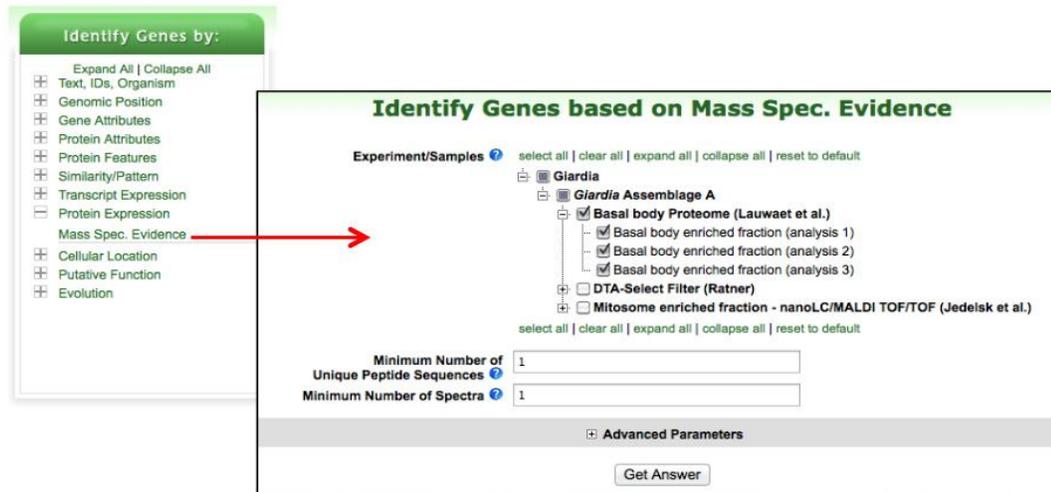


Proteomics

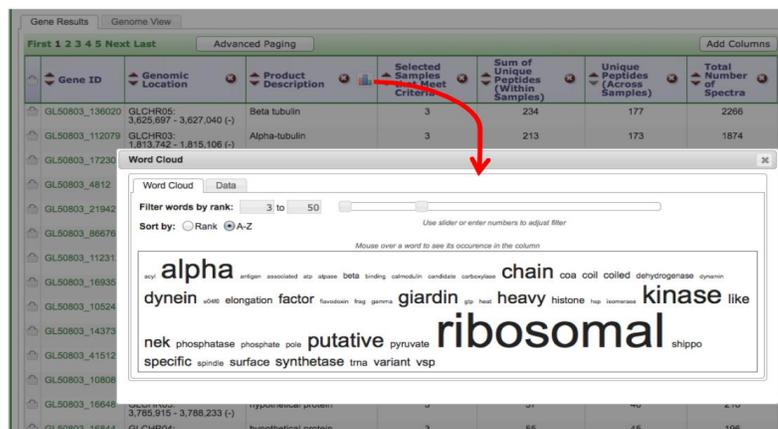
1. Find all *Giardia* genes with evidence of basal body expression based on mass spec/proteomics data. For this exercise use <http://giardiadb.org>

The proteomics search is available under the heading “Protein Expression” in the “Identify gene by” section.



- a. How many genes did you identify?
- b. What kinds of genes are in your result list? Are there genes with similar functions?
Hint: analyze the Product Description column.

The **word cloud tool** counts the number of times a word appears in the column and then draws a word cloud in which the size of the word reflects how many times the word appears in the product description column. Click on the little graphic icon next to the column called “Product description”.



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.

265 Genes from Step 1
Strategy: Mass Spec

All Results	Ortholog Groups	G.Assemblage A2	G.Assemblage A	G.Assemblage B	G.Assemblage E
265	250	0	265	0	0

Assemblage A Genes

Not Deprecated	Deprecated
265	0

Gene Results Genome View **Analyze Results** BETA

Gene ID Genomic Location Product Description

GL50803_136020	GLCHR05: 3,625,897 - 3,627,040 (-)	Beta tubulin
GL50803_112079	GLCHR03: 1,813,742 - 1,815,106 (-)	Alpha tubulin
GL50803_17230	GLCHR03: 483,217 - 484,152 (-)	Gamma giardin

Gene Ontology Enrichment

Parameters

Organism: Giardia Assemblage A isolate WB

Ontology: Cellular Component Molecular Function Biological Process

GO Association Sources: Select all Clear all InterPro predictions

P-Value Cutoff ($\alpha = 1.0$): 0.05

Submit

- 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?

