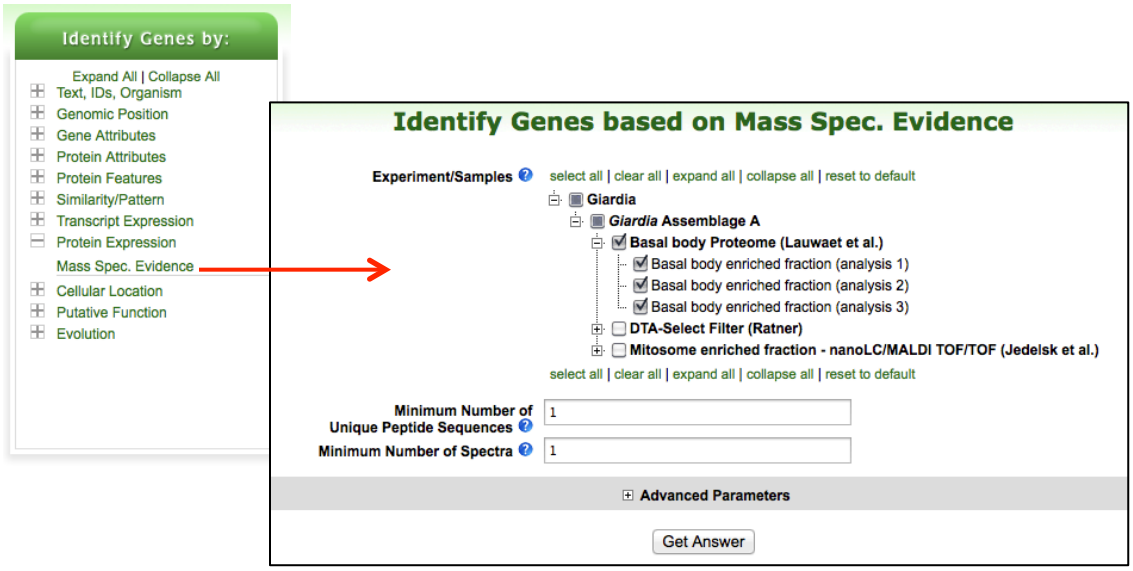
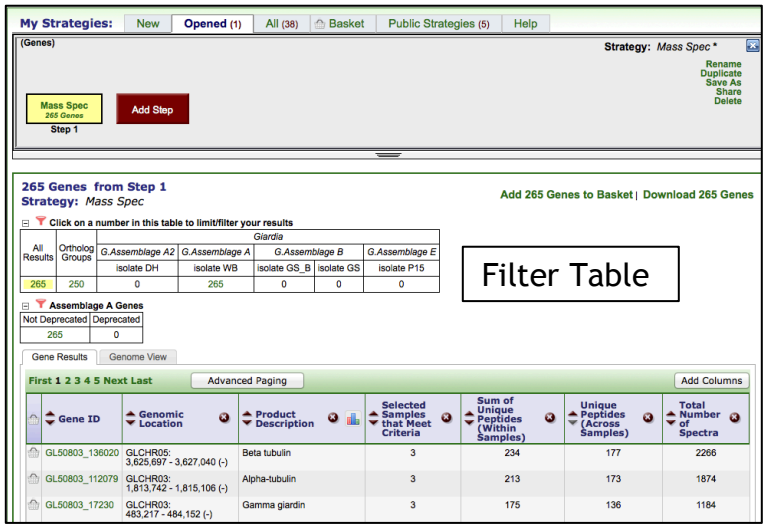
Exploring Proteomics Data
(Exercise 12)

1. **Find all genes with mass spec evidence from the basal body in Giardia. (For this exercise use http://giardiadb.org)**

