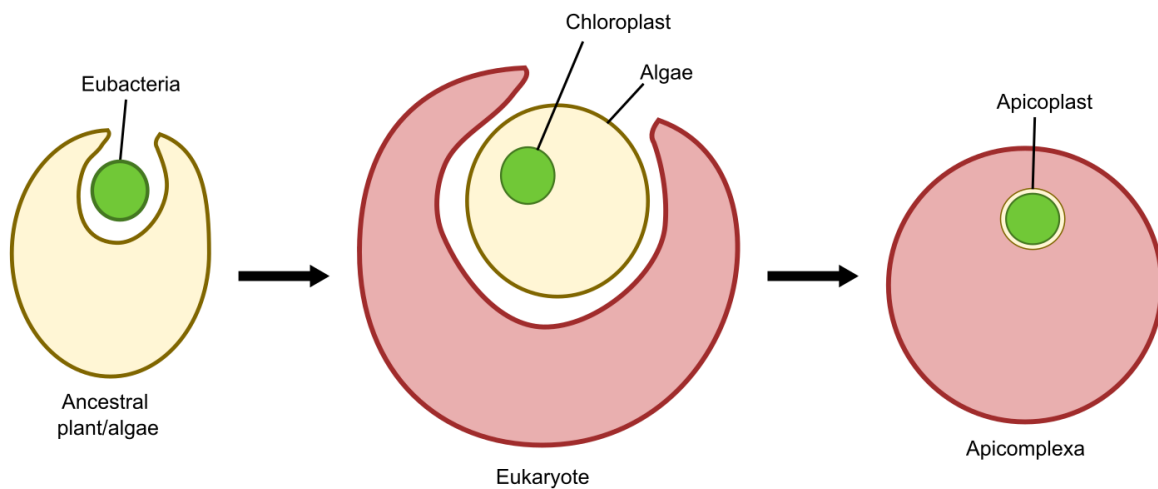


Integrate Phyletic and Genomic data in a complex Search Strategy

Goal: Identify apicoplast targeted genes in *Toxoplasma* and *Neurospora*. Note: For this exercise use <https://veupathdb.org/veupathdb/app>

What is an apicoplast?

The apicoplast is an organelle unique to apicomplexans. 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 the apicoplast.
Hint: Navigate to the Pfal 3D7 Subcellular Localization search for Apicoplast. You can filter the types of search by text query.

The screenshot shows a web application interface. At the top, there is a search bar with the placeholder text "Site search, e.g. PF3D7_1133400 or *reductase or 'binding protein'". Below the search bar is a navigation menu with links: "My Strategies", "Searches", "Tools", "My Workspace", "Data", "About", and "Help". The "Searches" link is circled in red. A dropdown menu is open under "Searches", showing a search box with "pf" and a list of options: "Genes", "Protein features and properties", "InterPro Domain", "Protein targeting and localization", "Pf3D7 Subcellular Localization", and "Structure analysis". A red arrow points from "Pf3D7 Subcellular Localization" to a "Localization" section on the right. The "Localization" section has a dropdown menu with "Apicoplast" selected. Below this is a "Get Answer" button, which is also circled in red. A red arrow points from the "Get Answer" button to a yellow box labeled "Subcell Loc 499 Genes" with "Step 1" below it. To the right of this box is a dashed blue box with a plus sign and the text "Add a step".

Site search, e.g. PF3D7_1133400 or *reductase or "binding protein"

My Strategies **Searches** Tools My Workspace Data About Help

pf

Genes

Protein features and properties

InterPro Domain

Protein targeting and localization

Pf3D7 Subcellular Localization

Structure analysis

Find genes whose protein products are likely targeted to the apicoplast or the RBC membrane.

Identify Genes based on P.f. Subcellular Localization

Localization

Apicoplast

Get Answer

Subcell Loc
499 Genes
Step 1

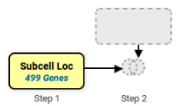
+ Add a step

- b. Expand your list of potentially Apicoplast targeted proteins by adding a GO terms search for the term "apicoplast" or the GO ID: "GO:0020011" in *P. falciparum* 3D7 (Which Boolean operation should you use? Union or intersect?)




