

## Exploring proteomics data in VEuPathDB Resources

Data from proteomics experiments are integrated into VEuPathDB resources under three categories:

### **1. Mass spec. evidence**

*Peptides from proteomics experiments are mapped to a reference genome enabling searches for genes based on that mapping.*

### **2. Quantitative mass spec. evidence**

*Data from quantitative proteomic experiments are loaded and made available for searching based on fold change or differential expression.*

### **3. Post-translational modification (PTM)**

*PTM data from proteomics experiments are loaded on genes enabling searches for genes based on the type and number of the PTM.*

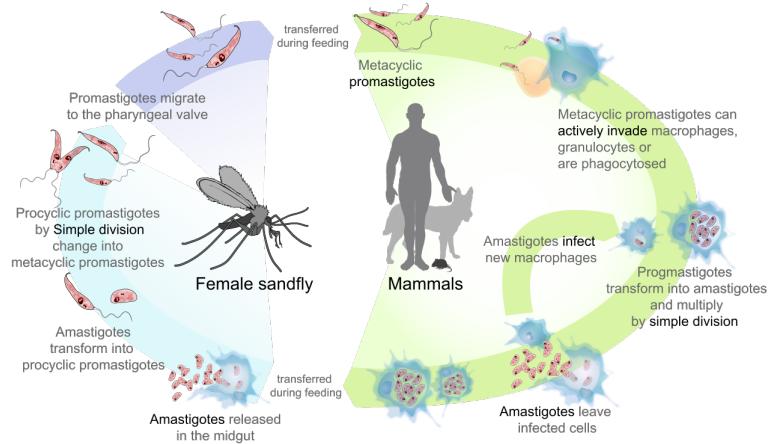
The exercises below explore the different categories and searches available for proteomics in VEuPathDB.

### **Learning objectives:**

- Understand the different categories of proteomics data
- Learn how to run searches to identify genes based on peptide evidence
- Learn how to identify differentially expressed genes based on quantitative data
- Learn how to identify genes with different PTMs

**1. Find genes that have peptide evidence from metacyclic stages but not amastigote or promastigote stages of *Leishmania infantum*.**

Note: for this exercise use <http://tritrypdb.org>



Life cycle of Leishmania. [https://commons.wikimedia.org/wiki/File:Leishmaniasis\\_life\\_cycle\\_diagram\\_en.svg](https://commons.wikimedia.org/wiki/File:Leishmaniasis_life_cycle_diagram_en.svg)

- Navigate to the mass spec. evidence search. How did you find it? You can use the search filter on the left of the home page or in the searches menu at the top of the page. Filter the searches by typing a word in the filter box.

The screenshot shows the TriTrypDB homepage. At the top, there is a navigation bar with links for My Strategies, Searches, Tools, My Workspace, Data, About, Help, and Contact Us. The main search bar contains the text "proteo". Below the search bar, there is a "Search for..." field containing "proteomi". To the right of this field, there is a dropdown menu titled "Genes" with options: Proteomics, Mass Spec. Evidence, Post-Translational Modification, and Quantitative Mass Spec. Evidence. The "Mass Spec. Evidence" option is highlighted with a red arrow. On the right side of the screen, there is a "Overview of Resources" section with a "Getting Started" guide. The guide explains that VEuPathDB is packed with data, tools, and visualizations to help answer research questions. It provides a link to the "Getting Started" tutorial.

- b. Select all *L. infantum* samples that come from the amastigote or promastigote stages. Note that you can filter the samples with key words like amastigote.

## Identify Genes based on Mass Spec. Evidence

10 selected, out of 151

[add these](#) | [clear these](#) | [select only these](#)  
[select all](#) | [clear all](#)

amasti

Leishmania

- Leishmania donovani
  - Leishmania donovani BPK282A1
    - Promastigote and amastigote stage proteomes (MHOM/IN/80/Dd8) (Nirujogi et al.)
      - amastigote
      - promastigote
- Leishmania infantum
  - Leishmania infantum JPCM5
    - Promastigote and Amastigote Phosphoproteomes (donovani) (Tsigankov et al.)
    - amastigote phosphopeptides
    - promastigote phosphopeptides
    - Promastigote and amastigote proteomes (MHOM/MA/67/ITMAP-263) (Brotherton et al.)
      - amastigote by 1DE, LC-MS/MS
      - amastigote by 2DE, LC-MS/MS, pH6-11
      - amastigote by 2DE, LC-MS/MS, pH6-9
      - promastigote by 2DE, LC-MS/MS, pH6-9
      - promastigote by 2DE, LC-MS/MS, pH control
      - promastigote by 2DE, LC-MS/MS, temp and pH stressed
      - promastigote secretome
- Leishmania mexicana
  - Leishmania mexicana MHOM/GT/2001/U1103
    - Intracellular Amastigotes (MNYC/BZ/62/M379) (Paape et al.)
      - amastigotes (FACS sorted, LC-MS/MS)
- Trypanosoma
  - Trypanosoma cruzi
    - Trypanosoma cruzi CL Brener Esmeraldo-like