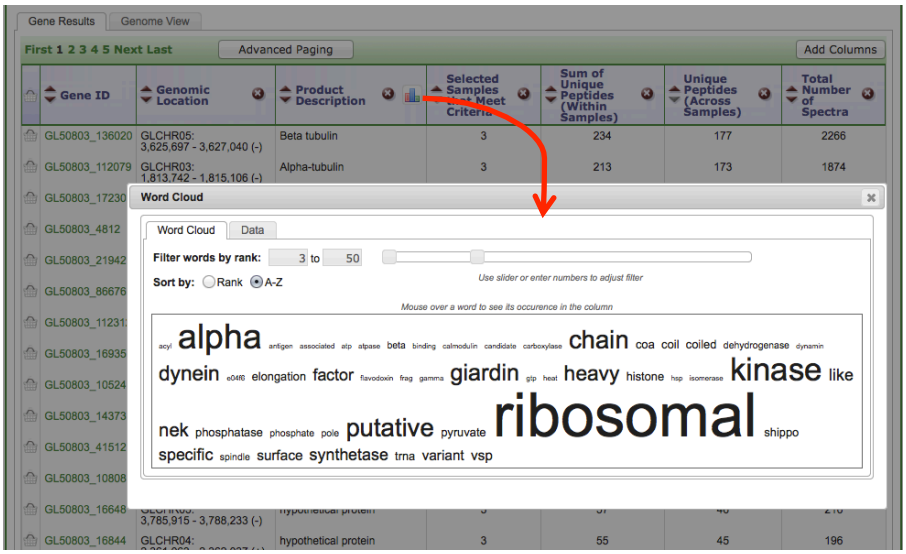
   - The proteomics search is available under the heading "Protein Expression" in the "Identify gene by" section.



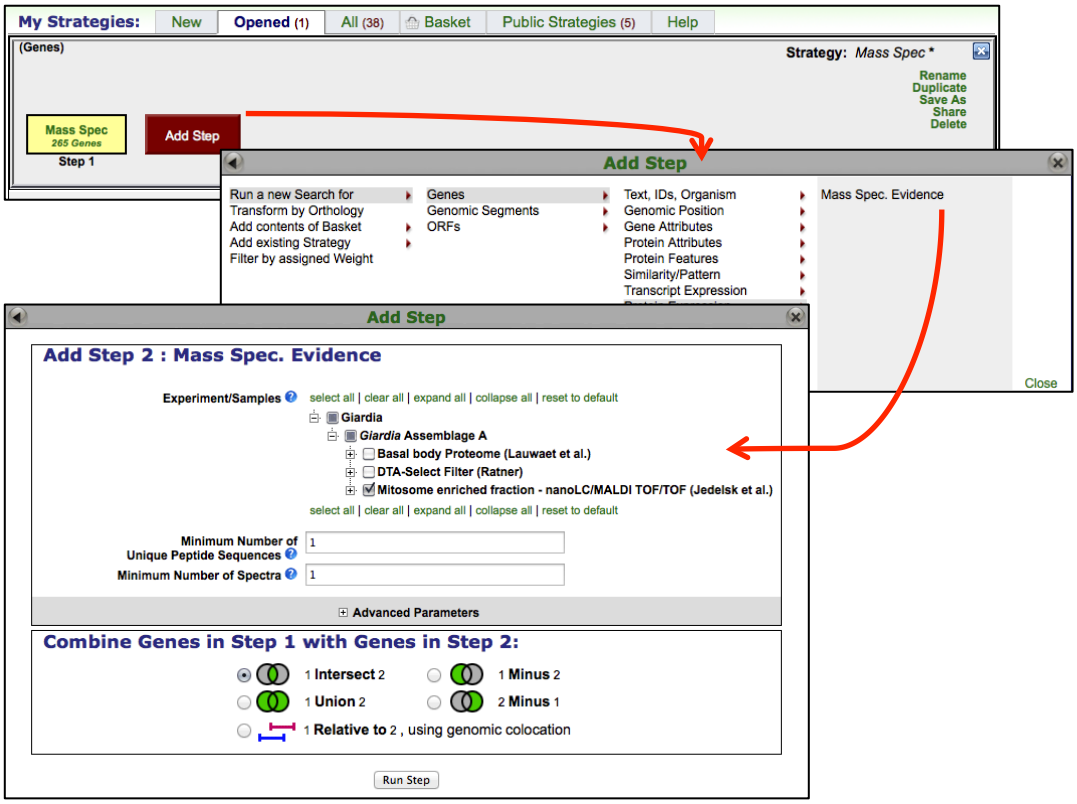
   - How many genes did you identify? Examine the filter table below the strategy section. Why is the number in the ortholog groups column less than the number in the Assemblage A column? Why do the other cells have a zero?



