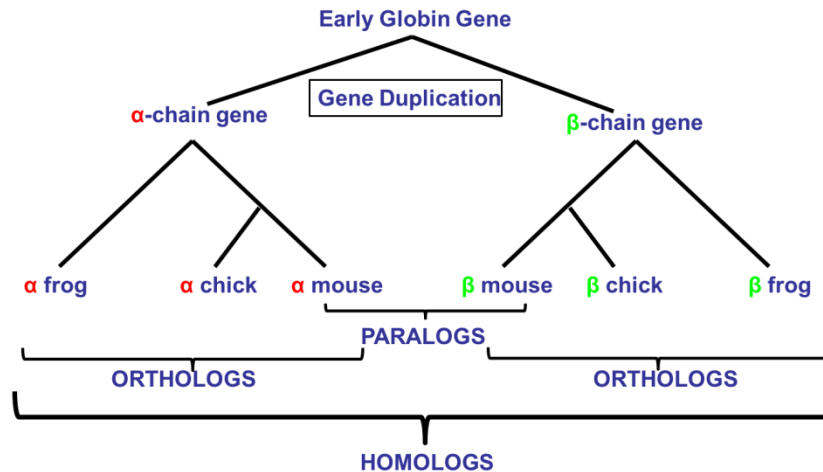


Orthology and Phyletic Patterns

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 orthologous protein sequences which not only share evolutionary history, but also share function. Thus, ortholog prediction is important in predicting the function of newly identified proteins. Detection of orthologs has become more widespread with the rapid progress in genome sequencing and the discovery of protein sequences (Glover et al. 2019). 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) (Li et al. 2003), 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; Dongen 2000; 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 cryptodb.org and orthomcl.org

- Use the CryptoDB site search to visit the gene page for the *Cryptosporidium muris* gene, CMU_034340, hypothetical protein, conserved.
- 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.

CMU_034340

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 Genetic variation

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

Proteins Properties and Features

Download

Data sets

Transcript ID	Isoelectric Point	Molecular Weight	Has SignalP	Has TMHMM	Protein Length	Pro Bro
CMU_034340-t26_1	10.5	23784	no	no	206	Interac

View in protein browser

Reference Sequence

InterPro Domains

Transmembrane Domains (TMHMM)

Signal Peptide

- 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: scan the organism column in the table) The orthologs may have defined functions (apparent from the product description) which can be used to infer the same function on CMU_034340.

7 Orthology and syteny

Ortholog Group: **OG6_101337**

▼ Orthologs and Paralogs within CryptoDB [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	Organism	Product	is syntenic	has comments
<input type="checkbox"/>	Cvel_467	Chromera velia CCMP2878	rRNA-processing protein FCF1 homolog, putative	no	no
<input type="checkbox"/>	cand_030400	Cryptosporidium andersoni isolate 30847	hypothetical protein	yes	no
<input type="checkbox"/>	Chro.70261	Cryptosporidium hominis TU502	hypothetical protein	yes	no
<input type="checkbox"/>	CHUDEA7_2290	Cryptosporidium hominis UdeA01	unspecified product	yes	no
<input type="checkbox"/>	GY17_00002025	Cryptosporidium hominis isolate 30976	rRNA-processing protein Fcf1/Utp23	yes	no
<input type="checkbox"/>	ChTU502y2012_407q1140	Cryptosporidium	Fcf1	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?

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





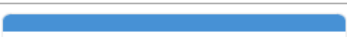


▼ Eukaryota (EUKA)	734
▶ Alveolates (ALVE)	130
▶ Amoebozoa (AMOE)	14
▶ Euglenozoa (EUGL)	73
▶ Fungi (FUNG)	299
▶ Metazoa (META)	143
▶ Other Eukaryota (OEUK)	56
▶ Viridiplantae (VIRI)	19
▼ Archaea (ARCH)	26
Nitrosopumilus maritimus (strain SCM1) (nmar)	1
▶ Crenarchaeota (CREN)	13
▶ Euryarchaeota (EURY)	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 760 proteins in this ortholog group? PF00149?