Add a step to your search strategy

Combine with other Genes



Step 1 Step 2

Transform into related records



Step 1 Step 2


Use **Genomic Colocation** to combine with other features



1 Choose *how* to combine with other Genes

- ☐ 1 INTERSECT 2 ☒ 1 UNION 2 ☐ 1 MINUS 2 ☐ 2 MINUS 1

2 Choose *which* Genes to combine. From...



- ☒ A new search ☐ An existing strategy ☐ My basket



[Function prediction](#)
[GO Term](#)
[Phenotype](#)
[CRISPR Phenotype](#)
[Text](#)
[Text \(product name, notes, etc.\)](#)

Search for Genes by GO Term

The results will be  unioned 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](#)

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


Evidence

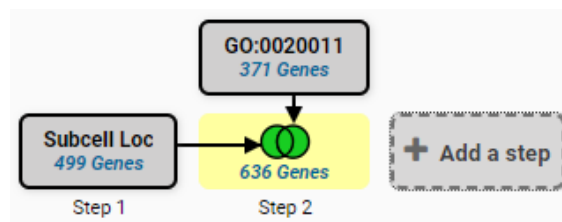
- ☒ Curated
☒ Computed
[select all](#) | [clear all](#)

Limit to GO Slim terms

- ☐ Yes
☒ No

GO Term or GO ID





- c. Add a step to your strategy that transforms the results with *Toxoplasma* and *Neospora* orthologs. Note that this finds orthologous genes from many strains and species.

Transform into related records

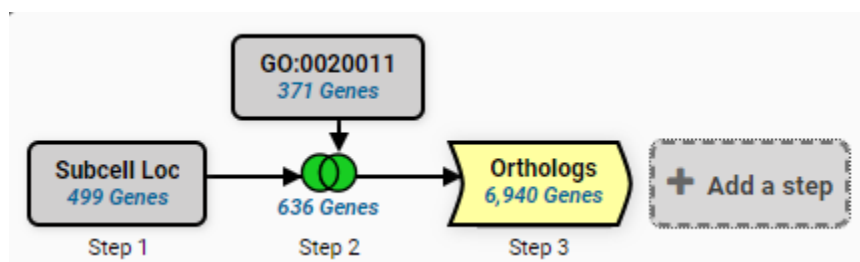
Organism

17 selected, out of 675
[select all](#) | [clear all](#) | [expand all](#) | [collapse all](#)

Filter list below... ☐ Reference only

- ☐ 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
 - ☐ Sarcocystis
 - ☒ Toxoplasma
 - ☐ Eugregarinorida
- ☐ Chromeraceae
- ☐ Euglenozoa
- ☐ Fornicata
- ☐ Fungi
- ☐ Heterolobosea
- ☐ Metazoa
- ☐ Oomycota
- ☐ Parabasalia
- ☐ Preaxostyla
- ☐ Vitrellaceae

Run Step



- d. Although *Cryptosporidium* is an apicomplexan parasite it has lost its apicoplast! Use this fact to refine your results from the above search and remove genes that also have orthologs in *Cryptosporidium*.

Hint: try subtracting out any orthologs present in *Cryptosporidium*. You will need to use a nested strategy. First retrieve all *Cryptosporidium* genes with the Genes by Taxonomy search

and then transform these to their Toxoplasma and Neospora orthologs for the subtraction to complete. Think about what kind of intersection you should be using!

← Add a step to your search strategy ?