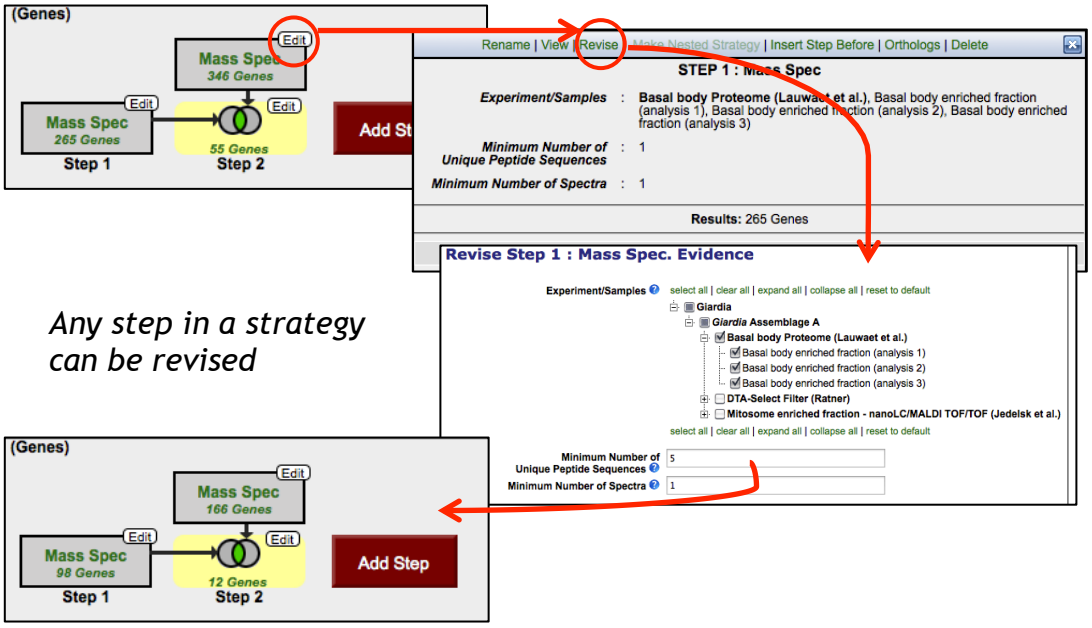
- What kinds of genes are in your result list? (hint: one option is to use the word cloud tool – click on the little graphic icon next to the column called “Product description”).



- How many genes with mass spec data from the basal proteome experiment also have mass spec evidence from the “mitosome enriched fraction experiment”?

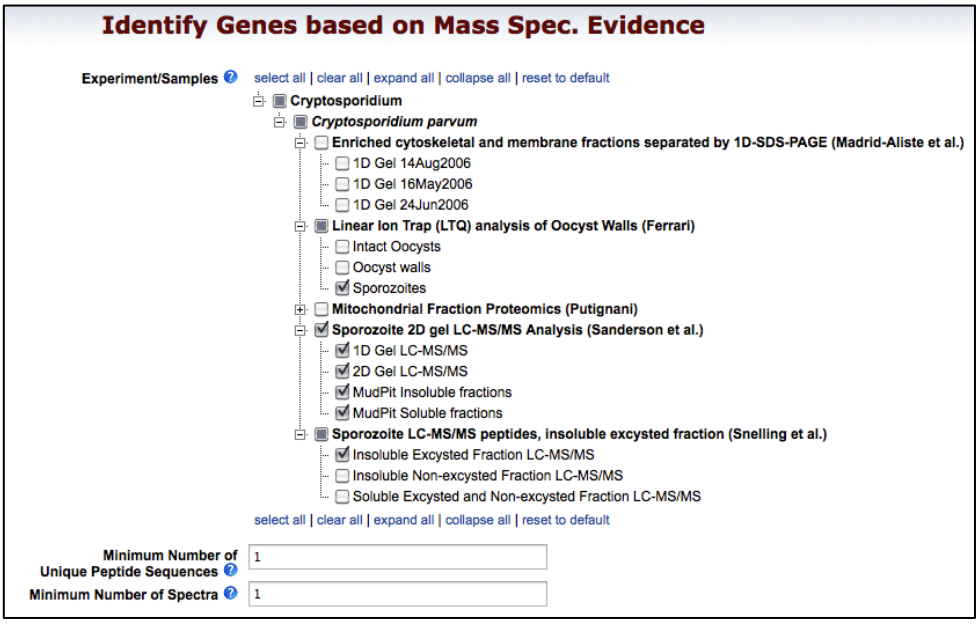


- The default parameters of the mass spec search is to identify any gene with at least 1 peptide identified. How will your results change if you revise the two steps in your search strategy to only return genes with at least 5 peptides identified?