4 PFam domains





▼ PFam Legend [Download](#)

Search this table... ?

↕ Accession	↕ Symbol	↕ Description	↕ Count ?	Legend
PF04900	Fcf1	Fcf1	731	
PF05811	DUF842	Eukaryotic protein of unknown function (DUF842)	5	
PF01850	PIN	PIN domain	3	
PF00227	Proteasome	Proteasome subunit	1	
PF00160	Pro_isomerase	Cyclophilin type peptidyl-prolyl cis-trans isomerase/CLD	1	
PF13638	PIN_4	PIN domain	1	
PF00149	Metallophos	Calcineurin-like phosphoesterase	1	

▼ PFam Architecture of Each Protein [Download](#)

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	
aalb AALFPA_064762	Aedes albopictus Foshan FPA	Peripheral	204	

- 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/>

- a. How many orthology groups in OrthoMCL do not have any orthologs in bacteria or archaea?
(Use the Groups by Phyletic Pattern search)

The screenshot shows the OrthoMCL website interface. On the left, the 'Search for...' section has a list of search criteria. A blue arrow points from 'Phyletic Pattern' in this list to the 'Identify Ortholog Groups based on Phyletic Pattern' tool window on the right.

Identify Ortholog Groups based on Phyletic Pattern

In the graphical tree display:

- Click on the icons to show or hide subtaxa and species.
- Click on the icons to specify which taxa or species to include or exclude in the profile.
- Refer to the legend below to understand other icons.

Expression:

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

expand all | collapse all

Type a taxonomic name

- 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)

Get Answer

The screenshot shows the OrthoMCL results page for a 'Phyletic' search. At the top, it says 'Phyletic 890,737 Ortholog Groups' with a 'Step 1' indicator and an 'Add a step' button. Below this, there's a 'Revise this search' button and a section titled 'Ortholog Group Results'.

Ortholog Group Results

1 2 3 ... 44,537 Rows per page: 20

Download Add to Basket Add Columns

Ortholog Group	Total Number Proteins	Keywords
OG6_100001	19270	unknown; hypothetical protein; conserved hypothetical protein; duf4106
OG6_100002	9386	unknown; conserved hypothetical protein; organellar; viral DNA polymerase type B

b. How many protein sequences do not contain orthologs from bacteria and archaea?

The screenshot shows the OrthoMCL web interface. The top panel, titled "Add a step to your search strategy", displays two options for Step 2: "Combine with other Ortholog Groups" and "Transform into related records". The "Transform into related records" option is selected, showing a diagram where "Phyletic 890,737 Ortholog Groups" (Step 1) is transformed into "Proteins" (Step 2). A blue arrow points from the "Proteins" box to the bottom panel.

The bottom panel, also titled "Add a step to your search strategy", shows the message "Your Ortholog Groups from Step 1 will be converted into Proteins". It includes links for "Configure Search" and "Learn More". A "Run Step" button is circled in blue.

Below the panels, a progress bar shows the workflow: "Phyletic 890,737 Ortholog Groups" (Step 1) leads to "To Proteins 6,277,497 Proteins" (Step 2), followed by a dashed box with a plus sign and the text "Add a step".

c. 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 in the Groups by Phyletic Pattern search. However, OrthoMCL has an added feature that allows you to enter an expression to define the phyletic pattern. This option provides additional flexibility. For example, the phyletic patterns search we just ran can be expressed as ARCH=OT AND BACT=OT. Can you figure out what expression to use to answer this question? (hint: scroll down to the bottom of the Phyletic Pattern search page to find additional information about expression parameters.

Identify Ortholog Groups based on Phyletic Pattern

In the graphical tree display:

- Click on the icons to show or hide subtaxa and species.
- Click on the icons to specify which taxa or species to include or exclude in the profile.
- Refer to the legend below to understand other icons.