Combine with other Genes

Step 3 Step 4

Transform into related records

Step 3 Step 4

1 Choose *how* to combine with other Genes

☐ 3 INTERSECT 4
 ☐ 3 UNION 4
 ☒ 3 MINUS 4
 ☐ 4 MINUS 3

2 Choose *which* Genes to combine. From...

☒ A new search
 ☐ An existing strategy
 ☐ My basket

Taxonomy
Q Organism



Add a step to your search strategy

The results will be subtracted from | v the resu

Configure Search Learn More View Data Sets Used

Reset values to default

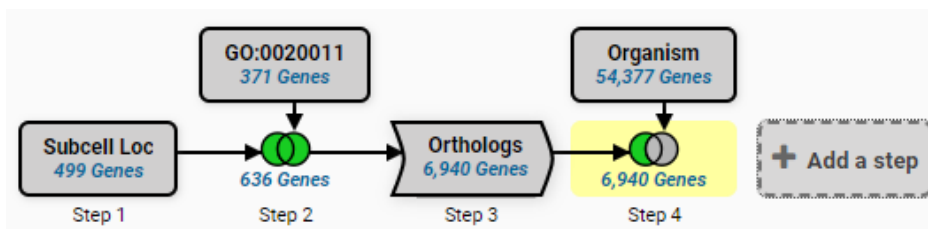
? Organism

14 selected, out of 675

select only these | add these | clear these

cryptos x ? ☐ Reference only

- [-] Apicomplexa
 - [-] Conoidasida
 - [-] Coccidia
 - ☒ Cryptosporidiidae
 - ☒ Cryptosporidium andersoni isolate 30847 [Reference]
 - ☒ Cryptosporidium bovis isolate 45015 [Reference]
 - ☒ Cryptosporidium hominis
 - ☒ Cryptosporidium hominis TU502 [Reference]
 - ☒ Cryptosporidium hominis UdeA01
 - ☒ Cryptosporidium hominis isolate 30976
 - ☒ Cryptosporidium hominis isolate TU502_2012
 - ☒ Cryptosporidium meleagridis strain UKMEL1 [Reference]
 - ☒ Cryptosporidium muris RN66 [Reference]
 - ☒ Cryptosporidium parvum
 - ☒ Cryptosporidium parvum IOWA-ATCC
 - ☒ Cryptosporidium parvum Iowa II [Reference]
 - ☒ Cryptosporidium ryanae 45019 [Reference]
 - ☒ Cryptosporidium sp. chipmunk genotype I strain 37763 [Reference]
 - ☒ Cryptosporidium tyzzeri isolate UGA55 [Reference]
 - ☒ Cryptosporidium ubiquitum isolate 39726 [Reference]



My Search Strategies

Opened (1) All (1) Public (11) Help

Unnamed Search Strategy *

Subcell Loc 499 Genes → GO:0020011 371 Genes → Orthologs 6,940 Genes → Organism 54,377 Genes → 6,940 Genes

Details for step Organism 54377 Genes

Organism ▶ Cryptosporidium andersoni isolate 30847, Cryptosporidium bovis isolate 45015, Cryptosporidium hominis TU502, Cryptosporidium hominis UdeA01, Cryptosporidium hominis isolate 30976, Cryptosporidium hominis isolate TU502_2012, Cryptosporidium meleagridis strain UKMEL1, Cryptosporidium muris RN66, Cryptosporidium parvum IOWA-ATCC, Cryptosporidium parvum Iowa II, Cryptosporidium ryanae 45019, Cryptosporidium sp. chipmunk genotype I strain 37763, Cryptosporidium tyzzeri isolate UGA55, Cryptosporidium ... Show more

Give this search a weight

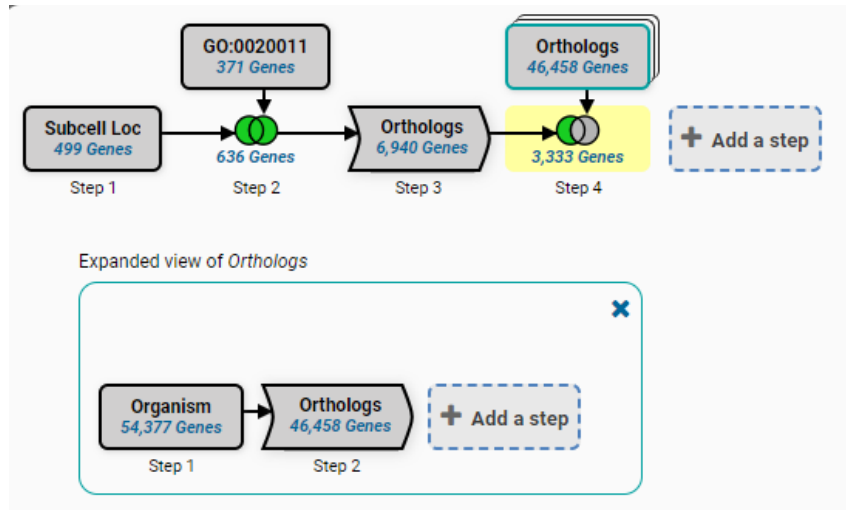
Organism

17 selected, out of 675

select all | clear all | expand all | collapse all

Filter list below... ? Ref:

- ▶ ☐ 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
 - ▶ ☐ Sarcocystis
 - ▶ ☒ Toxoplasma
 - ▶ ☐ Eugregarinorida
- ▶ ☐ Chromeraceae
- ▶ ☐ Euglenozoa
- ▶ ☐ Fornicata
- ▶ ☐ Fungi
- ▶ ☐ Heterolobosea
- ▶ ☐ Metazoa
- ▶ ☐ Oomycota
- ▶ ☐ Parabasalia
- ▶ ☐ Preaxostyla
- ▶ ☐ Vitrellaceae



This leaves you with apicoplast specific genes for *Toxoplasma* and *Neospora* that you could target in future research.

<https://veupathdb.org/veupathdb/app/workspace/strategies/import/543f14bfab645f7e>