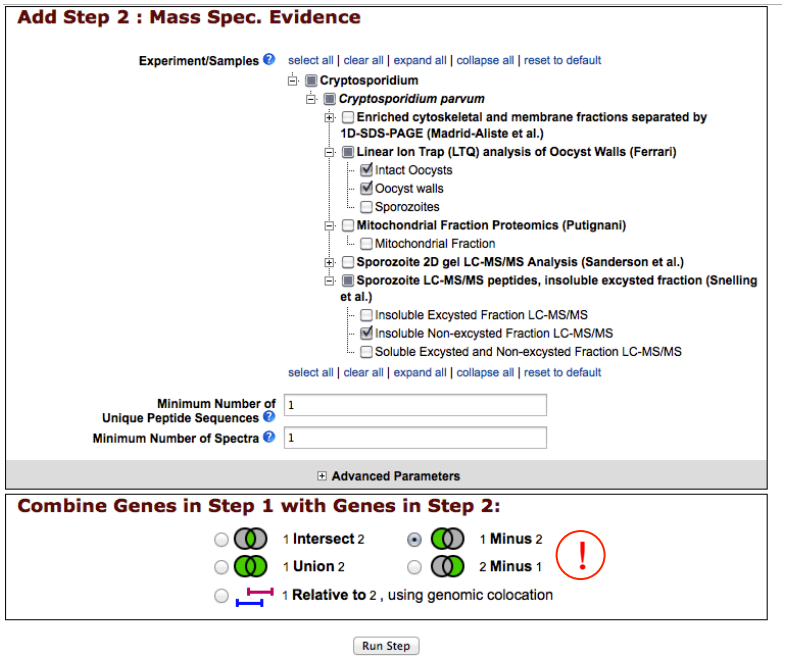


2. **Find gene in Cryptosporidium that have mass spec evidence from any of the sporozoite proteomics experiments available in CryptoDB. (For this exercise use http://cryptodb.org)**

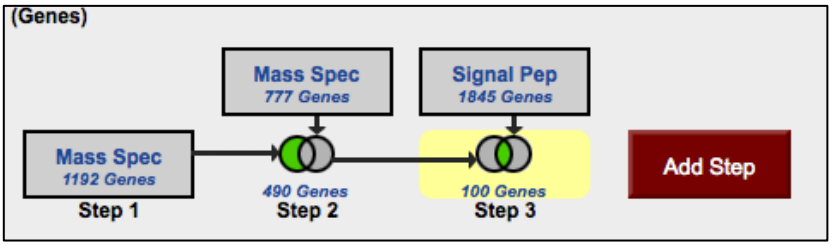
- Explore the available proteomics data and select samples that make sense. You may need to click on the '+' sign to expand experiments to see the underlying samples.



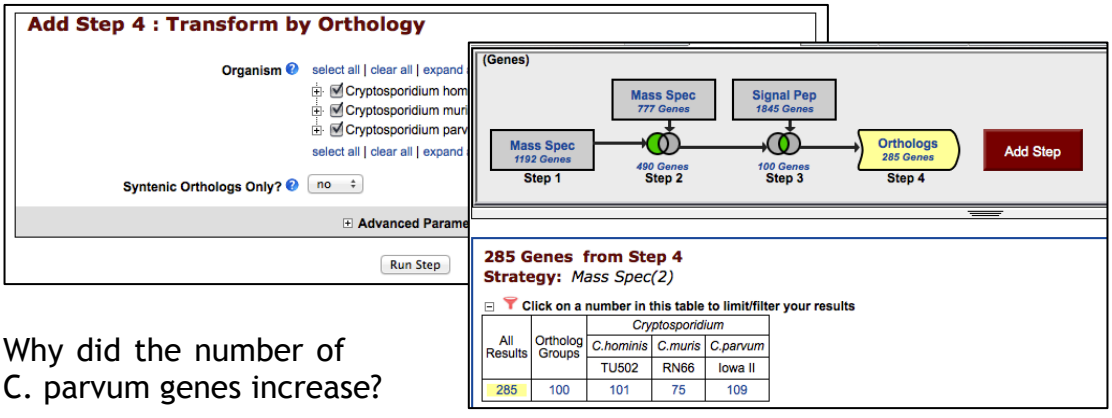
- How can you remove any gene with peptide evidence from non-sporozoite samples? (hint: add a step for mass spec data and think about how you will combine your results.)