- c. Keep the default search parameters and click on the Get Answer button.

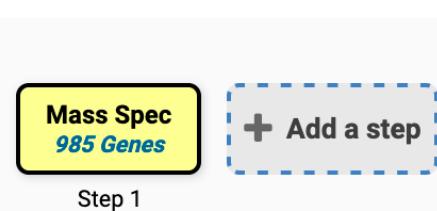
Minimum Number of Unique Peptide Sequences

Apply min # peptide sequences / sample OR across samples

[Advanced Parameters](#)

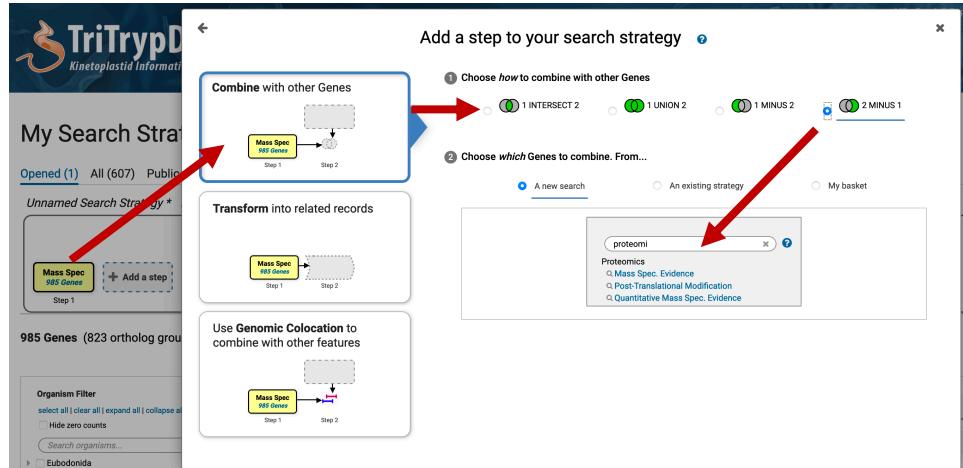
**Get Answer**

- d. How many genes did you get?

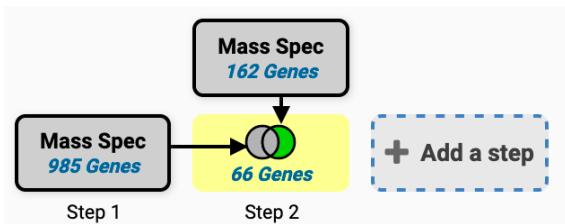


e. Can you remove these results from any gene with peptide evidence from the metacyclic stage of *L. infantum*? Try the following:

- Click on add step
- Select how to combine the results
- Find and click on the mass spec. evidence search
- Select the metacyclic stage proteome data and click on the Get Answer button.



f. How many genes did you get? Explore the results, do they make sense from a biological standpoint?



g. How can you increase the stringency of your results? One way is to increase the minimum number of unique peptides. The default returns any gene with a minimum of one peptide. What happens if you change this to a minimum of 5 peptides in both steps?

- Click on the edit button
- Click on the revise option in the popup
- Change the value from 1 to 5 and click on the Revise button.

- Remember you need to do this for each step.