My Strategies: New Opened (1) All (38) Basket Public Strategies (5) Help

Strategy: Mass Spec * 265 Genes

Mass Spec 265 Genes **Add Step**

Step 1

265 Genes from Step 1
Strategy: Mass Spec

Add 265 Genes to Basket | Download 265 Genes

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

All Results	Ortholog Groups	G.Assemblage A2	G.Assemblage A	G.Assemblage B	G.Assemblage E
265	250	0	265	0	0

Assemblage A Genes

Not Deprecated	Deprecated
265	0

Gene Results Genome View **Analyze Results** BETA

First 1 2 3 4 5 Next Last **Advanced Paging** Add Columns

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: 3,625,897 - 3,627,040 (-)	Beta tubulin	3	234	177	2266
GL50803_112079	GLCHR03: 1,813,742 - 1,815,106 (-)	Alpha-tubulin	3	213	173	1874
GL50803_17230	GLCHR03: 483,217 - 484,152 (-)	Gamma giardin	3	175	136	1184

Filter Table

- 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 a strategy editor interface. At the top, a menu bar includes 'My Strategies: New, Opened (1), All (38), Basket, Public Strategies (5), Help'. Below this, a 'Strategy: Mass Spec *' window shows a list of search criteria: Genes, Text, IDs, Organism, Genomic Segments, Genomic Position, ORFs, Gene Attributes, Protein Attributes, Protein Features, Similarity/Pattern, and Transcript Expression. A red arrow points from the 'Add Step' button in the strategy editor to the 'Add Step' dialog box. The 'Add Step' dialog box is titled 'Add Step 2 : Mass Spec. Evidence' and shows a tree view of 'Giardia' samples, including 'Basal body proteome (WB) (Lauwaet et al.)' and 'Mitosome enriched proteome (WB) (Jedelsk et al.)'. Below the tree, there are input fields for 'Minimum Number of Unique Peptide Sequences' (set to 1) and 'Minimum Number of Spectra' (set to 1). At the bottom, there are options to 'Combine Genes in Step 1 with Genes in Step 2:' with radio buttons for 'Intersect 2', 'Union 2', '1 Minus 2', '2 Minus 1', and '1 Relative to 2, using genomic colocation'. A 'Run Step' button is at the bottom right.

- 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 same strategy editor interface as in part d, but with a 'Revise' dialog box open. The 'Revise' dialog box is titled 'STEP 1 : Mass Spec' and shows the current search parameters: 'Experiment/Samples : Basal body Proteome (Lauwaet et al.), Basal body enriched fraction (analysis 1), Basal body enriched fraction (analysis 2), Basal body enriched fraction (analysis 3)', 'Minimum Number of Unique Peptide Sequences : 1', and 'Minimum Number of Spectra : 1'. Below this, there are input fields for 'Minimum Number of Unique Peptide Sequences' (set to 5) and 'Minimum Number of Spectra' (set to 1). A red arrow points from the 'Revise' button in the strategy editor to the 'Revise' dialog box. The 'Revise' dialog box also shows a tree view of 'Giardia' samples, including 'Basal body Proteome (Lauwaet et al.)', 'Basal body enriched fraction (analysis 1)', 'Basal body enriched fraction (analysis 2)', 'Basal body enriched fraction (analysis 3)', 'DTA-Select Filter (Ratner)', and 'Mitosome enriched fraction - nanoLC/MALDI TOF/TOF (Jedelsk et al.)'. Below the tree, there are input fields for 'Minimum Number of Unique Peptide Sequences' (set to 5) and 'Minimum Number of Spectra' (set to 1). A 'Run Step' button is at the bottom right. The text 'Any step in a strategy can be revised' is written in the middle of the image.

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>
- 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

- 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

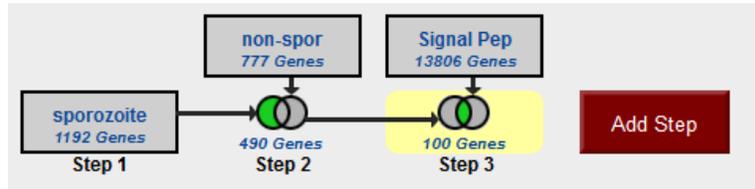
Minimum Number of Spectra

Advanced Parameters

Combine Genes in Step 1 with Genes in Step 2:

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

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



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?