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



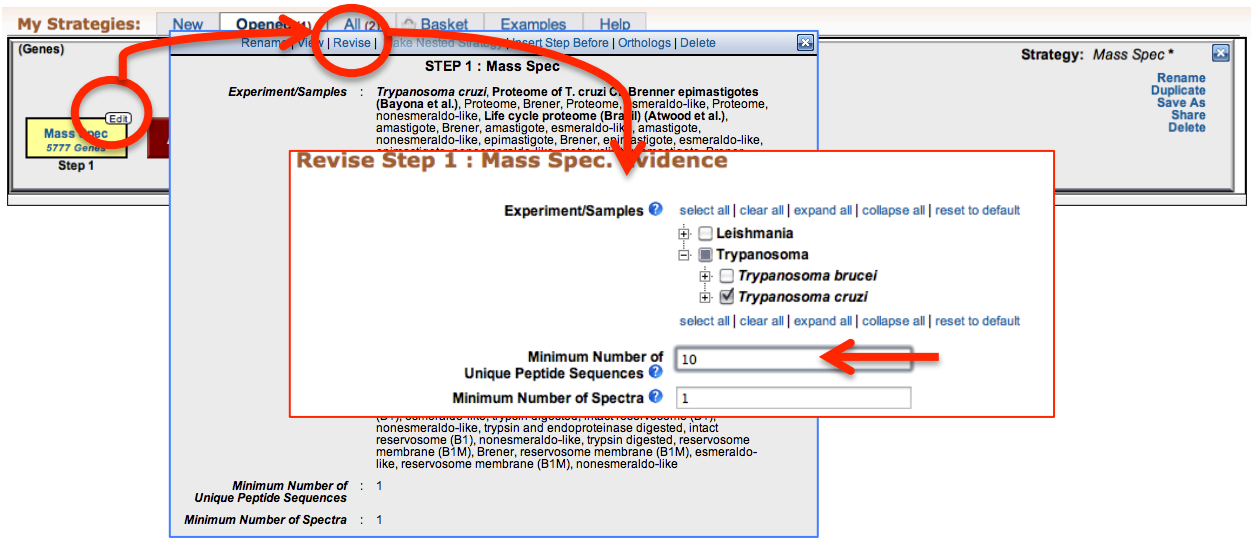
- Note that so far you have been searching for *C. parvum* genes because we only have proteomics data from this species. However, what if you are interested in the orthologs of these genes in *C. muris*. How can you transform your *C. parvum* results to *C. musris* genes? (hint: add a step then select the "transform by orthology" option).



- Why did the number of C. parvum genes increase?

