1

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?

Add a step to your search strategy

3D7 [×](#) [?](#)

Apicomplexa  
 Aconoidasida  
 Haemosporida  
 Plasmodium  
 Plasmodium falciparum  
 Plasmodium falciparum 3D7 [Reference]

[select only these](#) | [add these](#) | [clear these](#)

---

[?](#) Evidence

Curated  
 Computed  
[select all](#) | [clear all](#)

---

[?](#) Limit to GO Slim terms

Yes  
 No

```

graph TD
    A[Subcell Loc  
513 Genes] --> B(( ))
    B --> C[GO Term  
370 Genes]
    C --> D[649 Genes]
    
```

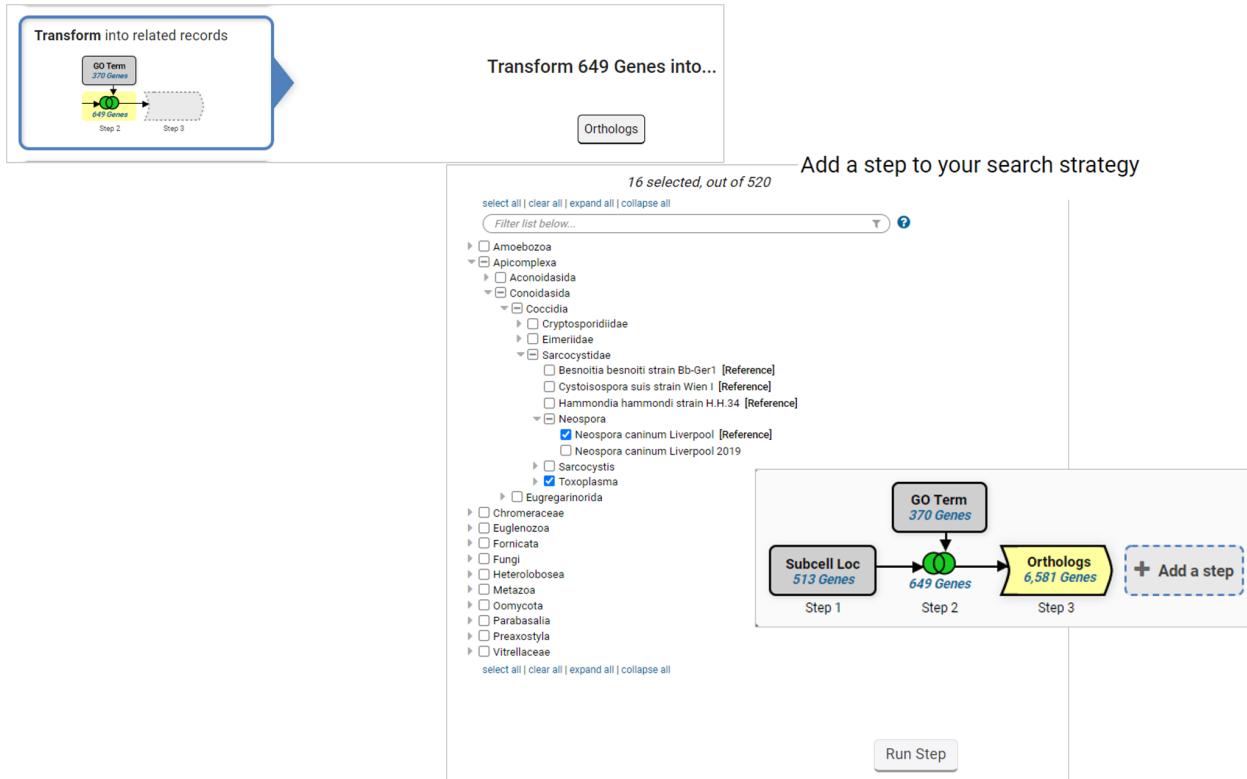
The flowchart illustrates a two-step gene selection process. Step 1, labeled "Subcell Loc 513 Genes", leads to Step 2, which is a green circle containing a green dot, labeled "649 Genes". Above Step 2 is a box labeled "GO Term 370 Genes". A dashed blue box labeled "+ Add a step" is located to the right of Step 2.

