

Hint: analyze the Gene Ontology terms assigned to the genes in your result list.
The Result Analysis/Enrichment tool applies the Fischer's Exact test to compare your gene result to the entire genome. Use the Gene Ontology Enrichment to find Biological Process ontology terms that are enriched in your gene result.

The screenshot displays the Gene Ontology Enrichment tool. At the top, there are two tables showing gene counts for different assemblies. The 'Gene Ontology Enrichment' section is highlighted, showing a 'Submit' button circled in red. Below it, a 'GO' logo is also circled in red. The interface includes tabs for 'Gene Results', 'Genome View', and 'Gene Ontology Enrichment'.

- 1 c. Examine the filter table below the strategy section. Why is the number in the ortholog groups column less than the number in the Assemblage A column? Why do the other cells have a zero?

The screenshot shows the 'My Strategies' page with a 'Filter Table' for 265 genes from Step 1. The table has columns for Gene ID, Genomic Location, Product Description, Selected Samples that Meet Criteria, Sum of Unique Peptides (Within Samples), Unique Peptides (Across Samples), and Total Number of Spectra.

Gene ID	Genomic Location	Product Description	Selected Samples that Meet Criteria	Sum of Unique Peptides (Within Samples)	Unique Peptides (Across Samples)	Total Number of Spectra
GL50803_136020	GLCHR05: 3,625,697 - 3,627,040 (-)	Beta tubulin	3	234	177	2266
GL50803_112079	GLCHR03: 1,813,742 - 1,815,106 (-)	Alpha-tubulin	3	213	173	1674
GL50803_17230	GLCHR03: 483,217 - 484,152 (-)	Gamma giardin	3	175	136	1184

- 1 d. How many genes with mass spec data from the basal proteome experiment also have mass spec evidence from the “Mitosome enriched proteome (WB) (Jedelsk et al.)”?

The screenshot shows the 'Add Step' dialog for 'Mass Spec. Evidence'. The 'Experiment/Samples' section shows a tree view with 'Giardia Assemblage A' expanded, and 'Mitosome enriched proteome (WB) (Jedelsk et al.)' selected. The 'Minimum Number of Unique Peptide Sequences' and 'Minimum Number of Spectra' are both set to 1. The 'Combine Genes in Step 1 with Genes in Step 2' section shows 'Intersect 2' selected.

- 1 e. The default parameters of the mass spec search are set to identify any gene with at least 1 mapped peptide identified. How will your results change if you **revise both steps** in your search strategy to only **return genes with at least 5 peptides identified**?

The screenshot shows the 'Revise Step' dialog for 'Mass Spec. Evidence'. The 'Minimum Number of Unique Peptide Sequences' parameter is changed from 1 to 5. The 'Results' section shows 265 Genes. The 'Revise Step' dialog also shows the 'Minimum Number of Unique Peptide Sequences' parameter changed from 1 to 5.

Any step in a strategy can be revised

2. Find genes in *Cryptosporidium* that have mass spec evidence from any of the sporozoite proteomics experiments available in CryptoDB.
For this exercise use <http://cryptodb.org>
- 2 a. Explore the available proteomics data and select samples that make sense. You may need to click on the '+' sign to expand experiments to see the underlying samples.

Identify Genes based on Mass Spec. Evidence

Experiment/Samples [select all](#) | [clear all](#) | [expand all](#) | [collapse all](#) | [reset to default](#)

- [-] **Cryptosporidium**
 - [-] **Cryptosporidium parvum**
 - Enriched cytoskeletal and membrane fractions (Madrid-Aliste et al.)
 - Mitochondrial Fraction Proteomics (Putignani)
 - Oocyst Wall Proteome (Iowall) (Ferrari)**
 - Intact Oocysts
 - Oocyst walls
 - Sporozoites
 - Proteome during Sporozoite Excystation (ISSC162) (Snelling et al.)**
 - Insoluble Excysted Fraction LC-MS/MS
 - Insoluble Non-excysted Fraction LC-MS/MS
 - Soluble Excysted and Non-excysted Fraction LC-MS/MS
 - Sporozoite Proteome (Iowall) (Sanderson et al.)**
 - 1D Gel LC-MS/MS
 - 2D Gel LC-MS/MS
 - MudPit Insoluble fractions
 - MudPit Soluble fractions