Expression:

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

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

Type a taxonomic name

- * 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)

[Get Answer](#)

Build a Web Services URL from this Search >>

Give this search a name (optional)
 Give this search a weight (optional)

Description

Find Ortholog Groups by phyletic pattern.

Phyletic Pattern Expression is a flexible and powerful way to identify ortholog groups with a certain conservation pattern. The pattern is used to identify groups with a certain copy number (e.g., duplications) within specified taxa.

Examples

These expressions find ortholog groups in which...

Before looking at the answer below, try this on your own or with the people in your breakout room.

Identify Ortholog Groups based on Phyletic Pattern

Find Ortholog Groups that have a particular phyletic pattern, i.e., that include or exclude taxa or species that you specify.

The search is controlled by the Phyletic Pattern Expression (PPE) shown in the text box. Use either the text box or the graphical tree display, or both, to specify your pattern. The graphical tree display is a friendly way to generate expression. You can always edit the expression directly. For PPE help see the instructions at the bottom of this page.

In the graphical tree display:

- Click on the icons to show or hide subtaxa and species.
- Click on the icons to specify which taxa or species to include or exclude in the profile.
- Refer to the legend below to understand other icons.

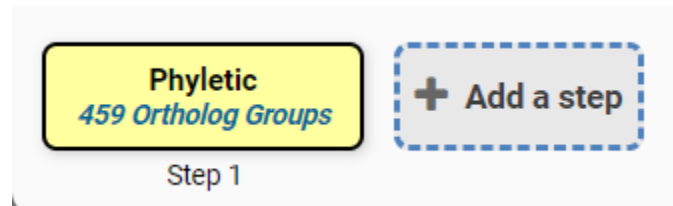
Expression:

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

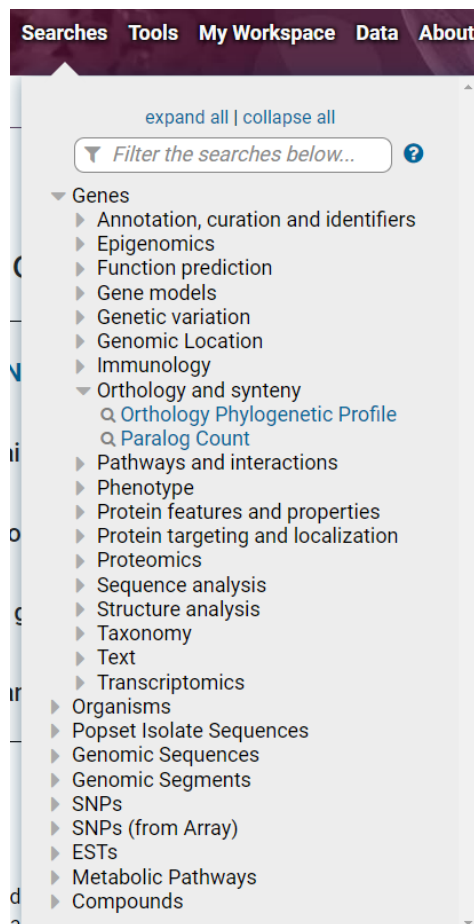
- * Root (ALL)
 - * Eukaryota (EUKA)
 - * Other Eukaryota (OEUK)
 - ✓ Giardia Assemblage A isolate WB (gass)
 - ✓ Giardia Assemblage A isolate WB (old build 2013-02-08) (gass-olc)
 - ✓ Giardia Assemblage A2 isolate DH (gadn)
 - ✓ Giardia Assemblage B isolate GS (gabn)
 - ✓ Giardia Assemblage B isolate GS_B (gabn)
 - ✓ Giardia Assemblage E isolate P15 (gase)
 - ✓ Giardia muris strain Roberts-Thomson (gmur)

[Get Answer](#)

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



- d. All VEuPathDB sites also have a phyletic pattern search that uses OrthoMCL data (PlasmoDB, Genes -> Orthology and synteny -> Orthology Phylogenetic Profile). This search is very useful to identify genes in your organism of interest that have a restricted orthology profile. For example, genes may make good drug targets or vaccine candidate may be conserved among organisms in your genus but not present in the host as these s. Optional: go to your favorite VEuPathDB site and run this search to identify all genes that are not present in human or mouse.



