

# Exploring Proteomics Evidence

## Exercise 12

### 11.1 Exploring proteomics data in TriTrypDB.

- a. How many organisms have mass spectrometry evidence in TriTrypDB?  
Hint: take a look at the “Mass Spec. Evidence section”, under protein expression.

The image shows two panels from the TriTrypDB interface. The left panel, titled "Identify Genes by:", lists various search criteria: Expand All | Collapse All, Text, IDs, Organism, Genomic Position, Gene Attributes, Protein Attributes, Protein Features, Similarity/Pattern, Transcript Expression, Protein Expression, and Mass Spec. Evidence. The "Mass Spec. Evidence" option is circled in red. A red arrow points from this option to the right panel. The right panel, titled "Identify Genes based on Mass Spec. Evidence", shows a search tree for "Experiment/Samples". The tree is partially expanded, showing "Leishmania braziliensis" with a checked box, and "Promastigote secretome (Cuervo et al.)" with a checked box. Below the tree are input fields for "Minimum Number of Unique Peptide Sequences" and "Minimum Number of Spectra", both set to 1. There are also checkboxes for "Give this search a weight" and "Give this search a name", and a "Get Answer" button.

- b. What kinds of experiments and parasite stages are represented?  
Hint: expand species and experiments by clicking on the plus signs.

The image shows the same "Identify Genes based on Mass Spec. Evidence" search results panel as in the previous screenshot, but with the search tree fully expanded. A red box labeled "Expand" is positioned to the left of the tree. Several plus signs in the tree are circled in red, indicating where the user clicked to expand the tree. The expanded tree shows "Leishmania braziliensis" with a checked box, and "Promastigote secretome (Cuervo et al.)" with a checked box. Below this, there are several other species and stages, each with a plus sign and a checked box: "secretome", "Leishmania donovani", "Leishmania infantum", "Leishmania mexicana", "Amastigote secretome (Paape et al.)", "Trypanosoma brucei", "Trypanosoma cruzi", "Insect form membrane proteins (Cordero et al.)", "Life cycle stages (Attwood et al.)", and "Reservosome proteome (Sant'Anna)". The "Life cycle stages" section is further expanded, showing sub-categories like "amastigote, esmeraldo-like", "amastigote, nonesmeraldo-like", "epimastigote, esmeraldo-like", "epimastigote, nonesmeraldo-like", "metacyclic trypomastigote, esmeraldo-like", "metacyclic trypomastigote, nonesmeraldo-like", "trypomastigote, esmeraldo-like", and "trypomastigote, nonesmeraldo-like". Below the tree are the same input fields for "Minimum Number of Unique Peptide Sequences" and "Minimum Number of Spectra", both set to 1. There are also checkboxes for "Give this search a weight" and "Give this search a name", and a "Get Answer" button.

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

### a. How many genes in *T. cruzi* have mass spec evidence?

Hint: select *Trypanosoma cruzi* from the Mass Spec experiment list you explored in 10.1.

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

The screenshot shows the workflow interface. In the 'My Strategies' panel, 'Mass Spec' is selected with 2990 genes. The 'STEP 1: Mass Spec' configuration shows 'Experiment/Samples' with 'Trypanosoma cruzi' selected and 'Minimum Number of Unique Peptide Sequences' set to 1. The 'Revise Step' dialog shows the same configuration but with 'Minimum Number of Unique Peptide Sequences' changed to 10.

### c. Can you expand the list of results in 'b' to include possible paralogs in *T. cruzi*?

Hint: you will have to use the ortholog transform option when adding a step and select only *T. cruzi*. Explore the columns in your result set.

The screenshot shows the workflow interface with the 'Orthologs' step added. The 'My Step Result' table shows 1540 genes. The table has columns for Gene ID, Organism, Genomic Location, Product Description, Input Ortholog, Ortholog Group, Paralog count, and Ortholog count.