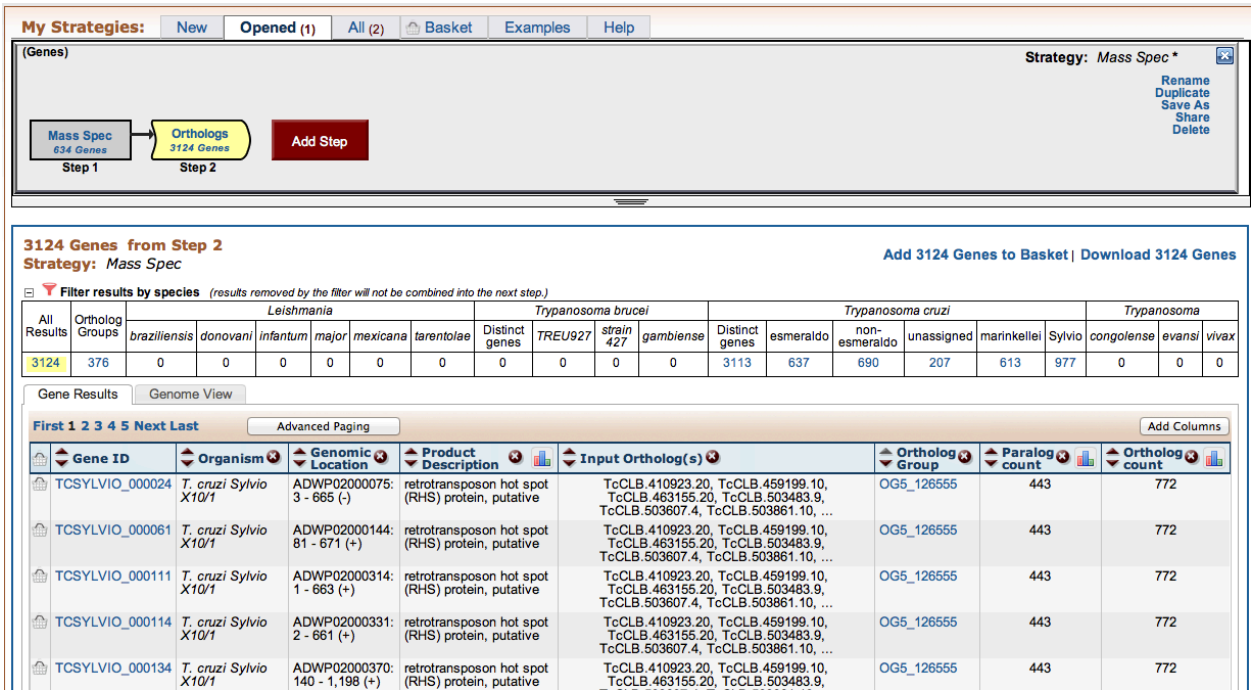
3. **Finding all genes with mass spec evidence in *T. cruzi*.**
   Note: for this exercise use http://TriTrypDB.org.

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



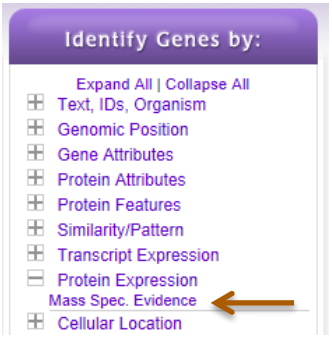
c. Can you expand the list of results in 'b' to include possible orthologs/paralogs in *T. cruzi*?

   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.



4. **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).

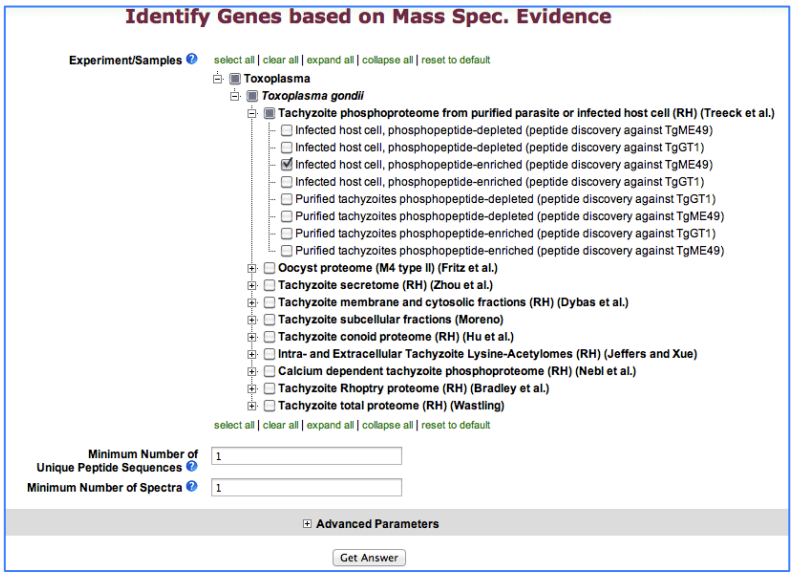
e. **Which female gametocyte gene has the highest number of peptide sequences?**
   (*hint*: look at the "number of peptide sequences" column in the list of results).

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).