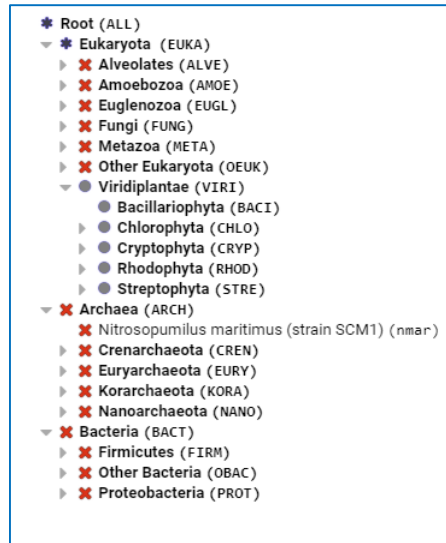
3. Combining searches in OrthoMCL (Use <http://orthomcl.org> for this exercise).

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).

The image shows a screenshot of the OrthoMCL web interface. On the left, a sidebar menu is open, showing a list of search criteria under the 'Proteins' category. A blue arrow points from the 'Text Terms' option in this menu to the main search area. The main search area is titled 'Identify Ortholog Groups based on Text Terms'. It has a text input field containing '*phosphatase*'. Below the input field is a 'Fields' section with several checkboxes: EC Number, Keywords, Ortholog group, Pfam Domains, Protein Description, Protein ID, Protein Previous Groups, Protein Taxon Abbreviation, and Protein Taxon Name. A 'Get Answer' button is located at the bottom of the search area. A blue arrow points from the 'Get Answer' button to a yellow box on the right that says 'Text 5,377 Ortholog Groups'. Below this box is the label 'Step 1'.

- 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. Examine your results. How many groups were returned by the search? What is the distribution of plant proteins in each orthology group? (use the Add Columns tool to turn on the Viridiplantae column if it is not already on)

Unnamed Search Strategy *



399 Ortholog Groups

Ortholog Group Results

Ortholog Group	Total Number Proteins	Viridiplantae	Top PFam Domains	Keywords
OG6_112109	26	10 / 14 (71%)	PF00481 (26), PF02148 (1), PF07576 (1), PF13639 (1)	phosphatase; ppm-type phosphatase domain containing protei
OG6_112423	24	7 / 14 (50%)	PF03372 (22)	ippc domain containing protein
OG6_130528	21	4 / 14 (29%)	PF00931 (15), PF13855 (1)	disease resistance; disease resistance protein; containing prot
OG6_153369	20	3 / 14 (21%)	PF13855 (14), PF00560 (13), PF08263 (13)	lrrnt_2 domain containing protein; plant-type; serine-threonine p
OG6_114630	18	1 / 14 (7%)	PF00481 (11)	ppm-type phosphatase domain containing protein; uncharacter

- 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

Site search: e.g. OG6_108961 or PF3027_11321 or binding protein

My Strategies | Searches | Tools | My Workspace | Data | About | Help | Contact Us

OG6_112109

expand all | collapse all

Search across reports

- 1 Phylogenetic distribution
- 2 Group summary
- 3 List of proteins
- 4 Pfam domains
- 5 Cluster graph

expand all | collapse all

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 genes' button.

