Orthology and Ontologies

Learning objectives:

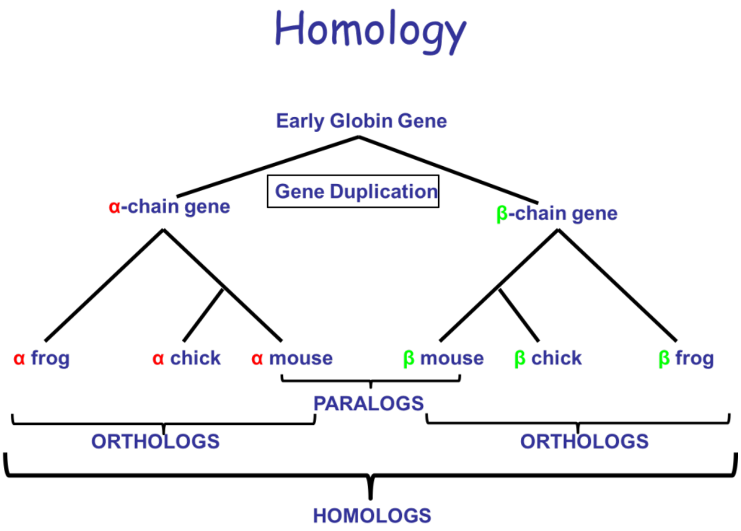
- Combine searches using the strategy system.
- GO enrichment

About Orthology and Phyletics

Homologs are genes that share ancestry either by speciation (orthologs), gene duplication (paralogs), or gene transfer events (xenologs). Paralogs of a conserved gene may occur in a single species or strain. Conserved sequences in genomes can be used to infer evolutionary history (e.g., ribosomal sequences), their similarities and differences can be used to trace the divergence and evolution of organisms. Genes that share function by convergent evolution, but do not share ancestry are known as analogs.

Ortholog groups can also allow you to explore the potential functions of a gene, or group of genes across species. In pathogens like *Plasmodium falciparum*, ortholog groups might facilitate the identification of potential targets for drug or vaccine development. A good place to start for more information about homology is Koonin, EV. Orthologs, paralogs and evolution. Annu Rev Genet 2005.

-



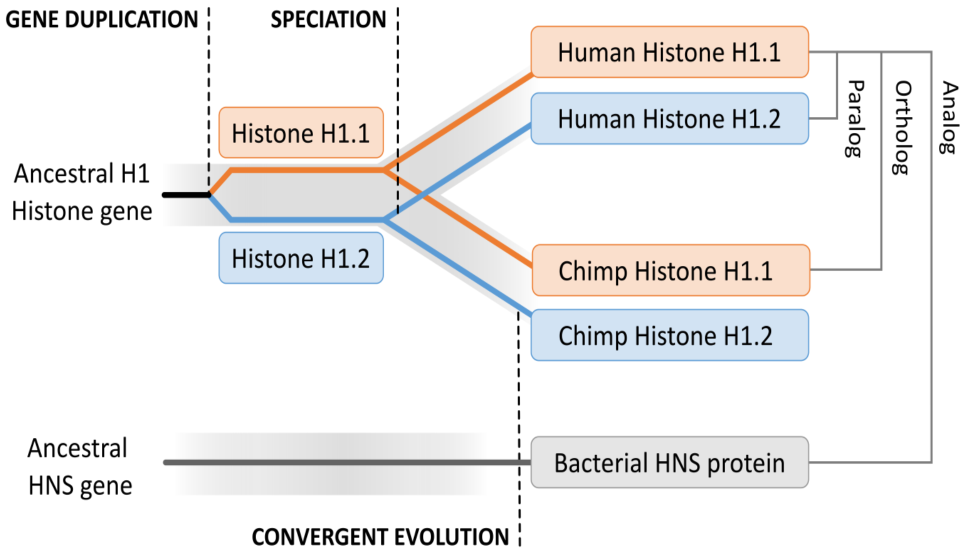


Figure SEQ Figure * ARABIC 1. Gene phylogeny (orange and blue) within species phylogeny (grey). Top shows an ancestral gene duplication event, producing two paralogs of the Histone H1 gene, producing H1.1 and H1.2. This is followed by a speciation event leading to Chimpanzee and Human Orthologs of the two genes. Bottom shows a gene with separate evolutionary origin that has evolved similar function to H1 Histones through convergent evolution, HNS (histone-like nucleoid-structuring protein). HNS is a bacterial analog to H1 Histone. Figure adapted from this image by Thomas Shafee (2018).

About OrthoMCL

OrthoMCL is a genome-scale database that groups orthologous protein sequences across the tree of life. An orthogroup contains genes descended from a common ancestor by a process of duplication and speciation (see figure above), so a single orthogroup may contain both genes across different species with similar function and paralogs within a single species. Each protein in every OrthoMCL species is assigned to precisely one ortholog group (e.g. OG6_162879). Importantly, proteins within a single OrthoMCL group 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). Orthology is important in predicting the function of the rapidly increasing number of newly identified proteins produced by genome sequencing and the automated discovery of protein sequences (Glover et al. 2019). Within VEuPathDB, orthology can be used to transform a list of genes from one species into their closest equivalents in another species.

OrthoMCL contains two sets of genomes. A **Core** set of 150 genomes have been chosen as well annotated reference organisms that broadly represent the major branches of the tree of life. The OrthoMCL algorithm uses BLAST to calculate pairwise distances among all proteins in the 150 core genomes, normalizes the scores for sequence length and evolutionary distance, then uses MCL clustering (Dongen 2000; www.micans.org/mcl) to create orthogroups of similar proteins. All of the non-core VEuPathDB species (pathogens, hosts, and vectors) have been added as **Peripheral** organisms, in some cases including multiple strains and genome assemblies for the same species. All proteins from the Peripheral organisms are assigned to the most similar Core cluster by best BLAST score, but proteins that do not match any Core protein with an e-value better than 1e^{-5} are set aside as

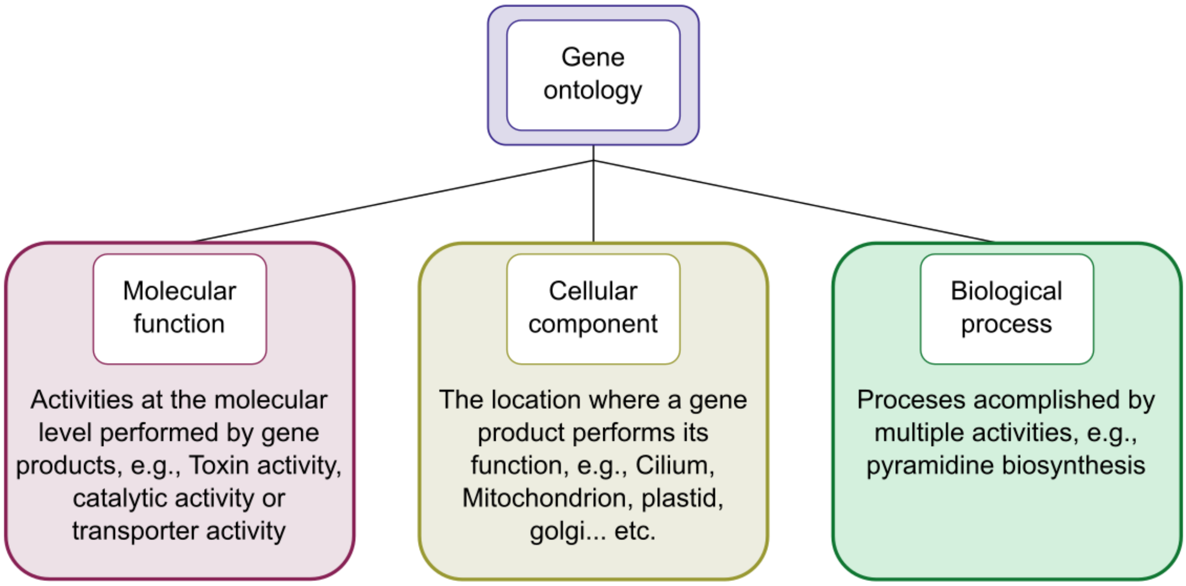
Residuals. Pairwise BLAST distances among all Residual proteins are computed and used for a second round of MCL clustering to create Residual groups (e.g. OG6r20_100305)

The OrthoMCL website offers the ability to explore ortholog groups by taxonomy, number of proteins or species, sequence similarity, EC numbers, PFAM domains, and text search of gene descriptions. Users can use the Ortholog Group or Protein queries in the grey Search box to the left or the Searches menu in the header bar, or just type a search term in the 'Site search' box above which will result in a list of proteins and groups to explore. In addition, users can use a VEuPathDB Galaxy workflow to map their own set of proteins (e.g. protein sequences derived from a genome sequence of an organism) to OrthoMCL groups. See the Assign Proteins to Groups page.