## My Search Strategies

**Opened (1) All (607) Public (41) Help**

**Unnamed Search Strategy \***

**Step 1** Mass Spec 985 Genes    **Step 2** Mass Spec 66 Genes    **+ Add a step**

**66 Genes (63 ortholog groups)**

**Organism Filter** select all | clear all | Hide zero count organisms

**Experiments and Samples**

amastigote phosphopeptides, promastigote by 1DE, LC-MS/MS, phosphopeptides, amastigote by 1DE, LC-MS/MS, amastigote by 2DE, LC-MS/MS, pH6-11, amastigote by 2DE, LC-MS/MS, pH6-9, amastigote by 2DE, LC-MS/MS, pH6-9, amastigote by 2DE, LC-MS/MS, temp and pH control, promastigote by 2DE, LC-MS/MS, temp and pH stressed, promastigote secretome

**Minimum Number of Unique Peptide Sequences** 1

Apply min # peptide sequences / sample OR across samples Per Sample

**Revise your step**

**Experiments and Samples**

10 selected, out of 151

select all | clear all | expand all | collapse all

Filter list below... ?

Leishmania (selected) Trypanosoma

select all | clear all | expand all | collapse all

**Minimum Number of Unique Peptide Sequences**

5

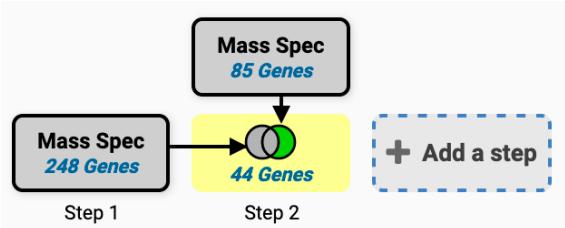
**Apply min # peptide sequences / sample OR across samples**

Per Sample

Advanced Parameters

**Revise**

- h. How did this change your results? Would you consider these results more stringent?



**2. Find genes in *Plasmodium falciparum* that are quantitatively present at a higher concentration in the apicoplast compared to the endoplasmic reticulum (ER). Note for this exercise use <https://plasmodb.org>**

- Go to the quantitative mass spec evidence searches
- Select the experiment called Apicoplast and ER Proteomes (Quantitative)(Dd2) (Boucher et al)

Search for... **quant**

Identify Genes based on Quantitative Mass Spec. Evidence

Organism	Data Set	Choose a Search
Plasmodium falciparum 3D7	Long-lived merozoite proteome (Kumar et al.)	FC
Plasmodium falciparum 3D7	Proteome and phosphoproteome during intraerythrocytic development (Quantitative) (Pease et al.)	FC
Plasmodium falciparum 3D7	Apicoplast and ER Proteomes (Quantitative)(Dd2) (Boucher et al)	FC

- Configure this search to return all genes that are upregulated by 1.5 fold in the apicoplast sample compared to the ER sample

Identify Genes based on P. falciparum 3D7 Apicoplast and ER Proteomes (Quantitative)(Dd2) Proteomics (fold change)

Reset values

For the Experiment  
 Apicoplast and ER Proteomes (Quantitative)(Dd2)

return protein coding  Genes

that are up-regulated

with a Fold change >= **1.5**

between each gene's minimum  expression value

in the following Reference Samples

Apicoplast  
 ER

select all | clear all

and its maximum  expression value

in the following Comparison Samples

Apicoplast  
 ER

select all | clear all

Example showing one gene that would meet search criteria  
(Dots represent this gene's expression values for selected samples)

Up-regulated

Expression

1.5 fold

Expression Value Comparison

Expression Value Reference

Reference Samples Comparison Samples

For each gene, the search calculates:  

$$\text{fold change} = \frac{\text{comparison expression value}}{\text{reference expression value}}$$

and returns genes when fold change >= 1.5.

You are searching for genes that are up-regulated between one reference sample and one comparison sample.

- How many genes did your search return?

PfDd2 Apico ER Prot (fc)  
**272 Genes**

+ Add a step

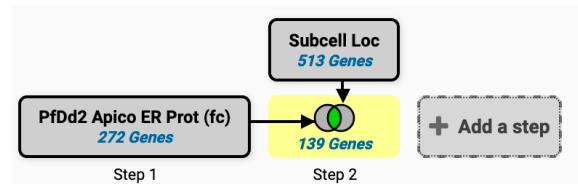
Step 1

- Can you further limit your results by leveraging available subcellular localization data?

- Click on the add step button and find the subcellular localization search

The screenshot shows the 'My Search Strategies' interface. A modal window titled 'Add a step to your search strategy' is open. In the 'Combine with other Genes' section, there is a 'Step 1' box containing 'PfDd2 Apico ER Prot (fc) 272 Genes'. A red arrow points to the '+ Add a step' button below it. To the right, under 'Choose how to combine with other Genes', the '1 INTERSECT 2' option is selected. Below that, under 'Choose which Genes to combine. From...', the 'subce' filter is applied, with 'Pf. Subcellular Localization' checked. A red arrow points to this checkbox.