Search this table

Clustal Omega	Accession	Description	Organism	Taxon	Core/Peripheral	Length
<input checked="" type="checkbox"/>	vcariD8UBL1	PPM-type phosphatase domain-containing protein	Volvox carteri f. nagariensis	Viridiplantae	Peripheral	1309
<input checked="" type="checkbox"/>	creiA0A2K3DZC7	PPM-type phosphatase domain-containing protein	Chlamydomonas reinhardtii (Chlamydomonas smithii)	Viridiplantae	Core	1237
<input checked="" type="checkbox"/>	vcariD8TYP9	Uncharacterized protein	Volvox carteri f. nagariensis	Viridiplantae	Peripheral	988
<input checked="" type="checkbox"/>	aproA0A087SRW5	PPM-type phosphatase domain-containing protein	Auxenochlorella protothecoides (Green microalga) (Chlorella protothecoides)	Viridiplantae	Core	708
<input checked="" type="checkbox"/>	cbraA0A388JMB4	PPM-type phosphatase domain-containing protein	Chara braunii (Braun's stonewort)	Viridiplantae	Core	704
<input checked="" type="checkbox"/>	aproA0A087SJZ6	PPM-type phosphatase domain-containing protein	Auxenochlorella protothecoides (Green microalga) (Chlorella protothecoides)	Viridiplantae	Core	543
<input checked="" type="checkbox"/>	creiA0A2K3DBF3	PPM-type phosphatase domain-containing protein	Chlamydomonas reinhardtii (Chlamydomonas smithii)	Viridiplantae	Core	491
<input checked="" type="checkbox"/>	osatiOQJMD4	Probable protein phosphatase 2C 3	Oryza sativa subsp. japonica (Rice)	Viridiplantae	Core	485

Check All Uncheck All

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

Output format: **Mismatches 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. Use the site search to navigate to OG6_131670.
- 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 ? Download

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

expand all | collapse all ☒ Hide zero counts

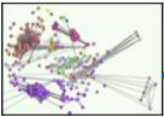
Type a taxonomic name	?
▼ Eukaryota (EUKA)	117
▶ Alveolates (ALVE)	114
▶ Metazoa (META)	3

- 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 Cluster graph

Click to open the Cluster graph in a new tab

Cluster graph of all proteins ?



Cluster Graph: OG6_131670 (97 proteins) ?

Back to Group page

Edge Options

Edge Type

- ☒ Ortholog
- ☒ Coortholog
- ☒ Inparalog
- ☒ Peripheral-Core
- ☐ Peripheral-Peripheral
- ☐ Other Similarities

E-Value Cutoff

Max E-Value: 1E-22

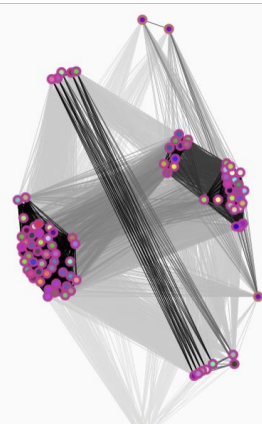
Node Options

Show Nodes By

- ☒ Taxa
- ☐ EC Numbers
- ☐ Pfam Domains

Mouse over a taxon legend to highlight sequences of that taxon.

gnip (1)	gnip-old (1)	hpl (1)
pad (1)	zber (1)	pbl (1)
gag (1)	zche (1)	psa (1)
pcym (1)	pcym (1)	pfal (1)
pfal-old (1)	pfag (1)	pfac (1)
pfad (1)	pfga (1)	pfpb (1)
pfgn (1)	pfho (1)	pfai (1)
pfke (1)	pfko (1)	pfkt (1)
pfml (1)	pfad (1)	pfan (1)