399 Genes from Step 4
Strategy: sporozoite

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

All Results	Ortholog Groups	Apicomplexa					Chromerida		
		Cryptosporidium			Gregarina	Chromera	Vitrella		
		C.hominis	C.muris	C.parvum	G.niphandrodes	C.velia	V.brassicaformis		
399	100	TU502	RN66	Iowa II	Unknown strain	CCMP2878	CCMP3155	44	44

3. Finding all genes with mass spec evidence in *T. cruzi*.

For this exercise use <http://TriTrypDB.org>

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

Identify Genes based on Mass Spec. Evidence

Experiment/Samples select all | clear all | expand all | collapse all | reset to default

- Leishmania
- Trypanosoma
 - Trypanosoma brucei*
 - Trypanosoma cruzi*
 - Epimastigote Cell Surface Proteome (Berenice) (Queiroz et al.)
 - Life cycle proteome (Brazil) (Atwood et al.)
 - Membrane Proteins from insect developmental forms (strain G) (Cordero et al.)
 - Phosphoproteome during Metocyclogenesis (Dm28c) (Marchini et al.)
 - Reservosomes (Dm28c) (Sant'Anna et al.)
 - SUMOylation enriched epimastigote proteome (CL Brener) (Bayona et al.)

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

b. How many genes from the results in 'a' have at least 10 uniquely mapped peptides? Hint: try revising the step in 'a' and change the "minimum number of unique peptide sequences" option to 10.

Mass Spec. Evidence
7892 Genes
Step 1

Edit

Revise Step

Revise Step 1 : Mass Spec. Evidence

Experiment/Samples select all | clear all | expand all | collapse all | reset to default

- Leishmania
- Trypanosoma
 - Trypanosoma brucei*
 - Trypanosoma cruzi*

select all | clear all | expand all | collapse all | reset to default

Minimum Number of Unique Peptide Sequences

Minimum Number of Spectra

Advanced Parameters

Run Step

c. Expand the list of results in 'b' to include possible orthologs/paralogs in *T. cruzi*.

Hint: use the ortholog transform option when adding a step and select only *T. cruzi*. Explore the columns in your result set. Pay close attention to the organism filter table.

My Strategies: [New](#) [Opened \(1\)](#) [All \(1\)](#) [Basket](#) [Public Strategies \(14\)](#) [Help](#)

(Genes) Strategy: Mass Spec

Mass Spec
983 Genes
Step 1 →
 Orthologs
5008 Genes
Step 2
Add Step

5008 Genes from Step 2 Add 5008 Genes to Basket

Strategy: Mass Spec

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

Trypanosoma												
.tropica		L.turana		T.brucei (nr Genes: 0)			T.congolense		T.cruzi (nr Genes: 4409)			
L590	strain LEM423	Lister strain 427	TREU927	gambiense DAL972	IL3000	CL Brener Esmeraldo-like	CL Brener Non-Esmeraldo-like	strain CL Brener	Dm28c	Sylvio X10/1	marinkellei strain B7	
0	0	0	0	0	0	872	949	214	942	1195	836	

Gene Results Genome View Analyze Results BETA

First 1 2 3 4 5 Next Last Advanced Paging

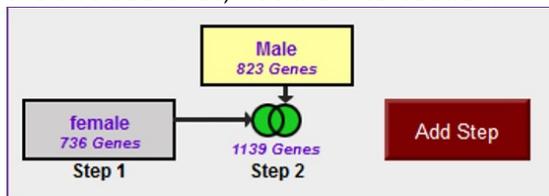
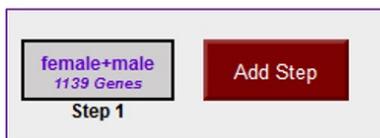
Gene ID	Organism	Genomic Location	Product Description	Input Ortholog(s)	Ortholog Group	Paralog count
TCDM_00399	<i>T. cruzi</i> Dm28c	AYLP01000002: 297,193 - 298,563 (-)	retrotransposon hot spot (RHS) protein	TcCLB.410923.20, TcCLB.459199.10, TcCLB.463155.20, TcCLB.503483.9, TcCLB.503607.4,	OG5_126555	23

4. Finding genes with evidence for protein level expression in *P. berghei* gametocytes.

For this exercise use <http://plasmodb.org>
<http://www.plasmodb.org/>

- 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.