---

[?](#) GO Term or GO ID

GO:0020011 : apicoplast : 7 [×](#)

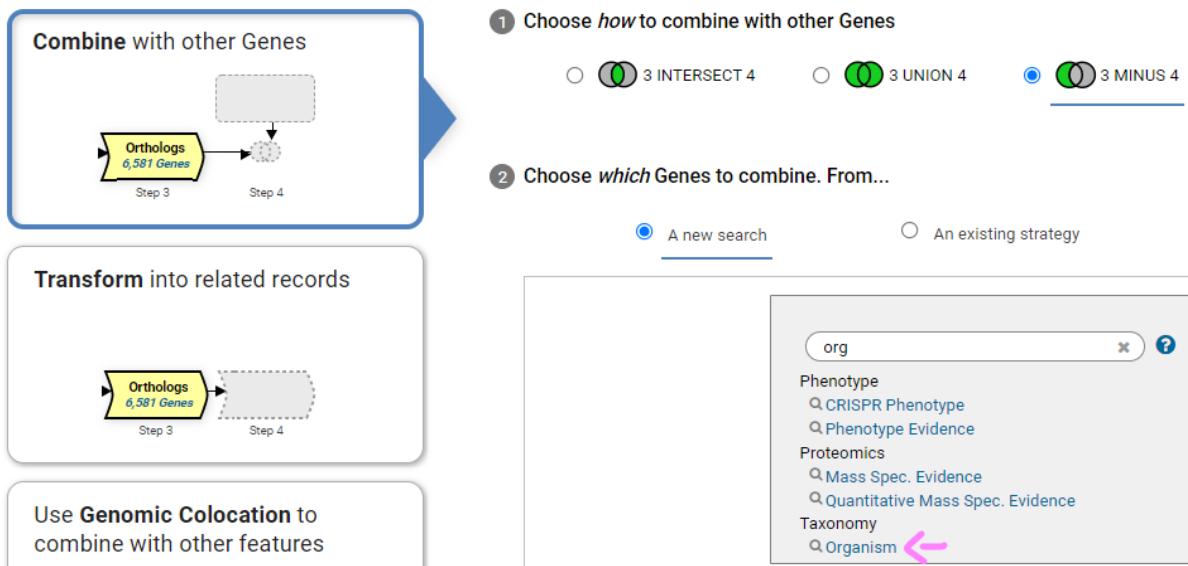
- b. Transform the results into their *Toxoplasma* and *Neospora* orthologs. Add a step to your strategy that transforms the results into *Toxoplasma* and *Neospora*.



- c. 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 ?



[←](#) Add a step to your search strategy [?](#)

Search for genes by organism

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

### Organism

11 selected, out of 520

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

Filter list below... [?](#)

- Amoebozoa
- Apicomplexa
  - Aconoidasida
  - Conoidasida
    - Coccidia
      - Cryptosporidiidae
      - Eimeriidae
      - Sarcocystidae
    - Eugregarinorida
  - Chromeraceae
  - Euglenozoa
  - Fornicata
  - Fungi
  - Heterolobosea
  - Metazoa
  - Oomycota
  - Parabasalia
  - Preaxostyla
  - Vitellaceae

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

Run Step 

[View](#) | [Analyze](#) | [Revise](#) **Make nested strategy** [Insert step before](#) | [Orthologs](#) | [Delete](#)

## My Search Strategies

Opened (1) All (2) Public (12) Help

Unnamed Search Strategy \* 



**Details for step Organism**   
43549 Genes

**Organism** Cryptosporidium andersoni isolate 30847, Cryptosporidium hominis TU502, Cryptosporidium hUdeA01, Cryptosporidium hominis isolate 30976, Cryptosporidium hominis isolate TU502\_2012, Cryptosporidium meleagridis strain UKMEL1, Cryptosporidium muris RN66, Cryptosporidium parvum ATCC, Cryptosporidium parvum Iowa II, Cryptosporidium tyzzeri isolate UGA55, Cryptosporidium isolate 39726



▶ Give this search a weight



Step 1: Subcell Loc 513 Genes → Step 2: GO Term 370 Genes (649 Genes) → Step 3: Orthologs 6,581 Genes (6,581 Genes) → Step 4: Organism 43,549 Genes

## Your Genes from Step 1 will be converted into Orthologs

### Organism

16 selected, out of 520

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

Filter list below... ?

- ▶  Amoebozoa
- ▼  Apicomplexa
  - ▶  Aconoidasida
  - ▼  Conoidasida
    - ▼  Coccidia
      - ▶  Cryptosporidiidae
      - ▶  Eimeriidae
      - ▼  Sarcocystidae
        - Besnoitia besnoiti strain Bb-Ger1 [Reference]
        - Cystoisospora suis strain Wien I [Reference]
        - Hammondia hammondi strain H.H.34 [Reference]
    - ▼  Neospora
      - Neospora caninum Liverpool [Reference]
      - Neospora caninum Liverpool 2019
    - ▶  Sarcocystis
    - ▶  Toxoplasma
    - ▶  Eugregarinorida
  - ▶  Chromeraceae
  - ▶  Euglenozoa
  - ▶  Fornicata
  - ▶  Fungi
  - ▶  Heterolobosea
  - ▶  Metazoa
  - ▶  Oomycota
  - ▶  Parabasalia
  - ▶  Preaxostyla
  - ▶  Vitellaceae

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