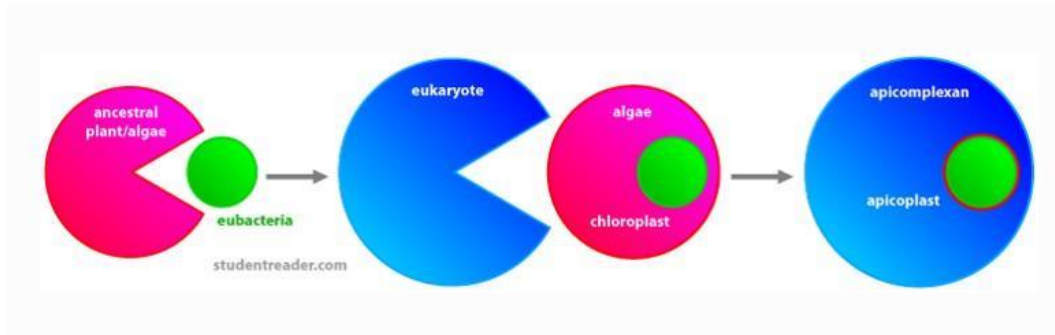


Sequence List ? Node Details ?

Search:

Accession	Taxon	Length
bbersBESB_010830	bbers	536
candicand_009760	cand	318
ccanycy_07054	ccay	553
ccnyfLO3442000	ccnyf	553
chocICHLRE4L_5030	chod	325
chocGY17_0000954	choi	325
chomChro05075	chom	325
chotCHTUS02012_400fg0200	chot	325
cmelCmelK0MEL1_07710	cmel	325
cmurCMU_011520	cmur	318
cpa-oldcpg8_5030	cpa-old	324
cpaoldcpg8_5030	cpa	324
cpaCPATCC_004380	cpa	324
csuICSU_001537	csu	577
ctyzICTYZ_00000545	ctyz	325
cubiCubi_03730	cubi	325
eaceIEAH_00023650	eace	550
eburEBH_0061790	ebur	551
efalIEAB_MNUS_13398.g1198	efal	547
emaxEMWY_00025950	emax	652
emilIEAH_0005590	emil	551
enecIEH_00050470	enec	553
epaIEPH_0014470	epa	551
etenIEH_00031620	eten	553
gnip-oldGNI_09890	gnip-old	387

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. Subcellular Localization search. 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?


The screenshot displays the search interface for *Plasmodium falciparum*. On the left, a sidebar titled "Search for..." lists various search categories. A red arrow points from the "P.f. Subcellular Localization" option in this list to the main search area. The main area is titled "Identify Genes based on P.f. Subcellular Localization" and features a "Localization" section with a dropdown menu set to "Apicoplast". A "Get Answer" button is located below the dropdown. At the bottom, a summary bar shows "Subcell Loc 499 Genes" as "Step 1" and a dashed box with a plus sign and the text "Add a step".

Search for...
expand all | collapse all
Filter the searches below...
Pathways and interactions
Phenotype
Protein features and properties
Protein targeting and localization
 Exported Protein
 P.f. Subcellular Localization
 Predicted Signal Peptide
 Transmembrane Domain Count
Proteomics

Identify Genes based on P.f. Subcellular Localization
Localization
Apicoplast
Get Answer

Subcell Loc
499 Genes
Step 1
+ Add a step

← Add a step to your search strategy ⓘ

The results will be  intersected with ▾ the results of Step 1.

Configure Search Learn More View Data Sets Used

Organism

1 selected, out of 622
select only these | add these | clear these

3d7 x ⓘ

- Apicomplexa
 - Aconoidasida
 - Haemosporida
 - Plasmodiidae
 - Plasmodium
 - Plasmodium falciparum
 - ☒ Plasmodium falciparum 3D7 [Reference]