[select all](#) | [clear all](#) | [expand all](#) | [collapse all](#) | [reset to default](#)

Minimum Number of Unique Peptide Sequences

Minimum Number of Spectra

[-] **Advanced Parameters**

[Get Answer](#)

- 2 b. Remove any gene with peptide evidence from non-sporozoite samples
Hint: add a step for mass spec data and think about how you will combine your results.

Add Step 2 : Mass Spec. Evidence

Experiment/Samples [select all](#) | [clear all](#) | [expand all](#) | [collapse all](#) | [reset to default](#)

- [-] **Cryptosporidium**
 - [-] **Cryptosporidium parvum**
 - Enriched cytoskeletal and membrane fractions (Madrid-Aliste et al.)
 - Mitochondrial Fraction Proteomics (Putignani)
 - Oocyst Wall Proteome (Iowall) (Ferrari)**
 - Intact Oocysts
 - Oocyst walls
 - Sporozoites
 - Proteome during Sporozoite Excystation (ISSC162) (Snelling et al.)**
 - Insoluble Excysted Fraction LC-MS/MS
 - Insoluble Non-excysted Fraction LC-MS/MS
 - Soluble Excysted and Non-excysted Fraction LC-MS/MS
 - Sporozoite Proteome (Iowall) (Sanderson et al.)**
 - 1D Gel LC-MS/MS
 - 2D Gel LC-MS/MS
 - MudPit Insoluble fractions
 - MudPit Soluble fractions

[select all](#) | [clear all](#) | [expand all](#) | [collapse all](#) | [reset to default](#)

Minimum Number of Unique Peptide Sequences

Minimum Number of Spectra

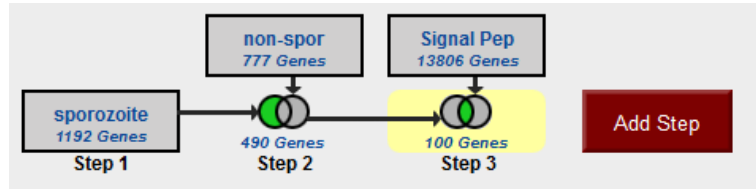
[-] **Advanced Parameters**

Combine Genes in Step 1 with Genes in Step 2:

1 Intersect 2
 1 Minus 2
 1 Union 2
 2 Minus 1
 1 Relative to 2, using genomic collocation

[Run Step](#)

2 c. How many of these genes are also predicted to be secreted?



2 d. So far you have been searching for *C. parvum* genes because we only have proteomics data from this species. However, what if you are studying *C. muris*? How can you garner information about the protein expression of *C. muris* genes from your *C. parvum* results? (Hint: add a step then select the “Transform by Orthology” option).

- Did the number of *C. parvum* genes increase or decrease? Why?

All Results	Ortholog Groups	Apicomplexa			Chromerida		
		Cryptosporidium			Gregarina	Chromera	Vitrella
		<i>C.hominis</i>	<i>C.muris</i>	<i>C.parvum</i>	<i>G.niphandrodes</i>	<i>C.vella</i>	<i>V.brassiciformis</i>
399	100	101	75	109	26	44	44

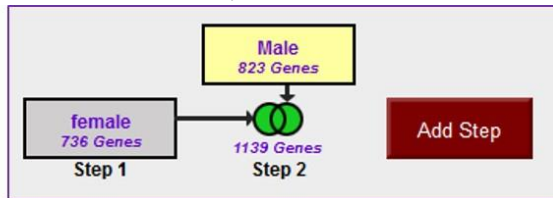
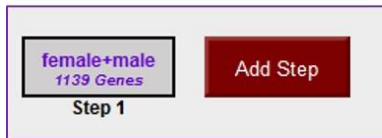
3 Finding all genes with mass spec evidence in *T. cruzi*.
For this exercise use <http://TriTrypDB.org>

3 a. How many genes in *T. cruzi* have expression evidence based on mass spec data?