- Make sure Apicoplast localization is selected and click on the Run Step button. How many genes did you identify?



### 3. I identify *Cryptococcus neoformans* genes that are upregulated in a protein kinase A dependent (PKA) manner and not in a non-PKA dependent manner.

Note for this exercise use <https://fungidb.org>

The expression of virulence factors in *C. neoformans*, including capsule and melanin, is in part regulated by the cyclic-AMP/protein kinase A (cAMP/PKA) signal transduction pathway. *C. neoformans* PGAL7::PKA1 strain can be used to induce the PKA pathway in galactose media and repress the pathway in glucose media.

- Go to the quantitative proteomic search section and find the experiment called “Analysis of the protein kinase A-regulated proteome of *Cryptococcus neoformans* (Geddes et al.)”

Identify Genes based on Quantitative Mass Spec. Evidence

The screenshot shows a table of experiments. The last row, 'Analysis of the protein kinase A-regulated proteome of Cryptococcus neoformans (Geddes et al.)', has a 'DC' button circled in red.

Organism	Data Set	Legend:
Aspergillus clavatus NRRL 1	Proteomics changes in response to human serum (Wiedner et al. 2013)	FC
Aspergillus fischeri NRRL 181	Proteomics changes in response to human serum (Wiedner et al. 2013)	FC
Aspergillus fumigatus Af293	Proteomics changes in response to human serum (Wiedner et al. 2013)	FC
Aspergillus fumigatus Af293	Aspergillus fumigatus response to hypoxia (Barker et al. 2012)	FC
Aspergillus fumigatus Af293	Development stage specific proteome (Suh et al.)	FC
Aspergillus fumigatus Af293	Adaptive mechanisms of Aspergillus fumigatus conidia to nutrient restriction Quant (Andjo et al.)	FC
Cryptococcus neoformans var. grubii H99	Secretome profiling of Pka-1 regulated proteins in Cryptococcus neoformans (16, 48, 72, and 120 hr post inoculation) (Geddes et al.)	FC
Cryptococcus neoformans var. grubii H99	Analysis of the protein kinase A-regulated proteome of Cryptococcus neoformans (Geddes et al.)	DC
Neurospora crassa OR74A	Circadian time course data from wild type and delta csp-1 (Hurley et al.)	DC

- b. Configure the direct comparison search to identify genes that are upregulated by 3 fold in galactose media

Identify Genes based on C. neoformans var. grubii H99 Analysis of the protein kinase A-regulated proteome of Cryptococcus neoformans Proteomics (direct comparison)

Analysis of the protein kinase A-regulated proteome of Cryptococcus neoformans

**Direction**

up-regulated  ←

**Comparison**

PGAL7::PKA1 + glucose  
 PGAL7::PKA1 + galactose ←

**Fold difference >=**

3 ←

Get Answer

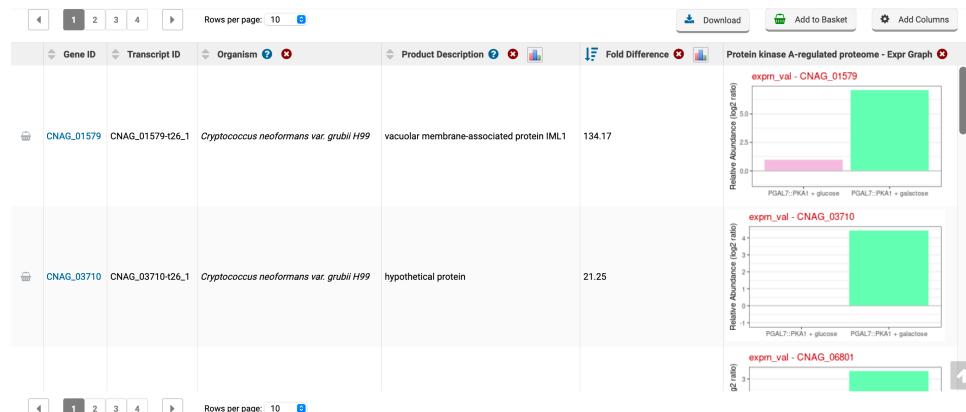
- c. How many genes did you get?

Protein kinase A-regulated prot...  
28 Genes

+ Add a step

Step 1

- d. Explore your results. Do the expression graphs meet the criteria you selected?