Evidence

☒ Curated
☒ Computed
 select all | clear all

Limit to GO Slim terms

☐ Yes
☒ No

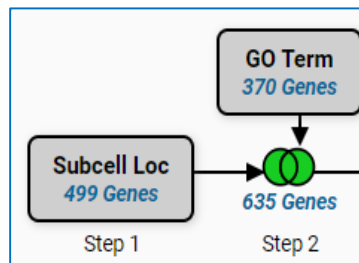
GO Term or GO ID

GO:0020011:apicoplast:7 x

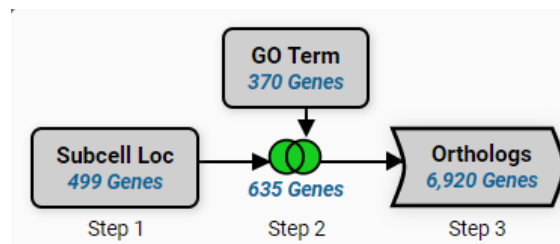
☐ GO Term or GO ID wildcard search

N/A

Run Step



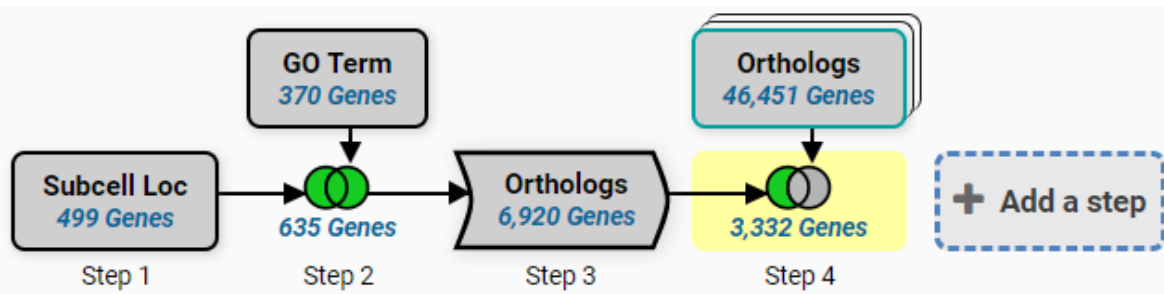
- b. Transform the results into their *Toxoplasma* and *Neospora* orthologs. Add a step to your strategy that transforms the results into *Toxoplasma* and *Neospora*.



- 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. First pull up all *Cryptosporidium* genes with the Genes by Taxonomy search and then transform these back to their *Toxoplasma* and *Neospora* orthologs for the subtraction to complete.

The screenshot shows the 'Add a step to your search strategy' dialog. On the left, there are three tabs: 'Combine with other Genes', 'Transform into related records', and 'Use Genomic Colocation to combine with other features'. The 'Combine with other Genes' tab is active, showing options for 'Choose how to combine with other Genes' (3 INTERSECT 4, 3 UNION 4, 3 MINUS 4) and 'Choose which Genes to combine. From...'. The 'From...' dropdown is set to 'A new search', and a search box contains 'organism'. A red arrow points to the 'Organism' option in the dropdown list. On the right, the 'Organism' panel shows '14 selected, out of 558' results. A search box contains 'crypto'. A red arrow points to the 'Cryptosporidiidae' checkbox, which is checked. Below it, a list of *Cryptosporidium* species and isolates is shown, each with a checkbox and a reference link.

The screenshot shows the 'My Search Strategy' page in VEuPathDB. The top navigation bar includes 'View', 'Analyze', 'Revise', 'Make nested strategy', 'Insert step before', 'Orthologs', and 'Delete'. The 'Make nested strategy' button is circled in red. Below the navigation bar, the 'Details for step Organism' are shown, listing 54,377 genes. A red arrow points to the 'Organism' step in the search strategy diagram. The diagram shows a sequence of steps: Step 1 (Subcell Loc, 499 Genes), Step 2 (GO Term, 370 Genes), Step 3 (Orthologs, 6,920 Genes), and Step 4 (Organism, 54,377 Genes). A red arrow points to the 'Organism' step in the diagram. The 'Organism' step is highlighted in yellow. A red arrow points to the 'Make nested strategy' button in the top navigation bar.



Expanded view of *Orthologs*