Gene ID	Organism	Genomic Location	Product Description	Input Ortholog	Ortholog Group	Paralog count	Ortholog count
LbrM.27.2620	<i>L. braziliensis</i>	LbrM.27.1,029,884 - 1,030,999 (+)	aldo-keto reductase-like protein	Tc00.1047053511287.49	OG5_126583	3	29
LbrM.31.2410	<i>L. braziliensis</i>	LbrM.31.1,114,033 - 1,114,887 (-)	prostaglandin (2.alpha. synthase/D-arabinose dehydrogenase (PGFS)	Tc00.1047053511287.49	OG5_126583	3	29
LbrM.31.3240	<i>L. braziliensis</i>	LbrM.31.1,430,920 - 1,431,780 (-)	aldehyde reductase, putative oxidoreductase, putative	Tc00.1047053511287.49	OG5_126583	3	29

### 11.3 Proteins with post-translational modifications.

- Find all genes whose proteins have evidence of post-translational modification in *L. donovani*. How many did you get?
- How many have evidence of phosphorylation?  
Hint: revise your step from 'a' and select only the phosphorylation option.
- How many of these have any phenotypic evidence?  
Hint: add a step for phenotype found under "Putative Function".

The image shows two screenshots from a bioinformatics software interface. The top screenshot is the 'Add Step' dialog box, which has a tree view of search categories. The 'Phenotype' category is circled in red, and a red arrow points from it to the bottom screenshot. The bottom screenshot is the 'Identify Genes based on Phenotype' configuration window. It shows the following settings:

- Organism:**  Trypanosoma brucei TREU927
- Phenotype:**  growth (425)
- cell cycle (271)
- cell morphogenesis (270)
- mitochondrion organization and biogenesis (213)
- golgi organization and biogenesis (209)
- endocytosis (209)
- regulation of cell motility (208)
- enzyme activity (50)
- proteasomal ubiquitin-dependent protein catabolic process (13)
- cell motility (10)
- RNA editing (9)
- peptidase activity (9)
- molecular function (7)
- proteasome assembly (7)
- cytokinesis (4)
- flagellum biogenesis (4)
- RNA splicing (4)
- cell movement (3)
- carbon utilization (2)
- cellular response to phosphate starvation (2)

- How many results did you get? Did you get zero results? Why?  
Hint: where does the phenotype data come from, which organism?

Is there anything that can be done to get some results?

Hint: how about finding the orthologs of the *L. donovani* genes in other kinetoplastida.

My Strategies: New Opened (1) All (1)

(Genes)

Phenotype 474 Genes

Mass Spec 19 Genes Step 1

0 Genes Step 2

Add Step

Rename | View | Revise | Make Nested Strategy | Insert Step Before | Orthology | Delete

STEP 1 : Mass Spec

Experiment/Samples : Posttranslationally modified proteins from promastigote, amastigote, and differentiating promastigotes (Rosenzweig et al.), acetylated proteins, glycosylated proteins, methylated proteins, phosphorylated proteins

Minimum Number of Unique Peptide Sequences : 1

Minimum Number of Spectra : 1

Revise Step

Transform by Orthology

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

Leishmania

Trypanosoma

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

Syntenic Orthologs Only?

Give this search a name

Run Step

My Strategies: New Opened (1) All (1) Basket

(Genes)

Phenotype 474 Genes

Mass Spec 19 Genes Step 1

Orthologs 344 Genes Step 2

1 Genes Step 3

Add Step

My Strategies: New Opened (1) All (1) Basket

(Genes)

Phenotype 474 Genes

Mass Spec 96 Genes Step 1

Orthologs 1359 Genes Step 2

19 Genes Step 3

Add Step

- e. **How many genes did you get?** What happens if you revise your first step to include all types of post-translational modifications?

My Strategies: New Opened (1) All (1) Basket

(Genes)

Phenotype 474 Genes

Mass Spec 96 Genes Step 1

Orthologs 1359 Genes Step 2

19 Genes Step 3

Add Step

#### 11.4 Finding all genes with mass spec evidence in *L. infantum*.

- a. **Find all genes that have mass spec evidence in *L. infantum*.** How many genes did you get?