4. Finding genes with evidence for protein level expression in *P. berghei* gametocytes.
For this exercise use <http://plasmodb.org>

- 4 a.** Find all *P. berghei* genes that have mass spec evidence in either or both male and female gametocytes.
- What proteomics experiment and samples did you search? How many genes did you get? How did you get to this number?
 - Try running this search in two different ways:
 - (i) Select both male and female gametocyte options and run the search.
 - (ii) Select one of them first, run the search then add the other one using the add step button. How did you combine the two steps? Do you get the same results as in (i)?

Hint: When using a two-step search to find genes that have either or both characteristics, take the union of the two searches, not the intersection.



- 4 b.** Find all genes that have mass spec evidence in both male and female gametocytes. Hint: use the strategy you developed in (ii) to get this answer, but change the union into an intersection.
- 4 c.** Find genes that have mass spec evidence only in male gametocytes and not in female ones. Hint: modify the set operation in b.
- 4 d.** Find genes that have mass spec evidence only in female gametocytes and not in male ones. Hint: modify the set operation in b.
- 4 e.** Which female gametocyte gene has the highest number of peptide sequences? Focus on the female gametocyte search step. Hint: look at the “Sum of Unique Peptides (Within Samples)” column in the list of results.

The screenshot shows the Plasmodb search interface. At the top, there are tabs for 'My Strategies: New, Opened (3), All (3), Basket, Public Strategies (8), Help'. Below this, a search strategy is shown with a box for 'female' (736 Genes) and a box for 'Male' (823 Genes) connected by a plus sign, resulting in '318 Genes Step 2'. A red circle highlights the 'female' box. Below the strategy is a table of results for '736 Genes from Step 1'. The table has columns for 'All Results', 'Ortholog Groups', 'Pberghi', 'Pchabaudi', 'Pcynomolgi', 'Pfalciiparum', 'Pgalinaceum', 'Pknowlesi', 'Preichenoai', and 'Pvivax'. The first row shows '736' results. Below the table is a 'Gene Results' section with a table of results. The table has columns for 'Gene ID', 'Genomic Location', 'Product Description', 'Selected Samples that Meet Criteria', 'Sum of Unique Peptides (Within Samples)', 'Unique Peptides (Across Samples)', and 'Total Number of Spectra'. The first row shows 'PBAfKA_145300' with a sum of 48 unique peptides. A red circle highlights the 'Add Columns' button in the bottom right corner. A 'Select Columns' dialog box is open on the right, showing options for 'Search-Specific', 'Selected Samples that Meet Criteria', 'Unique Peptides (Across Samples)', 'Sum of Unique Peptides (Within Samples)', 'Total Number of Spectra', 'Text, IDs, Species', 'Genomic Sequence ID', 'Organism', 'UniProt ID', 'Entrez Gene ID', and 'Gene Name or Symbol'. A red arrow points from the 'Sum of Unique Peptides (Within Samples)' option in the dialog to the corresponding column in the table.

Gene ID	Genomic Location	Product Description	Selected Samples that Meet Criteria	Sum of Unique Peptides (Within Samples)	Unique Peptides (Across Samples)	Total Number of Spectra
PBAfKA_145300	berg14 2,327,865 - 2,335,685 (-)	osmiophilic body protein (G377)	1	48	48	48
PBAfKA_071190	berg07 420,169 - 422,250 (-)	heat shock protein, putative (HSP70)	1	34	34	34
PBAfKA_080570	berg08 266,690 - 271,847 (+)	heat shock protein 90, putative (HSP90)	1	33	33	33

4 f. What does the distribution of peptides in the gene from 'e' look like? Hint: go to the gene page and look at the "Protein features" section, or go to the genome browser from the gene page and turn on the right tracks.