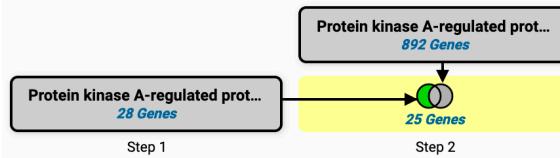


- e. Add a step and remove from this list any gene that is upregulated by 1.5 fold in glucose media.

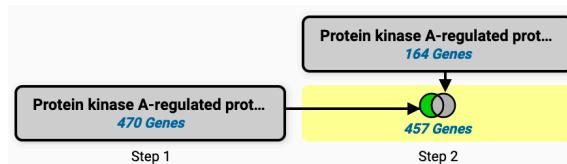
The screenshot shows the FungiDB search interface with the following configuration:

- Step 1:** "Protein kinase A-regulated prot... 28 Genes"
- Step 2:** "Combine with other Genes" (selected operation: "1 MINUS 2")
- Search Strategy:** "Proteomics" (selected) and "Quantitative Mass Spec. Evidence" (unchecked)
- Direction:** "up-regulated" (selected)
- Comparison:** "PGAL7::PKA1 + glucose" (selected)
- Fold difference >=:** "1.5" (entered)

- f. How many genes did you get?



- g. Can you reconfigure the above searches to identify genes that are downregulated as opposed to upregulated? Did your results change?



**4. Find genes with evidence of protein phosphorylation in intracellular *Toxoplasma* tachyzoites.**  
**Note for this exercise use <https://toxodb.org>**

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

- Find all genes that have at least 5 phosphorylation events from all available phosphoproteomic experiments. Navigate to the Post-Translational Modification search. Filter the samples using the key word *phospho* then select all the samples that are phosphopeptide enriched.
- Next make sure to change the number of modifications to 5.
- How many genes did you return? Which gene has the highest number of phosphorylation sites? Hint, examine the column called total modified residues.

5 selected, out of 9

[add these](#) | [clear these](#) | [select only these](#)  
[select all](#) | [clear all](#)

➡

**Toxoplasma gondii**  
 **Toxoplasma gondii GT1**

- Tachyzoite phosphoproteome from purified parasite or infected host cell (RH) (Treck et al.)
- Infected host cell, phosphopeptide-depleted (peptide discovery against TgGT1)
- Infected host cell, phosphopeptide-enriched (peptide discovery against TgGT1)
- Purified tachyzoites phosphopeptide-depleted (peptide discovery against TgGT1)
- Purified tachyzoites phosphopeptide-enriched (peptide discovery against TgGT1)

**Toxoplasma gondii ME49**

- Tachyzoite phosphoproteome - Calcium dependent (RH) (Nebi et al.)
- phosphopeptide-enriched (via Mascot)
- phosphopeptide-depleted (via Sequest)
- phosphopeptide-enriched (via Sequest)

Tachyzoite phosphoproteome from purified parasite or infected host cell (RH) (Treck et al.)

- Infected host cell, phosphopeptide-enriched (peptide discovery against TgME49)
- Purified tachyzoites phosphopeptide-enriched (peptide discovery against TgME49)

[add these](#) | [clear these](#) | [select only these](#)  
[select all](#) | [clear all](#)

**Number of modifications is**

Greater than or equal to  ➡

**Number of Modifications**

➡

Post-Translational Mod (250 items) + Add a step

Step 1

2,266 Genes (1,531 ortholog groups) Revise this search

Gene Results | Genome View | Analyze Results

Organism Filter  
 **Entamoebae**  
 **Microsporidia**  
 **Toxoplasma**  
 **Other organisms**