For more information, see the About OrthoMCL and OrthoMCL FAQ pages.

About Gene Ontology

Ontologies are a controlled vocabulary of terms and concepts with relationships between them. Gene Ontology describes the knowledge of biological sciences and divides this knowledge into three broad categories: Molecular function, cellular component, and biological process. To learn more about Gene Ontology, please visit: http://geneontology.org/docs/ontology-documentation/



Activities at the molecular level performed by gene products, e.g., Toxin activity, catalytic activity or transporter activity

The location where a gene product performs its function, e.g., Cilium, Mitochondrion, plastid, golgi... etc.

Proceses acomplished by multiple activities, e.g., pyramidine biosynthesis

A gene can be assigned a GO term either manually (by an annotator or curator when they evaluate experimental evidence from a publication) or computationally (based on the GO terms of genes that share sequence or functional domains). The origin of the assignment is documented; some researchers believe that manually assigned functional annotations are more accurate than those that are electronically transferred since a researcher has reviewed the manually annotated assignments. GO terms can be used to test whether your set of genes are enriched for a molecular function, cellular component, or biological process.

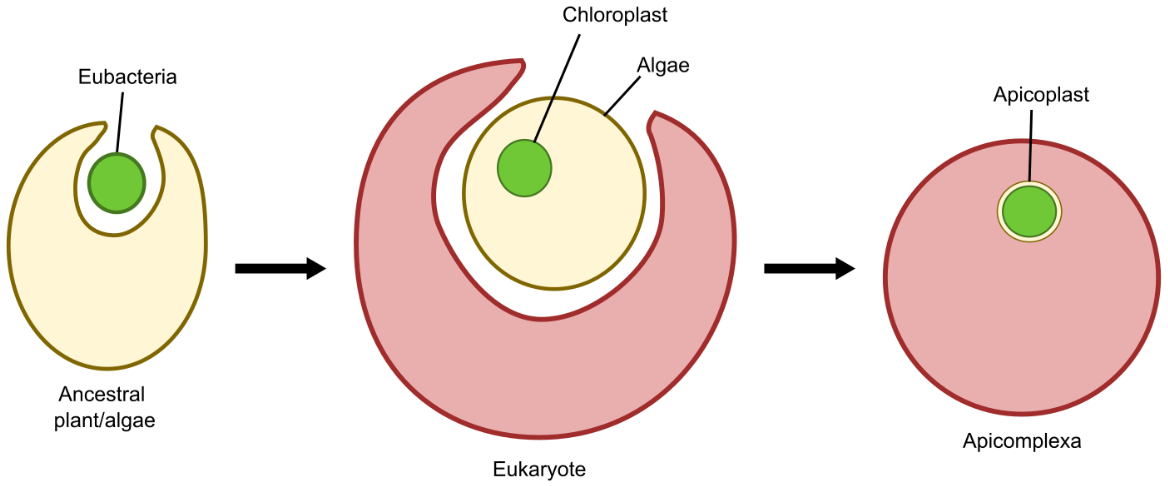
For example: A researcher performs a proteomics experiment on a protein fraction collected during an antimalarial treatment and identifies 100 proteins in total. When they examine the GO terms assigned to the gene set corresponding to the proteome, they see that 25 genes are assigned GO:0016301, kinase activity. Out of 5000 genes in the genome, only 100 are assigned GO:0016301. There is an overrepresentation of GO:0016301 in the researcher's proteome which is 'enriched' for kinase activity.

A standard enrichment determination method employs Fisher's exact test, a statistical test that evaluates a 2x2 contingency table (in this case, the number of genes in my set versus number of genes from the same genome not in my set, and number of genes with GO term X versus number of genes without term X). This test produces a p-value between 0 and 1, where $p \leq 0.05$ is considered significant (that is, less than 5% probability that the enrichment is due to chance).
However, the test is performed for each of the 100s of GO terms, increasing the chances that a GO term will be incorrectly considered enriched (a false positive, or type I, error). Thus, the original p-value must be adjusted for so-called multiple hypothesis testing, resulting in an adjusted p-value such as the Benjamini-Hochberg false discovery rate (FDR) or Bonferroni adjusted p-value.

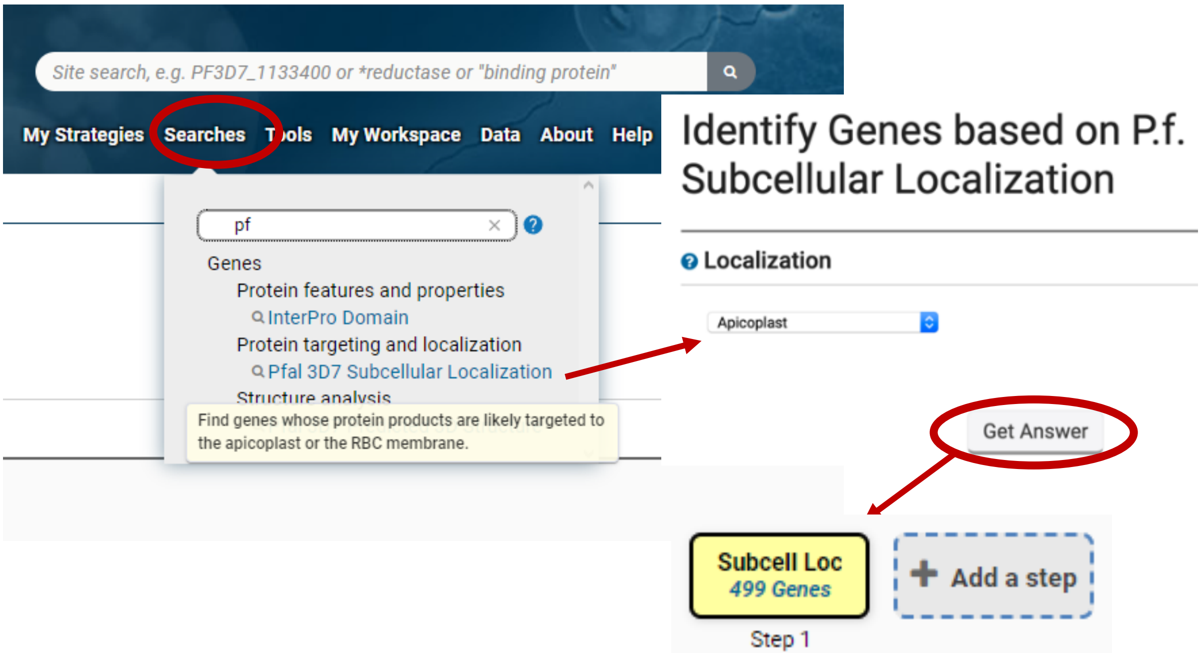
1. **Identify apicoplast targeted genes in Toxoplasma and Neurospora**. Note: For this exercise use https://veupathdb.org/veupathdb/app

What is an apicoplast?

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.



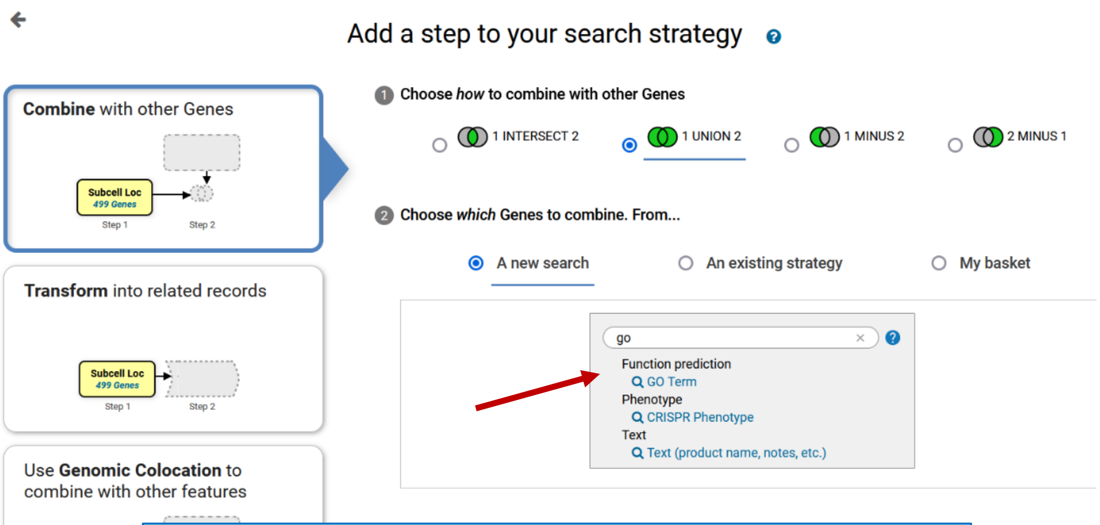
a. Start by finding genes in *Plasmodium* that are predicted to target the apicoplast. *Hint: Navigate to the Pfal 3D7 Subcellular Localization search for Apicoplast. You can filter the types of search by text query.*



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
Pfal 3D7 Subcellular Localization
Structure analysis
Identify Genes based on P.f. Subcellular Localization
Localization
Apicoplast
Get Answer
Subcell Loc
499 Genes
Step 1
Add a step

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

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

Transform into related records

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...