5. Find genes with evidence of phosphorylation in intracellular *Toxoplasma* tachyzoites. For this exercise use <http://www.toxodb.org>

Phosphorylated peptides can be identified by searching the appropriate experiments in the Mass Spec Evidence search page.

5 a. Find all genes with evidence of phosphorylation in intracellular tachyzoites. Select the "Infected host cell, phosphopeptide-enriched (peptide discovery against TgME49)" sample under the experiment called "Tachyzoite phosphoproteome from purified parasite or infected host cell (RH) (Treeck et al.)"

Identify Genes based on Mass Spec. Evidence

Experiment/Samples ? [select all](#) | [clear all](#) | [expand all](#) | [collapse all](#) | [reset to default](#)

- Eimeria
- Toxoplasma
 - Toxoplasma gondii*
 - Oocyst Partially Sporulated Proteome (VEG) (Possenti, et al.)
 - Oocyst proteome (M4 Typell) (Wastling)
 - Oocyst proteome - Fractionated (M4 type II) (Fritz et al.)
 - Proteome During Infection in *H. sapiens* (Wastling)
 - Tachyzoite Intra- and Extracellular Lysine-Acetylomes (RH) (Jeffers and Xue)
 - Tachyzoite Rhoptyr proteome (RH) (Bradley et al.)
 - Tachyzoite conoid proteome (RH) (Hu et al.)
 - Tachyzoite membrane and cytosolic fractions (RH) (Dybas et al.)
 - Tachyzoite phosphoproteome - Calcium dependent (RH) (Nebl et al.)
 - Tachyzoite phosphoproteome from purified parasite or infected host cell (RH) (Treeck et al.)
 - Infected host cell, phosphopeptide-depleted (peptide discovery against TgME49)
 - Infected host cell, phosphopeptide-depleted (peptide discovery against TgGT1)
 - Infected host cell, phosphopeptide-enriched (peptide discovery against TgME49)
 - Infected host cell, phosphopeptide-enriched (peptide discovery against TgGT1)
 - Purified tachyzoites phosphopeptide-depleted (peptide discovery against TgGT1)
 - Purified tachyzoites phosphopeptide-depleted (peptide discovery against TgME49)
 - Purified tachyzoites phosphopeptide-enriched (peptide discovery against TgGT1)
 - Purified tachyzoites phosphopeptide-enriched (peptide discovery against TgME49)
 - Tachyzoite secretome (RH) (Zhou et al.)
 - Tachyzoite subcellular fractions (Moreno)
 - Tachyzoite total proteome (RH) (Wastling)

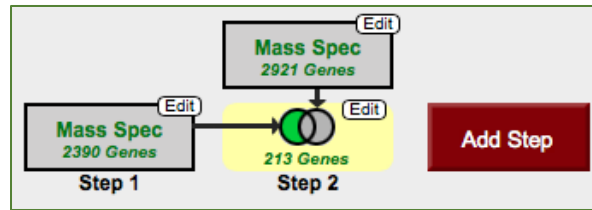
[select all](#) | [clear all](#) | [expand all](#) | [collapse all](#) | [reset to default](#)

Minimum Number of Unique Peptide Sequences ?

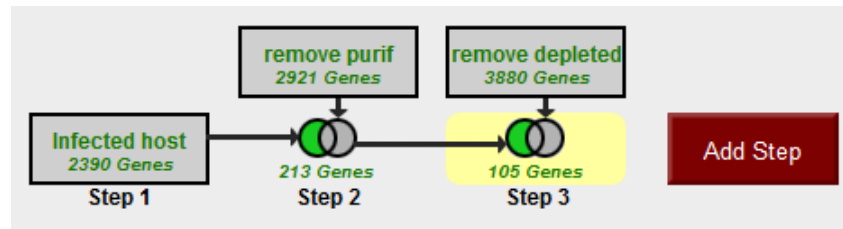
Minimum Number of Spectra ?

[Advanced Parameters](#)

5 b. Remove all genes with phosphorylation evidence from purified tachyzoites.



5 c. Remove all genes that are also present in the phosphopeptide-depleted fractions (select both intracellular and extracellular).