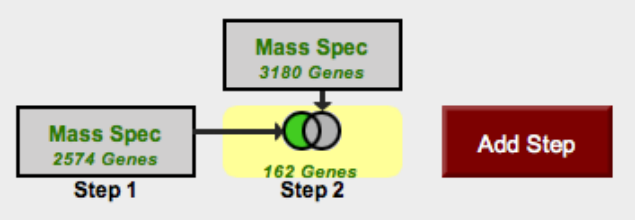
5. **Finding genes with evidence of phosphorylation in intracellular *Toxoplasma* tachyzoites.**
   Note: For this exercise use http://www.toxodb.org

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

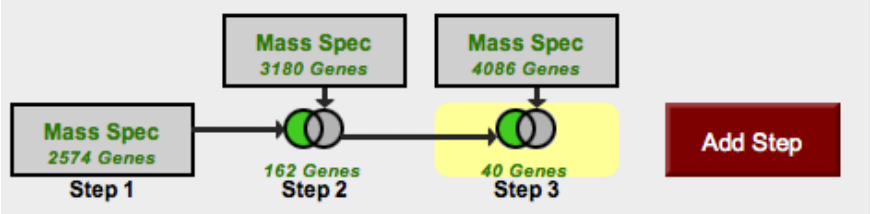
a. Find all genes with evidence of phosphorylation in intracellular tachyzoites. Select the "Infected host cell, phosphopeptide-enriched (peptide discovery against TgME49)" sample under the experiment called "Tachyzoite phosphoproteome from purified parasite or infected host cell (RH) (Treeck et al.)"



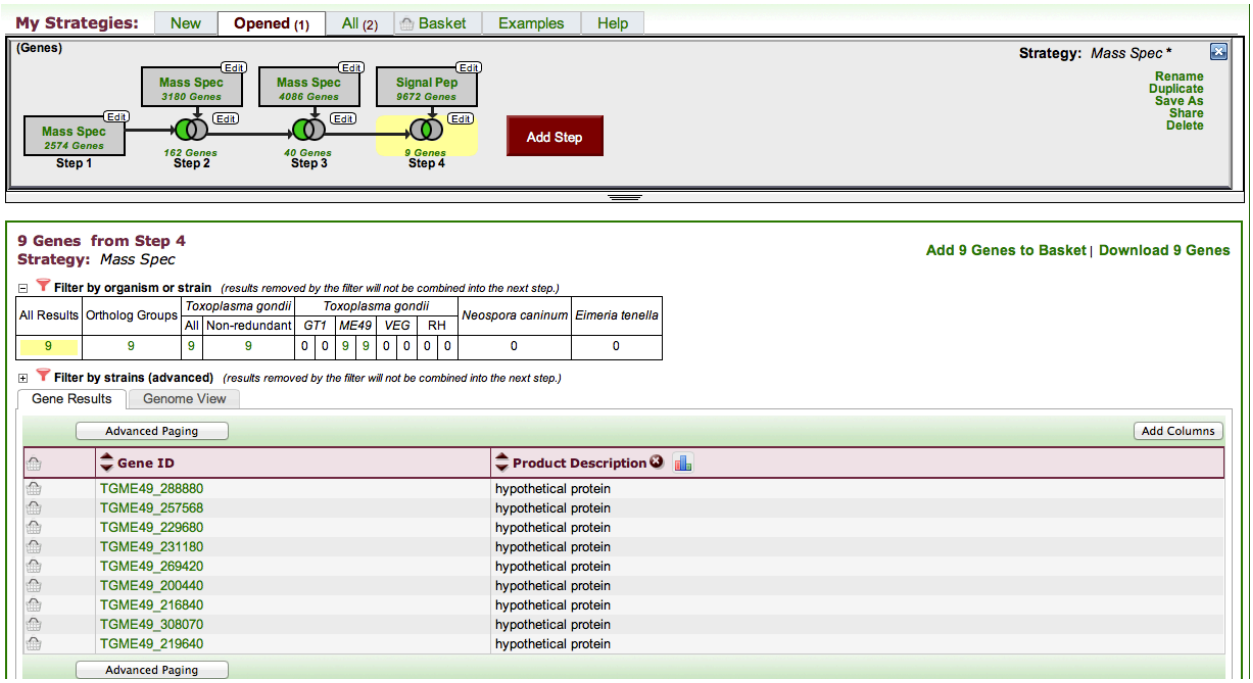
b. Remove all genes with phosphorylation evidence from purified extracellular tachyzoites.



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



d. Explore your results. What kids of genes did you find? Are any of these results to be secreted? (*Hint*: add a step searching for genes with secretory



signal peptides).

e. 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).

f. What about polymorphism data? Go back to your strategy and add columns for SNP data found under the population section. Explore the gene page for the gene that has the most number of nonsynonymous SNPs.