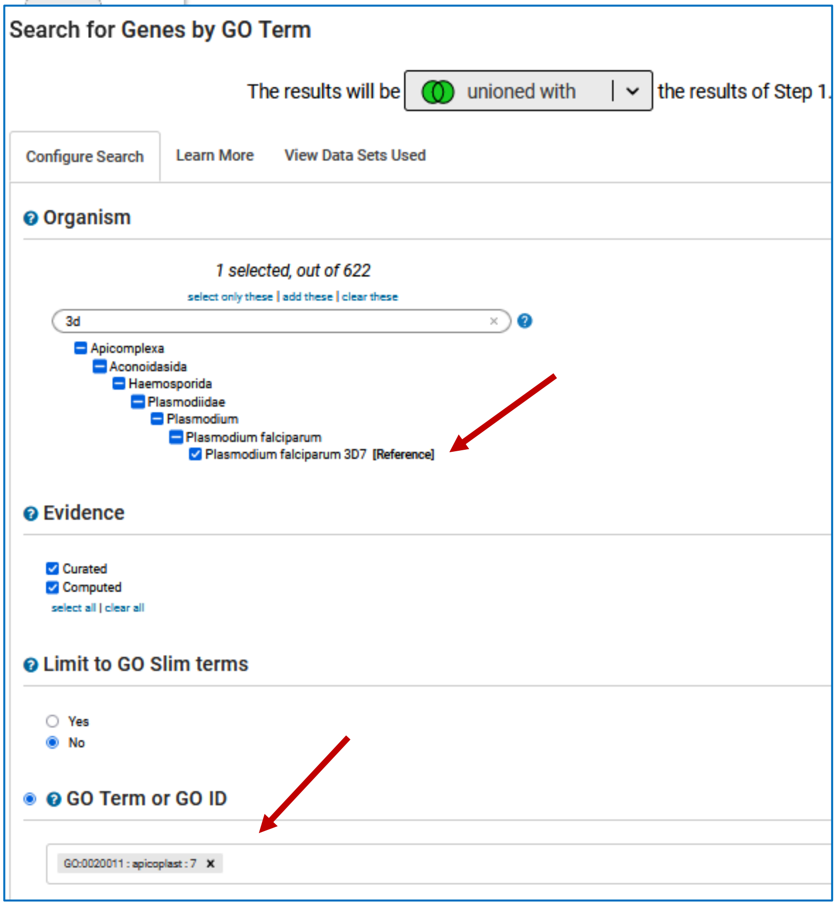
● A new search ○ An existing strategy ○ My basket

go

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.

Organism

1 selected, out of 622

select only these | add these | clear these

3d

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

Evidence

☑ Curated
☑ Computed

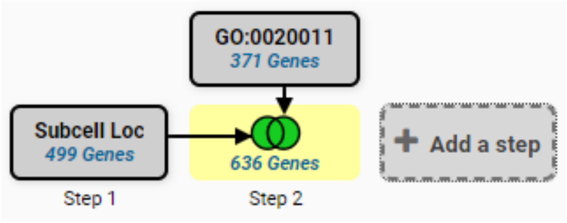
select all | clear all

Limit to GO Slim terms

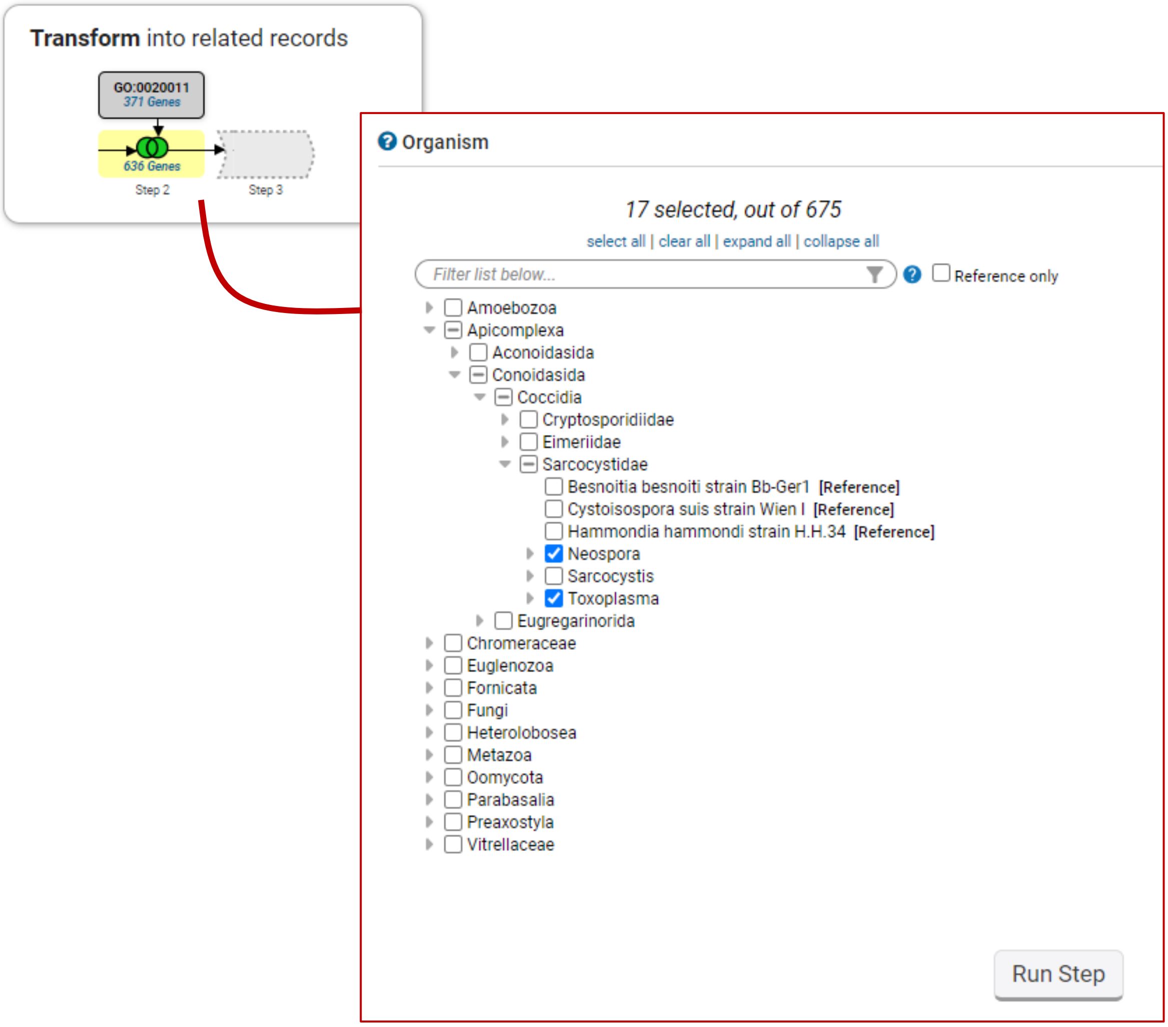
○ Yes
● No

● GO Term or GO ID

GO:0020011 : apicoplast : 7 ✖



c. Add a step to your strategy that transforms the results with *Toxoplasma* and *Neospora* orthologs.