- 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.
- Find genes that have mass spec evidence only in male gametocytes and not in female ones. Hint: modify the set operation in b.
- Find genes that have mass spec evidence only in female gametocytes and not in male ones. Hint: modify the set operation in b.
- 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 a search interface with a strategy diagram at the top. The strategy involves a 'female' step (736 Genes) and a 'male' step (823 Genes) combined via an intersection operation. Below the strategy is a table of 736 genes from Step 1, with columns for various Plasmodium species. A detailed 'Gene Results' table is shown below, with columns for Gene ID, Genomic Location, Product Description, Selected Samples, Sum of Unique Peptides (Within Samples), Unique Peptides (Across Samples), and Total Number of Spectra. The 'Sum of Unique Peptides (Within Samples)' column is highlighted with a red circle and an arrow pointing to the 'Add Columns' button in the table's header. The 'Add Columns' button is also circled in red.

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
PBANKA_146300	berg14: 2,327,865 - 2,335,685 (-)	osmiophilic body protein (G377)	1	48	48	48
PBANKA_071190	berg07: 420,169 - 422,250 (-)	heat shock protein, putative (HSP70)	1	34	34	34
PBANKA_080570	berg08: 268,690 - 271,847 (+)	heat shock protein 90, putative (HSP90)	1	33	33	33

- 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.

- 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 TypeII) (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 Rhostry 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) (Nebi 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](#)

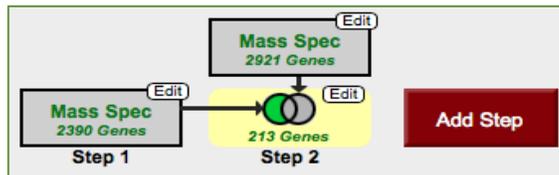
Minimum Number of Unique Peptide Sequences

Minimum Number of Spectra

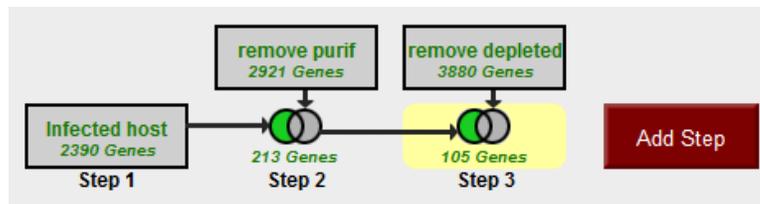
[Advanced Parameters](#)

[Get Answer](#)

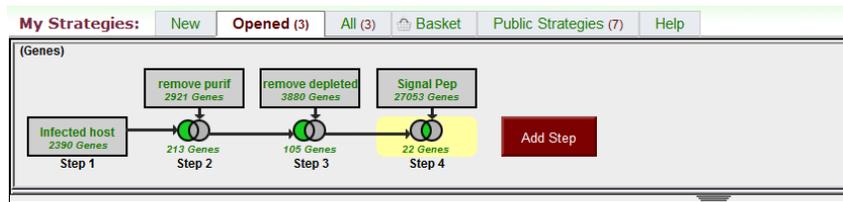
- b. Remove all genes with phosphorylation evidence from purified tachyzoites.



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



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



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> Houghton	<i>E.brunetti</i> Houghton	<i>E.falciformis</i> Bayer Haber Korn 1970	<i>E.maxima</i> Weybridge	<i>E.mitis</i> Houghton	<i>E.necatrix</i> Houghton	<i>E.praecox</i> Houghton	<i>E.tenella</i> strain Houghton	<i>H.hammondi</i> strain H.H.34	<i>N.</i>	
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_chrlfa: 1,282,608 - 1,287,925 (-)	hypothetical protein
TGME49_222870	TGGT1_222870	TGME49_chrlf: 1,271,864 - 1,275,140 (+)	hypothetical protein
TGME49_320150	TGGT1_320150	TGME49_chrlfv: 464,394 - 473,129 (-)	elongation factor Tu GTP binding domain-containing protein

- 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.
- 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** BETA

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