Hide Organism Filter

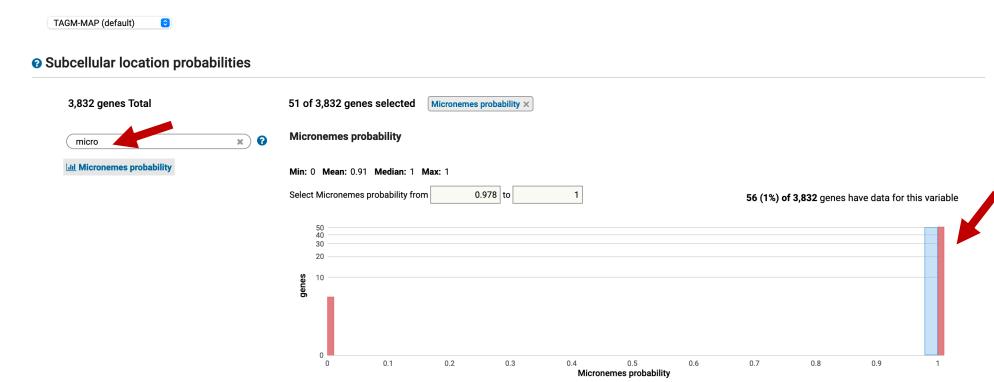
Gene ID	Transcript ID	Modified Residues	Total Modified Residues	Modifications By Type	Type
TGDT1_311230	TGDT1_311230	S130, S164, S174, S231, S240, S243, T248, T330, S404, T405, I26, I	126	phosphorylation site:128	hypothetical protein
TGDT1_235790	TGDT1_235790	Y129, S131, T170, S213, S222, S242, T243, S273, S278, S293, S294, T295, I26, I	115	phosphorylation site:115	PLU-1 family protein
TGDT1_209000	TGDT1_209000	S130, S134S, S130, S135A, S139, S134, S134, S140I, S140S, I26, I	110	phosphorylation site:113	HECT-Domain (ubiquitin-transferase) domain-containing protein
TGDT1_254940	TGDT1_254940	S103, S137, S143, S234, S678, S705, S769, S791, S809, S812, T813, S814, S815, S816, S817, S818, S819, S820, S821, S822, S823, S824, S825, S826, I26, I	106	phosphorylation site:106	MIF4G domain-containing protein
TGDT1_291180	TGDT1_291180	S198, S246, S255, S259, S263, S265, S266, T270, S271, T289, S304, S308, T341, S342, S343, S344, S345, S346, S347, S348, S349, S350, S351, T352, I26, I	103	phosphorylation site:103	hypothetical protein
TGDT1_232080	TGDT1_232080	S320, T440, S442, S445, S503, S568, S699, S990, S1042, S1043, S1044, S1045, S1046, S1047, S1048, S1049, S1050, S1051, T352, I26, I	103	phosphorylation site:103	hypothetical protein

- How many of these phosphorylated genes are also localized to the microneme organelles?

- Add a step and locate the Protein Targeting and Localization searches. Select the one called Localization by LOPIT Mass Spec.

The screenshot shows the 'Unnamed Search Strategy' interface. On the left, there's a summary of 'Post-Translational Mod' genes (2,266 Genes). In the center, a 'Combine with other Genes' step is shown, followed by a 'Transform into related records' step. On the right, a modal window titled 'Add a step to your search strategy' is open. It has two sections: 'Choose how to combine with other Genes' (with options 1 INTERSECT 2, 1 UNION 2, 1 MINUS 2, 2 MINUS 1) and 'Choose which Genes to combine. From...' (with options A new search, An existing strategy, and My basket). A search bar contains the text 'local'. Below it, a list includes 'Protein targeting and localization' (selected), 'Predicted Signal Peptide', and 'Transmembrane Domain Count'. A red arrow points from the 'Add a step' button to the 'local' search term.

- Filter the localization categories using the word microneme. Select all genes with a probability of 1 (or close to 1).



- Explore your results.

The screenshot shows the 'Gene Results' section. At the top, it says 'Opened (1) All (1) Public (9) Help' and 'Unnamed Search Strategy \*'. The search strategy summary shows 'Post-Translational Mod' genes (2,266 Genes) and 'LOPIT' genes (51 Genes). A 'Gene Results' table is displayed with columns: Gene ID, Transcript ID, Genomic Location (Gene), Product Description, and # Transcripts. The table contains five rows of data. A red arrow points to the 'Gene ID' column header.

Gene ID	Transcript ID	Genomic Location (Gene)	Product Description	# Transcripts
TGME49_221180	TGME49_221180-126..1	TGME49_chrl:108,336,114,763(-)	hypothetical protein	1
TGME49_205680	TGME49_205680-126..1	TGME49_chrl:1,214,495,1,219,931(-)	hypothetical protein	1
TGME49_226020	TGME49_226020-126..1	TGME49_chrxII:1,862,972,1,869,285(-)	transporter, major facilitator family protein	1
TGME49_245490	TGME49_245490-126..1	TGME49_chrxII:2,306,394,2,309,968(+)	microneme protein MIC8	1