Organism

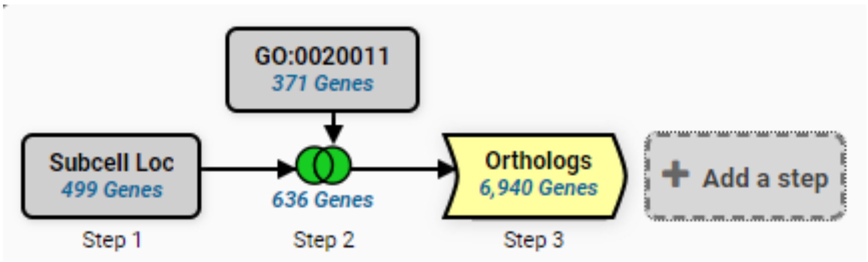
17 selected, out of 675

select all | clear all | expand all | collapse all

Filter list below... □ Reference only

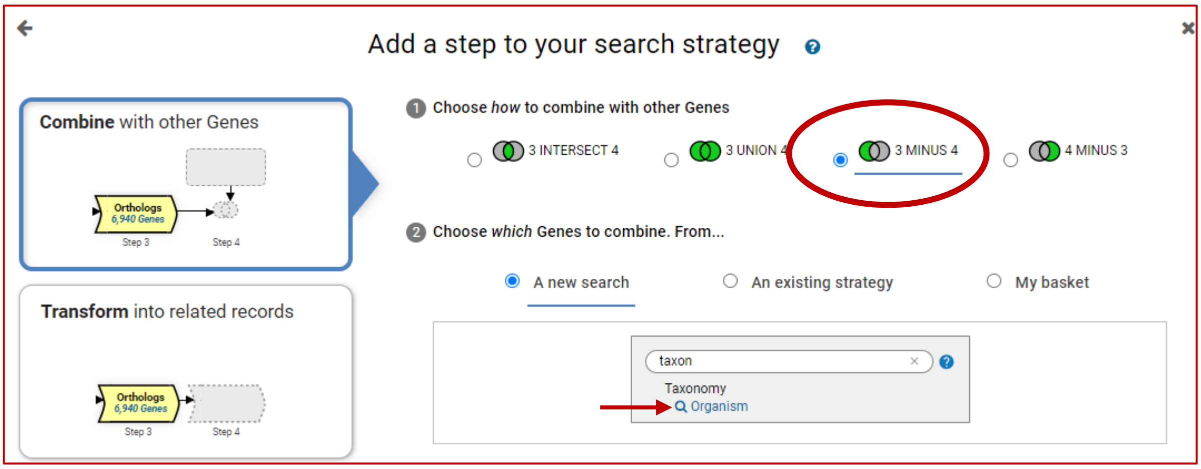
- □ Amoebozoa
- ⊟ Apicomplexa
  - □ Aconoidasida
  - ⊟ Conoidasida
    - ⊟ Coccidia
      - □ Cryptosporidiidae
      - □ Eimeriidae
      - ⊟ Sarcocystidae
        - □ Besnoitia besnoiti strain Bb-Ger1 **[Reference]**
        - □ Cystoisospora suis strain Wien I **[Reference]**
        - □ Hammondia 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

Transform into related records

1 Choose *how* to combine with other Genes

○ 3 INTERSECT 4 ○ 3 UNION 4 ● 3 MINUS 4 ○ 4 MINUS 3

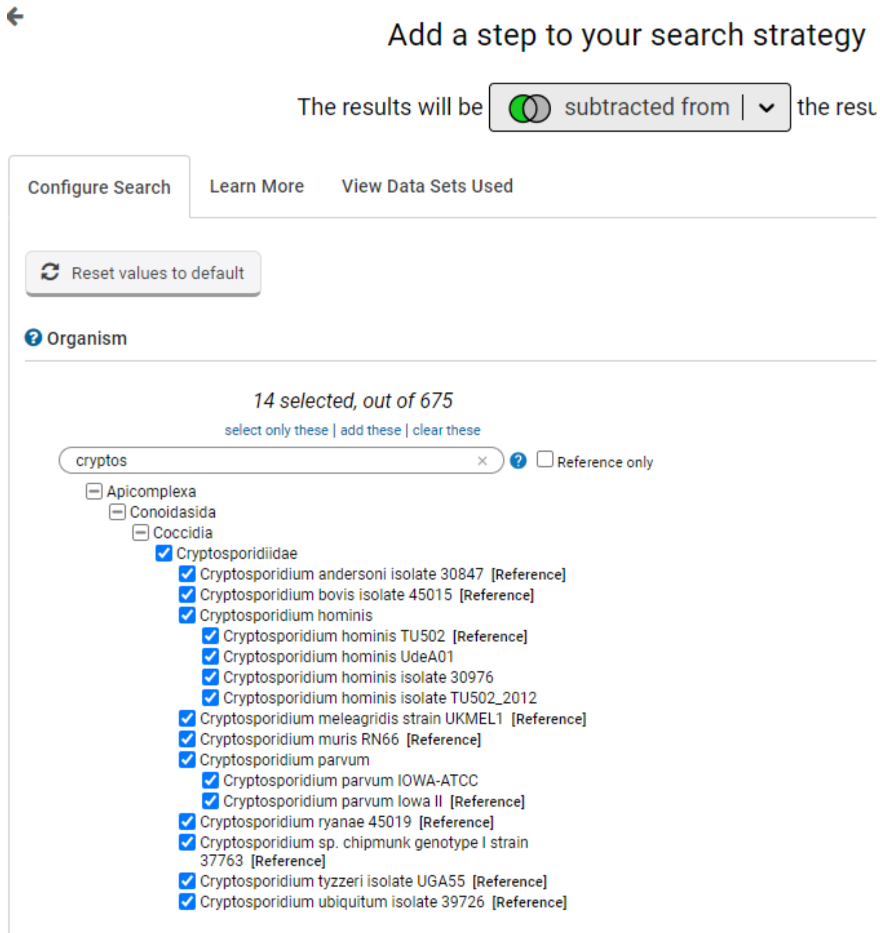
2 Choose *which* Genes to combine. From...

● A new search ○ An existing strategy ○ My basket

taxon ×

Taxonomy

Organism



Add a step to your search strategy

The results will be subtracted from the resu

Configure Search | Learn More | View Data Sets Used

Reset values to default

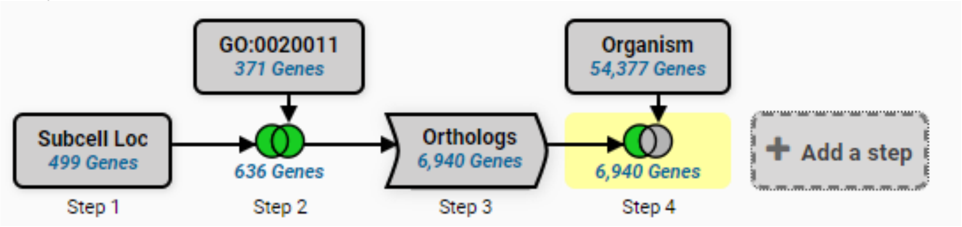
Organism

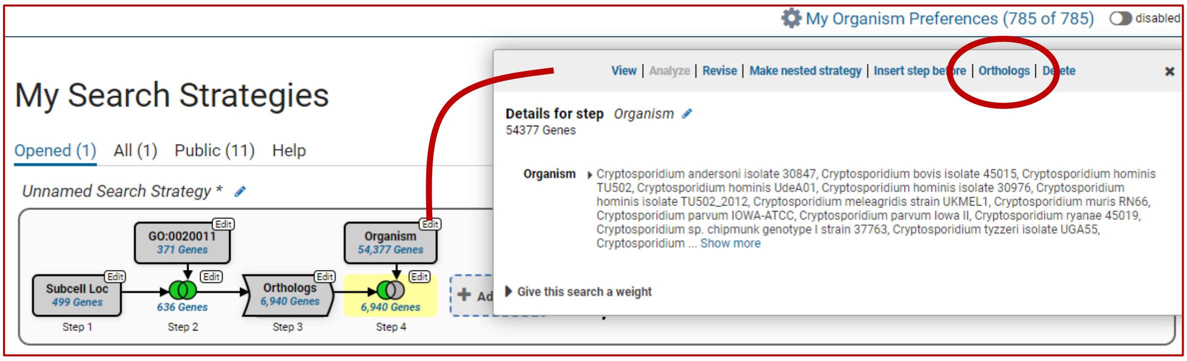
14 selected, out of 675

select only these | add these | clear these

cryptos × □ Reference only

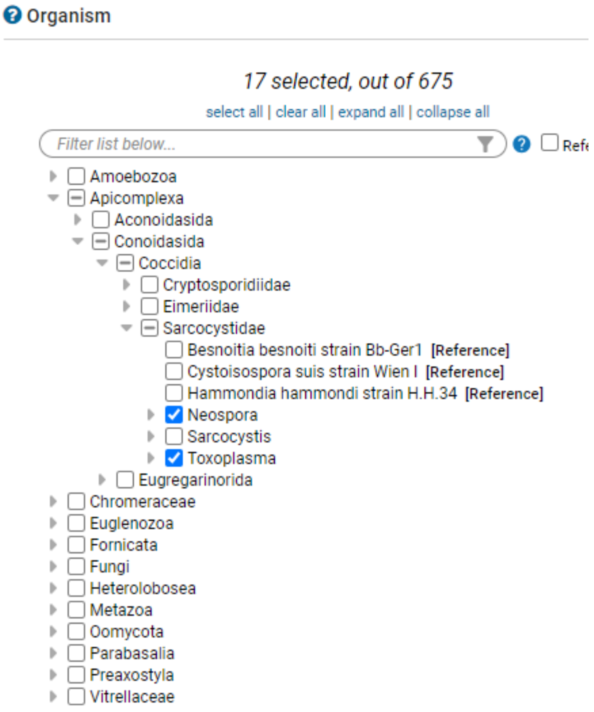
- Apicomplexa
  - Conoidasida
    - Coccidia
      - ☑ Cryptosporidiidae
        - ☑ Cryptosporidium andersoni isolate 30847 [Reference]
        - ☑ Cryptosporidium bovis isolate 45015 [Reference]
        - ☑ Cryptosporidium hominis
          - ☑ Cryptosporidium hominis TU502 [Reference]
          - ☑ 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]