5 d. Explore your results. What kinds of genes did you find? Hint: use the Product description word column or perform a GO enrichment analysis of your results.

5 e. Are any of these genes likely to be secreted? Hint: add a step searching for genes with secretory signal peptides.

My Strategies: [New](#) [Opened \(3\)](#) [All \(3\)](#) [Basket](#) [Public Strategies \(7\)](#) [Help](#)

(Genes)

22 Genes from Step 4
Strategy: *Infected host*

Click on a number in this table to limit/filter your results

All Results	Ortholog Groups	<i>Eimeria</i>									<i>Hammondia</i>	<i>N.</i>
		<i>E.acervulina</i>	<i>E.brunetti</i>	<i>E.falciformis</i>	<i>E.maxima</i>	<i>E.mitis</i>	<i>E.necatrix</i>	<i>E.praecox</i>	<i>E.tenella</i>	<i>H.hammondi</i>	<i>N.</i>	
		Houghton	Houghton	Bayer HaberKorn 1970	Weybridge	Houghton	Houghton	Houghton	Houghton	strain Houghton	strain H.H.34	L
22	22	0	0	0	0	0	0	0	0	0	0	

Filter by strains (advanced)

Gene Results [Genome View](#) [Analyze Results](#) **BETA**

First 1 2 Next Last [Advanced Paging](#)

Gene ID	Gene Group (representative gene)	Genomic Location	Product Description
TGME49_294940	TGGT1_294940	TGME49_chrla: 1,282,608 - 1,287,925 (-)	hypothetical protein
TGME49_222870	TGGT1_222870	TGME49_chrl1: 1,271,864 - 1,275,140 (+)	hypothetical protein
TGME49_320150	TGGT1_320150	TGME49_chrlV: 464,394 - 473,129 (-)	elongation factor Tu GTP binding domain-containing protein

5 f. Pick one or two of the hypothetical genes in your results and visit their gene pages. Can you infer anything about their function? Hint: explore the protein and expression sections.

- 5 g. What about polymorphism data? Go back to your strategy and add columns for SNP data found under the population biology section. Explore the gene page for the gene that has the most number of non-synonymous SNPs. Hint: you can sort the columns by clicking on the up/down arrows next to the column names.

Gene Results Genome View Analyze Results SEIA

First 1 2 Next Last Advanced Paging Add Columns

Gene ID	Product Description	Total SNPs All Strains	NonSynonymous SNPs All Strains	Synonymous SNPs All Strains	Non-Coding SNPs All Strains	SNPs with Stop Codons All Strains	NonSyn/Syn SNP Ratio All Strains
TGME49_271110	hypothetical protein	890	157	44	679	10	3.57
TGME49_257595	hypothetical protein	317	123	51	131	12	2.41
TGME49_219640	hypothetical protein	382	85	34	263	0	2.5
TGME49_288370	hypothetical protein	224	82	35	105	2	2.34
TGME49_216840	hypothetical protein	189	75	23	89	2	3.26
TGME49_257640	hypothetical protein	110	66	12	31	1	5.5
TGME49_320150	elongation factor Tu GTP binding domain-containing protein	378	65	22	286	5	2.95
TGME49_235960	hypothetical protein	155	58	14	77	6	4.14
TGME49_288880	hypothetical protein	220	56	17	147	0	3.29
TGME49_269750	CrcB family protein	95	54	20	18	3	2.7
TGME49_315700	hypothetical protein	338	54	14	265	5	3.86
TGME49_308070	hypothetical protein	188	43	22	123	0	1.95
TGME49_269420	hypothetical protein	45	37	8	0	0	4.63
TGME49_200440	hypothetical protein	72	35	11	24	2	3.18
TGME49_259830	diacylglycerol kinase catalytic domain-containing protein	176	32	3	139	2	10.67
TGME49_236220	PCI domain-containing protein	383	28	18	332	5	1.56
TGME49_231180	hypothetical protein	54	25	9	18	2	2.78
TGME49_294940	hypothetical protein	137	16	7	111	3	2.29