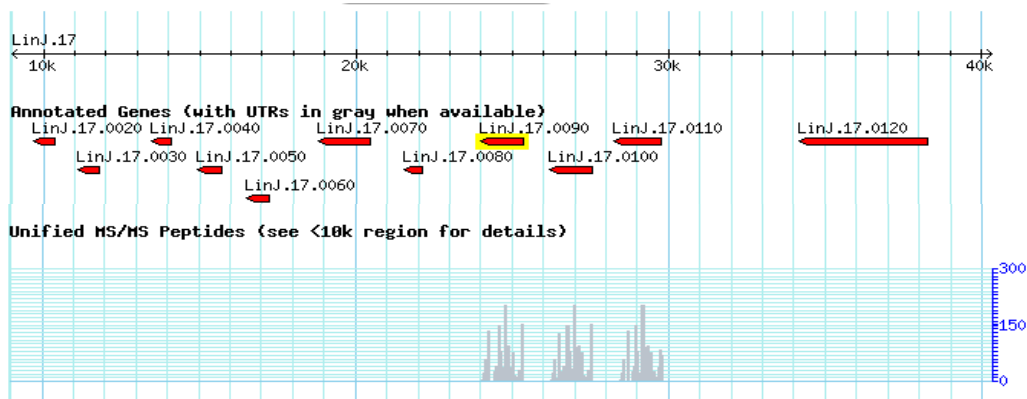
**b. Which gene has the largest number of peptide hits?**

Hint: sort “Number of Peptide Sequences” column.

**c. Which gene has the largest number of spectra?**

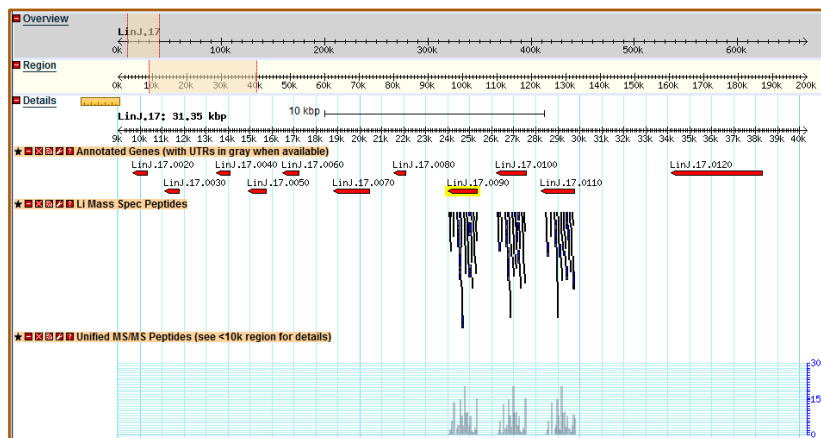
- Did the gene with the largest number of peptides also have the largest number of spectra?
- Is this surprising or plausible?

**d. Go to the gene page for one of the genes with the largest number of peptide hits (from step b). Take a look at the “Unified MS/MS Peptides” track in the genomic context view. What is this graphic telling you?**



**e. View this gene in GBrowse. Turn on the tracks for L. infantum MS/MS Peptides and for Unified MS/MS Peptides. Do you see a correlation between the graph and the peptides?**

Hint: you may wish to turn off all other tracks to make it easier to visualize.

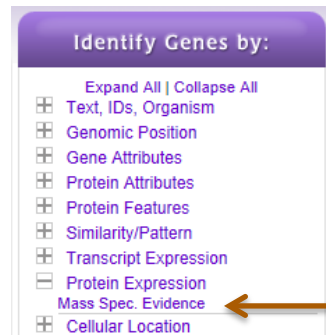


## 11.5 Finding genes with mass spec evidence in *P. berghei* gametocytes.

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

- a. Find all *P. berghei* genes that have mass spec evidence in either or both male and female gametocytes.

Hint: mass spec searches are in the “protein expression” expression section. Either or both is the Union of both results, not the intersection.



- 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)?
- 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.
- c. Find genes that have mass spec evidence only in male gametocytes and not in female ones.  
Hint: modify the set operation in b.
- d. Find genes that have mass spec evidence only in female gametocytes and not in male ones.  
Hint: modify the set operation in b.
- e. Which female gametocyte gene has the highest number of peptide sequences?  
Hint: look at the “number of peptide sequences” column in the list of results.
- 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.