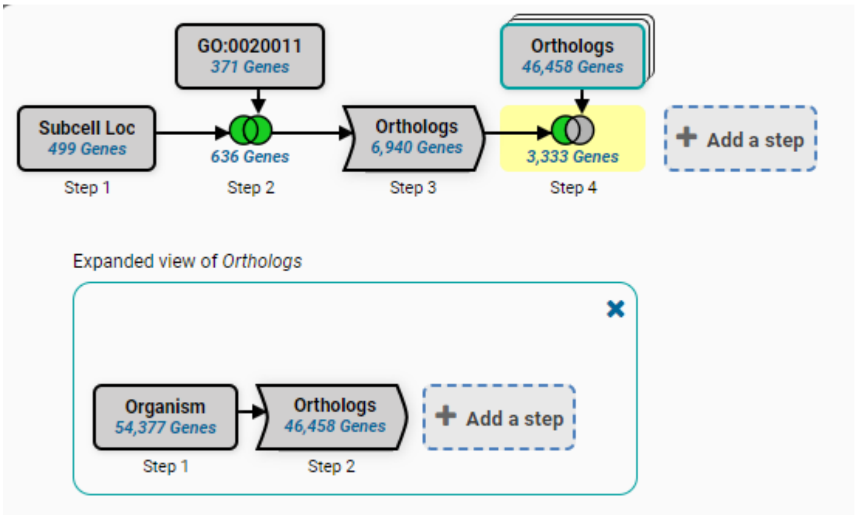
Details for step *Organism*
54377 Genes

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



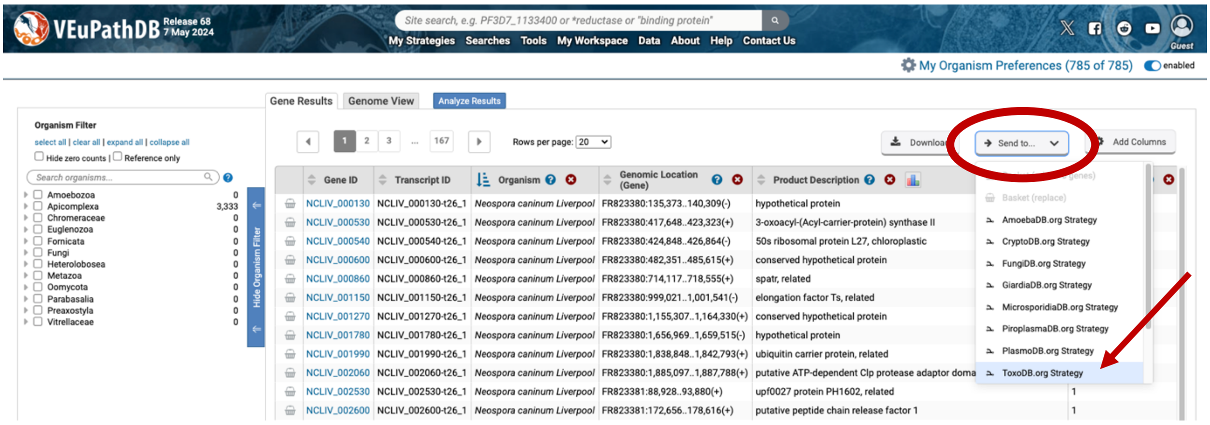
17 selected, out of 675

- □ 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

e. Use the **Send to** button to explore the data in ToxoDB.



Organism Filter

select all | clear all | expand all | collapse all

□ Hide zero counts | □ Reference only

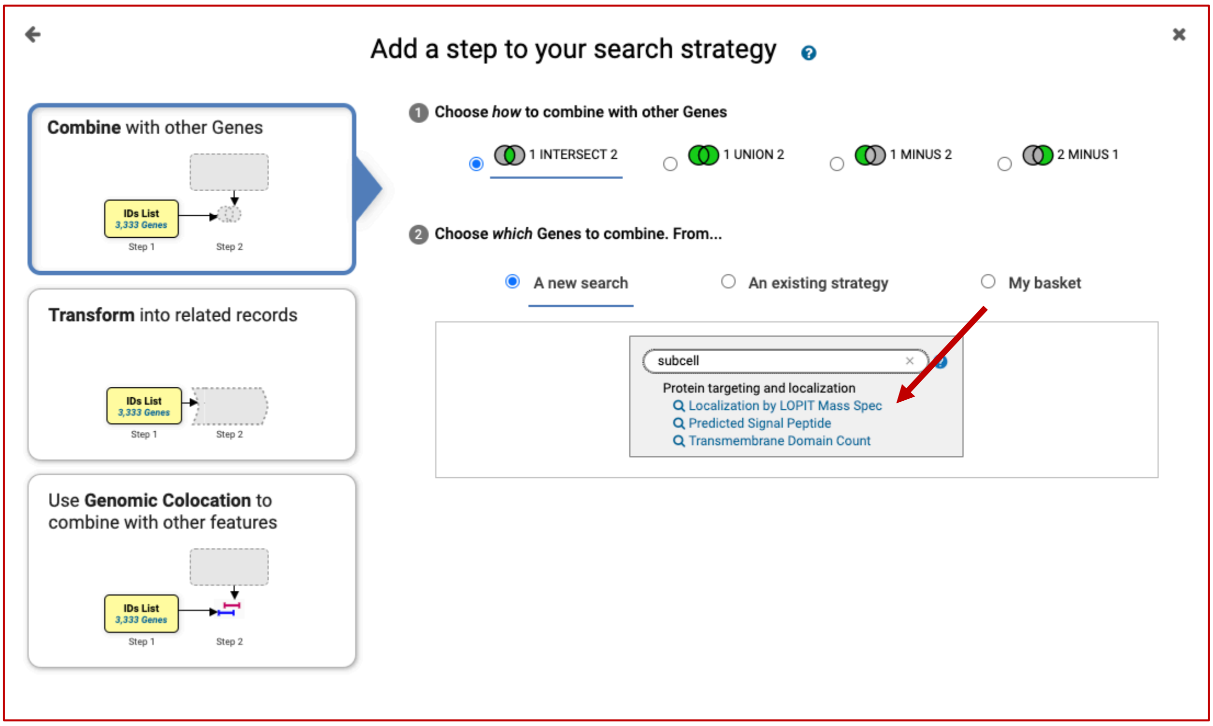
□ Amoebozoa	0
□ Apicomplexa	3,333
□ Chromeraceae	0
□ Euglenozoa	0
□ Fornicata	0
□ Fungi	0
□ Heterolobosea	0
□ Metazoa	0
□ Oomycota	0
□ Parabasalia	0
□ Preaxostyla	0
□ Vitrellaceae	0

Gene ID	Transcript ID	Organism	Genomic Location (Gene)	Product Description	
NCLIV_000130	NCLIV_000130-t26_1	*Neospora caninum Liverpool*	FR823380:135,373..140,309(-)	hypothetical protein	
NCLIV_000530	NCLIV_000530-t26_1	*Neospora caninum Liverpool*	FR823380:417,648..423,323(+)	3-oxoacyl-(Acyl-carrier-protein) synthase II	
NCLIV_000540	NCLIV_000540-t26_1	*Neospora caninum Liverpool*	FR823380:424,848..426,864(-)	50s ribosomal protein L27, chloroplastic	
NCLIV_000600	NCLIV_000600-t26_1	*Neospora caninum Liverpool*	FR823380:482,351..485,615(+)	conserved hypothetical protein	
NCLIV_000860	NCLIV_000860-t26_1	*Neospora caninum Liverpool*	FR823380:714,117..718,555(+)	spatr, related	
NCLIV_001150	NCLIV_001150-t26_1	*Neospora caninum Liverpool*	FR823380:999,021..1,001,541(-)	elongation factor Ts, related	
NCLIV_001270	NCLIV_001270-t26_1	*Neospora caninum Liverpool*	FR823380:1,155,307..1,164,330(+)	conserved hypothetical protein	
NCLIV_001780	NCLIV_001780-t26_1	*Neospora caninum Liverpool*	FR823380:1,656,969..1,659,515(-)	hypothetical protein	
NCLIV_001990	NCLIV_001990-t26_1	*Neospora caninum Liverpool*	FR823380:1,838,848..1,842,793(+)	ubiquitin carrier protein, related	
NCLIV_002060	NCLIV_002060-t26_1	*Neospora caninum Liverpool*	FR823380:1,885,097..1,887,788(+)	putative ATP-dependent Clp protease adaptor doma	
NCLIV_002530	NCLIV_002530-t26_1	*Neospora caninum Liverpool*	FR823381:88,928..93,880(+)	upf0027 protein PH1602, related	1
NCLIV_002600	NCLIV_002600-t26_1	*Neospora caninum Liverpool*	FR823381:172,656..178,616(+)	putative peptide chain release factor 1	1

Send to... menu:

- Basket (replace)
- AmoebaDB.org Strategy
- CryptoDB.org Strategy
- FungiDB.org Strategy
- GiardiaDB.org Strategy
- MicrosporidiaDB.org Strategy
- PiroplasmaDB.org Strategy
- PlasmoDB.org Strategy
- ToxoDB.org Strategy

f. Once your gene list is in ToxoDB intersect your results with the hyper_LOPIT subcellular localization data. Does the hyper_LOPIT data confirm the apicoplast location of *T. gondii* ME49 genes?



Add a step to your search strategy

Combine with other Genes

Transform into related records

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

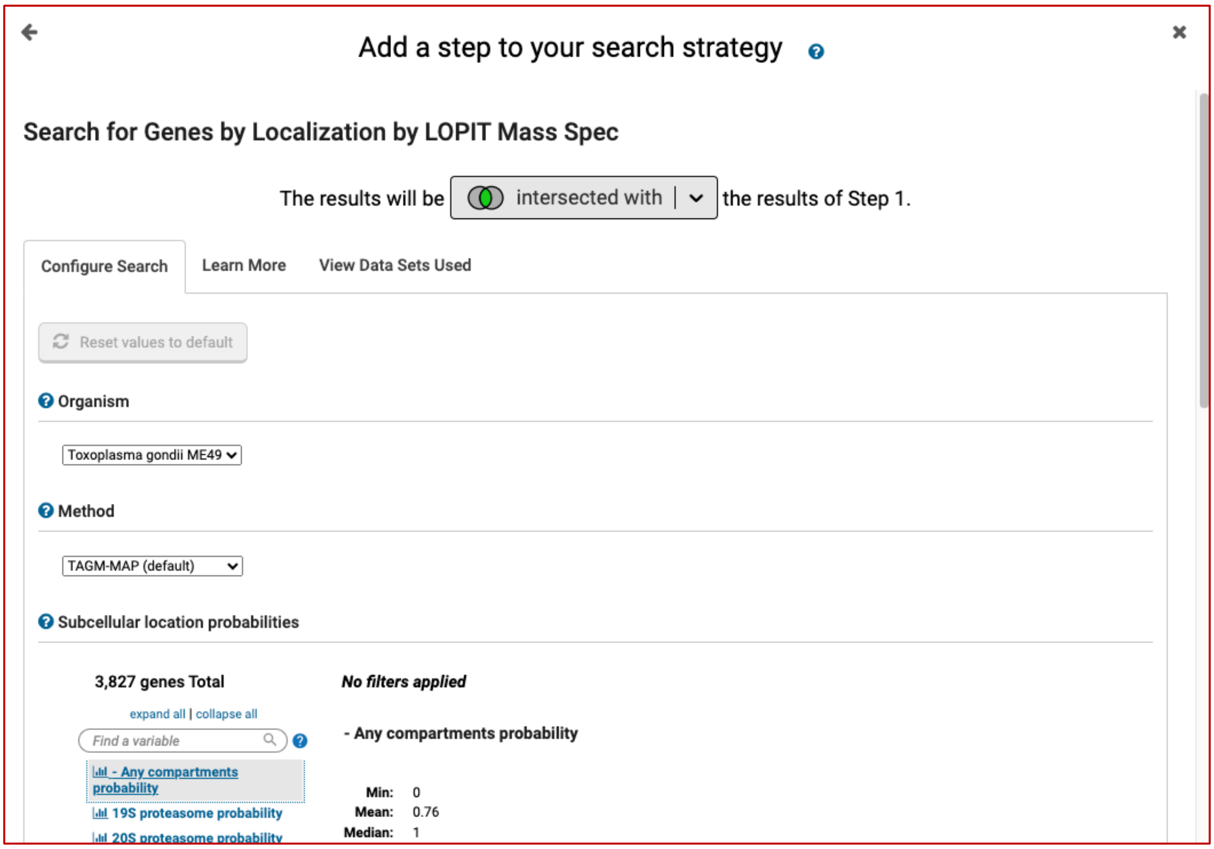
2 Choose *which* Genes to combine. From...

● A new search ○ An existing strategy ○ My basket

subcell

Protein targeting and localization

- Localization by LOPIT Mass Spec
- Predicted Signal Peptide
- Transmembrane Domain Count



Add a step to your search strategy

Search for Genes by Localization by LOPIT Mass Spec

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

Configure Search
Learn More
View Data Sets Used
Reset values to default

Organism

Toxoplasma gondii ME49

Method

TAGM-MAP (default)

Subcellular location probabilities

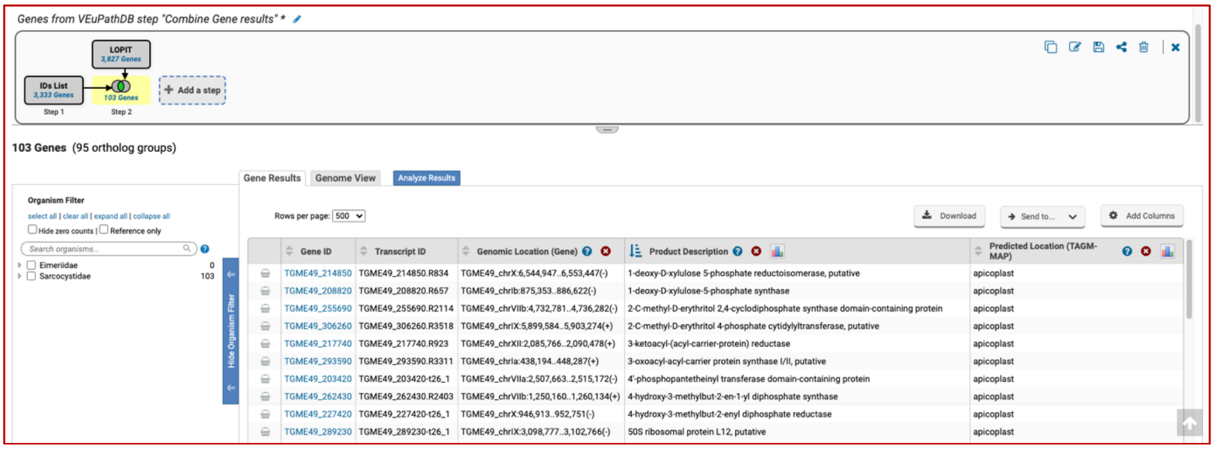
3,827 genes Total

No filters applied

expand all | collapse all
Find a variable
- Any compartments probability
19S proteasome probability
20S proteasome probability

- Any compartments probability

Min: 0
Mean: 0.76
Median: 1



*Genes from VEuPathDB step "Combine Gene results" **

103 Genes (95 ortholog groups)

Organism	
□ Eimeriidae	0
□ Sarcocystidae	103

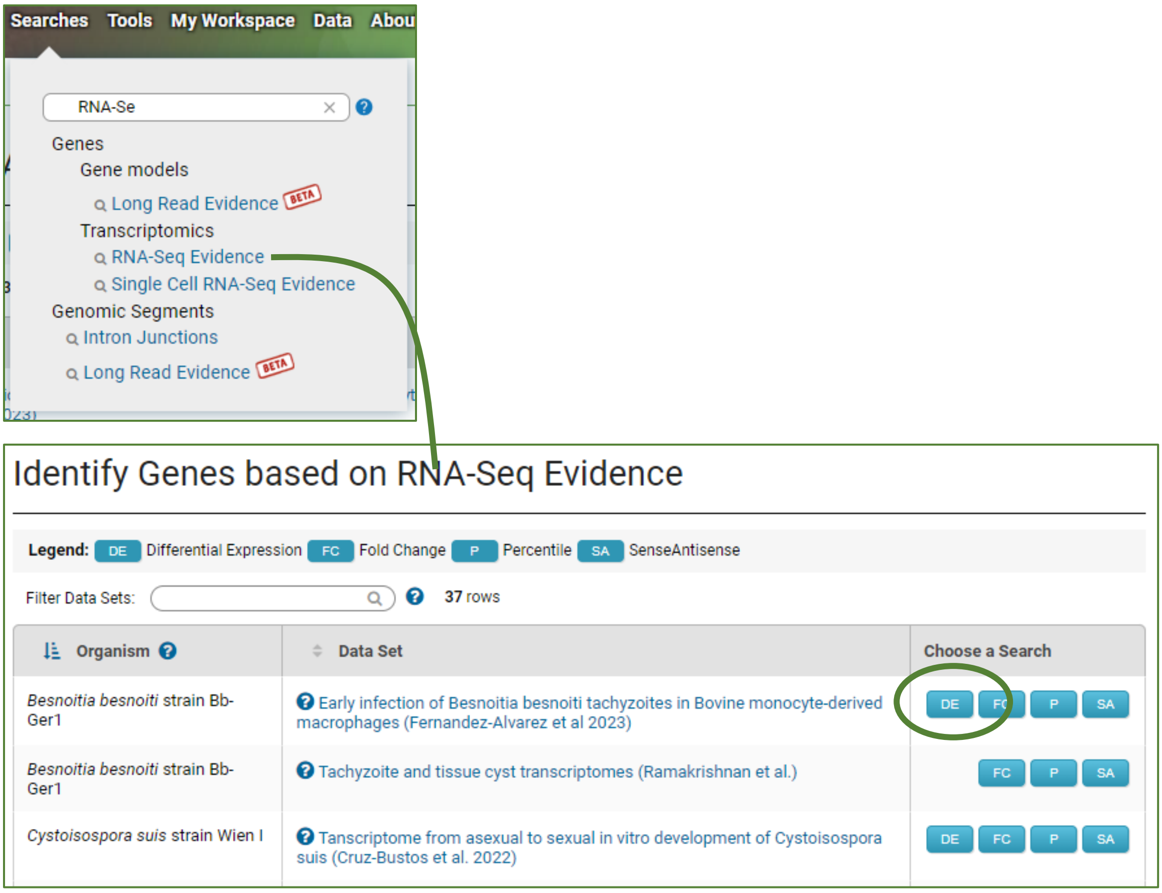
Rows per page: 500

Gene ID	Transcript ID	Genomic Location (Gene)	Product Description	Predicted Location (TAGM-MAP)
TGME49_214850	TGME49_214850.R834	TGME49_chrX:6,544,947..6,553,447(-)	1-deoxy-D-xylulose 5-phosphate reductoisomerase, putative	apicoplast
TGME49_208820	TGME49_208820.R657	TGME49_chrIb:875,353..886,622(-)	1-deoxy-D-xylulose-5-phosphate synthase	apicoplast
TGME49_255690	TGME49_255690.R2114	TGME49_chrVIIb:4,732,781..4,736,282(-)	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase domain-containing protein	apicoplast
TGME49_306260	TGME49_306260.R3518	TGME49_chrIX:5,899,584..5,903,274(+)	2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase, putative	apicoplast
TGME49_217740	TGME49_217740.R923	TGME49_chrXII:2,085,766..2,090,478(+)	3-ketoacyl-(acyl-carrier-protein) reductase	apicoplast
TGME49_293590	TGME49_293590.R3311	TGME49_chrIa:438,194..448,287(+)	3-oxoacyl-acyl-carrier protein synthase I/II, putative	apicoplast
TGME49_203420	TGME49_203420-t26_1	TGME49_chrVIIa:2,507,663..2,515,172(-)	4'-phosphopantetheinyl transferase domain-containing protein	apicoplast
TGME49_262430	TGME49_262430.R2403	TGME49_chrVIIb:1,250,160..1,260,134(+)	4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	apicoplast
TGME49_227420	TGME49_227420-t26_1	TGME49_chrX:946,913..952,751(-)	4-hydroxy-3-methylbut-2-enyl diphosphate reductase	apicoplast
TGME49_289230	TGME49_289230-t26_1	TGME49_chrIX:3,098,777..3,102,766(-)	50S ribosomal protein L12, putative	apicoplast

2. **Determine functional enrichment of the set of genes that are upregulated at least 2-fold in an early infection of *Besnoitia besnoiti* tachyzoites in Bovine monocyte-derived macrophage.** For this exercise use ToxoDB.

ToxoDB has RNA-sequence data from a study of the early interaction between *B. besnoiti* tachyzoites and primary bovine monocyte-derived macrophages *in vitro*. Data set record. Dual transcriptomic profiling of *B. besnoiti* tachyzoites and *Bos taurus* macrophages was conducted at early infection 4 h and 8 h post infection by high-throughput RNA sequencing. Bovine macrophages inoculated with heat-killed tachyzoites (MO-hkBb) and non-infected macrophages (MO) were used as controls. In this exercise we will find a list of genes that are differentially expressed based on the RNA Seq data and investigate the functional enrichment (if any) for the gene set.

a. Run a differential expression search looking for genes that are upregulated at least 2-fold at 8 hours post infection compared to 4 hours with p<0.001. Navigate to the RNA-seq searches in ToxoDB and choose the DE search for **Early infection of Besnoitia besnoiti tachyzoites in Bovine monocyte-derived macrophages (Fernandez-Alvarez et al 2023).**



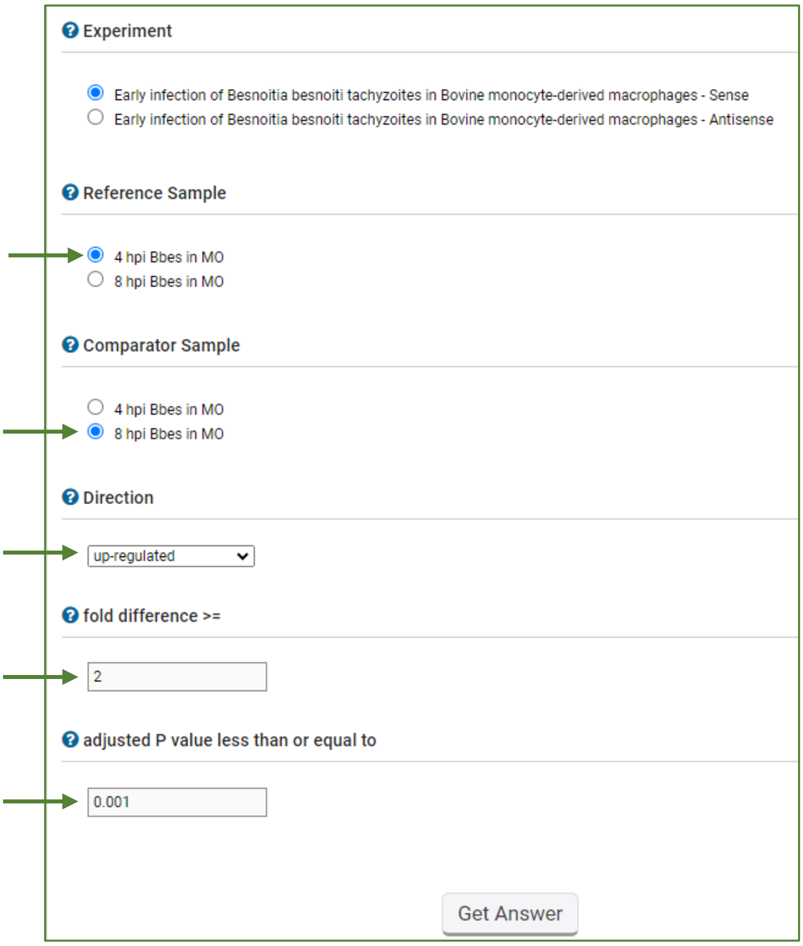
Identify Genes based on RNA-Seq Evidence

Legend: DE Differential Expression FC Fold Change P Percentile SA SenseAntisense

Filter Data Sets: 37 rows

Organism	Data Set	Choose a Search
Besnoitia besnoiti strain Bb-Ger1	Early infection of Besnoitia besnoiti tachyzoites in Bovine monocyte-derived macrophages (Fernandez-Alvarez et al 2023)	DE FC P SA
Besnoitia besnoiti strain Bb-Ger1	Tachyzoite and tissue cyst transcriptomes (Ramakrishnan et al.)	FC P SA
Cystoisospora suis strain Wien I	Tanscriptome from asexual to sexual in vitro development of Cystoisospora suis (Cruz-Bustos et al. 2022)	DE FC P SA

b. Arrange the search to return genes upregulated at least 2-fold (p<0.001) in the 8hr Bbes in MO sample compared to 4hr Bbes in MO.



Experiment

● Early infection of Besnoitia besnoiti tachyzoites in Bovine monocyte-derived macrophages - Sense
○ Early infection of Besnoitia besnoiti tachyzoites in Bovine monocyte-derived macrophages - Antisense

Reference Sample

● 4 hpi Bbes in MO
○ 8 hpi Bbes in MO

Comparator Sample

○ 4 hpi Bbes in MO
● 8 hpi Bbes in MO

Direction

up-regulated

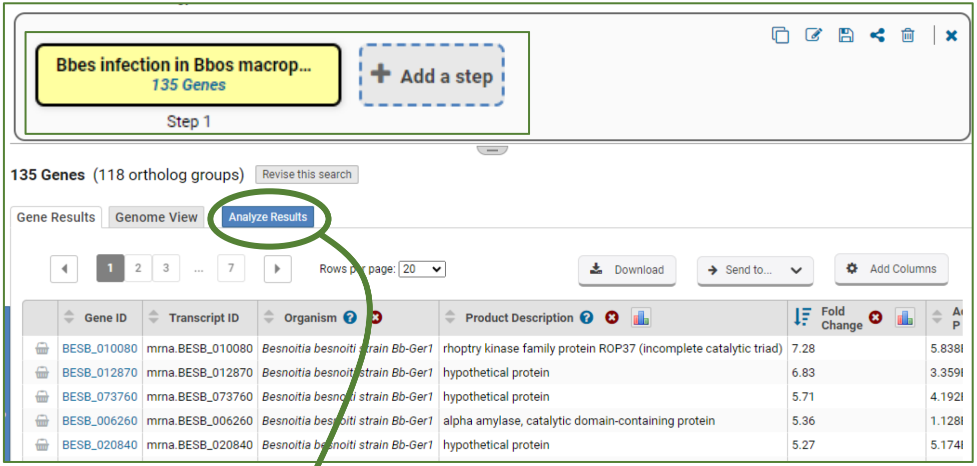
fold difference >=

2

adjusted P value less than or equal to

0.001

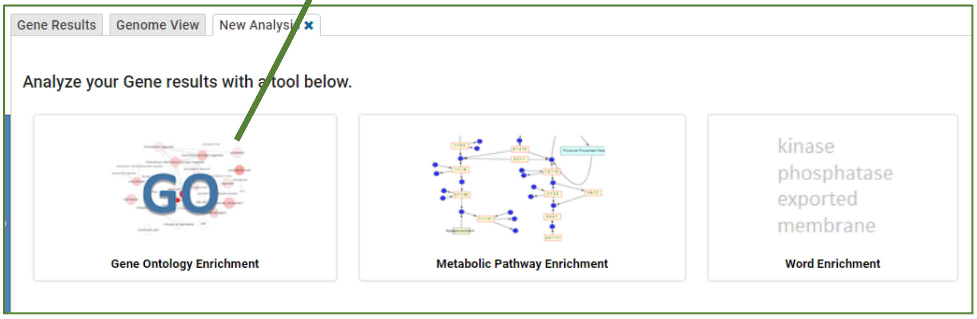
c. From the result page, choose the Analyze Results tab to go to the Enrichment Tool.



135 Genes (118 ortholog groups)

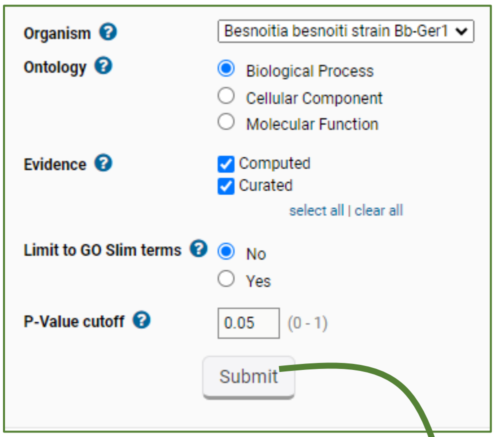
Rows per page: 20

	Gene ID	Transcript ID	Organism	Product Description	Fold Change	A P
	BESB_010080	mrna.BESB_010080	*Besnoitia besnoiti strain Bb-Ger1*	rhoptry kinase family protein ROP37 (incomplete catalytic triad)	7.28	5.838
	BESB_012870	mrna.BESB_012870	*Besnoitia besnoiti strain Bb-Ger1*	hypothetical protein	6.83	3.359
	BESB_073760	mrna.BESB_073760	*Besnoitia besnoiti strain Bb-Ger1*	hypothetical protein	5.71	4.192
	BESB_006260	mrna.BESB_006260	*Besnoitia besnoiti strain Bb-Ger1*	alpha amylase, catalytic domain-containing protein	5.36	1.128
	BESB_020840	mrna.BESB_020840	*Besnoitia besnoiti strain Bb-Ger1*	hypothetical protein	5.27	5.174



Analyze your Gene results with a tool below.

d. Run a GO enrichment analysis on the Biological Process ontology terms associated with your gene list.



Organism: Besnoitia besnoiti strain Bb-Ger1

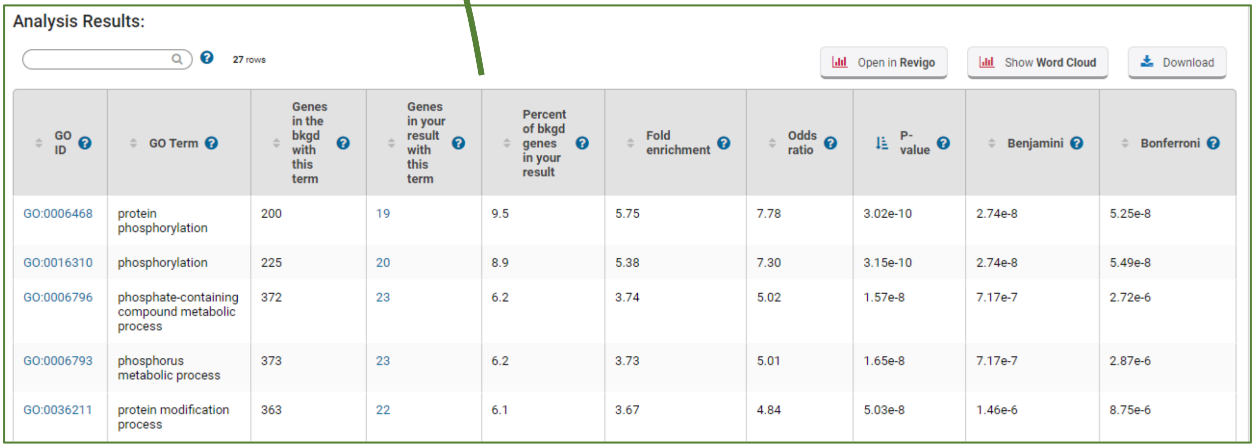
Ontology: ● Biological Process ○ Cellular Component ○ Molecular Function

Evidence: ☑ Computed ☑ Curated

select all | clear all

Limit to GO Slim terms: ● No ○ Yes

P-Value cutoff: 0.05 (0 - 1)



Analysis Results:

27 rows

GO ID	GO Term	Genes in the bkgd with this term	Genes in your result with this term	Percent of bkgd genes in your result	Fold enrichment	Odds ratio	P-value	Benjamini	Bonferroni
GO:0006468	protein phosphorylation	200	19	9.5	5.75	7.78	3.02e-10	2.74e-8	5.25e-8
GO:0016310	phosphorylation	225	20	8.9	5.38	7.30	3.15e-10	2.74e-8	5.49e-8
GO:0006796	phosphate-containing compound metabolic process	372	23	6.2	3.74	5.02	1.57e-8	7.17e-7	2.72e-6
GO:0006793	phosphorus metabolic process	373	23	6.2	3.73	5.01	1.65e-8	7.17e-7	2.87e-6
GO:0036211	protein modification process	363	22	6.1	3.67	4.84	5.03e-8	1.46e-6	8.75e-6

e. What is the top enriched GO term from this analysis? Does this make sense for an enrichment analysis of the biological processes associated with your differentially expressed genes? Notice that the p-value with Benjamini or Bonferroni correction is very low. What do each of the columns in the analysis table represent? Hint: move your mouse over the question mark next to each column header

- Fold enrichment -The ratio of the proportion of genes in the list of interest with a specific GO term over the proportion of genes in the background with that term.
- Odds ratio -The odds of the GO term appearing in the gene list are the same as that for the background list.
- P-value –The null hypothesis or the probability of getting a result that is equal or greater than what was observed.
- Benjamini-Hochburg false discovery rate – A method for controlling false discovery rates for type 1 errors.

f. Click on the link in the ‘Genes in your result with this term’ column. This will create a one-step strategy that returns only the genes with this GO term.

g. Perform the enrichment analysis on the other ontologies. Its possible to modify the paramaters of the current enrichment analysis but you will overwrite your current results. To run a new analysis and save the old, start with the Analyze Results tab.

h. Is there a cellular location or molecular function that is enriched? Do these support the biological process enrichments?