

## Data Integration II

### Complex strategy: finding vaccines and drug targets

#### Exercise 11

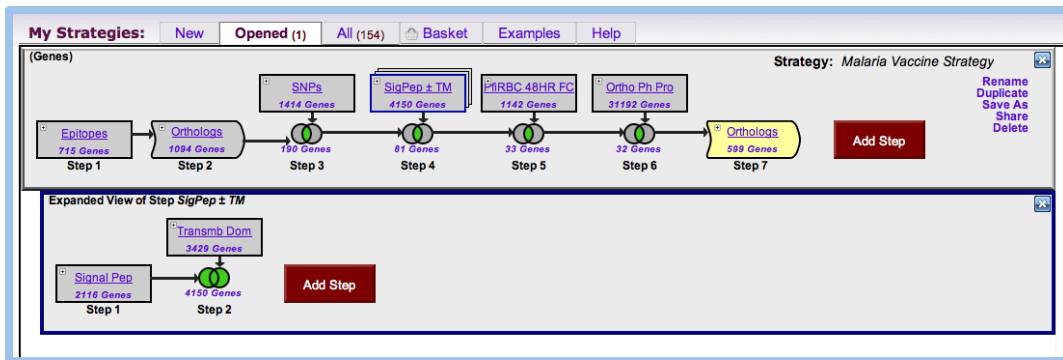
11.1 Defining possible vaccine candidates in malaria.

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

- a. Considering all of the many 'Queries & Tools' available on PlasmoDB, how many criteria can you define that might be useful for identifying candidate vaccine targets?
- b. After *first* trying to develop your own query, you might be interested to look at an example query:

Malaria Vaccine Strategy - <http://plasmodb.org/plasmo/im.do?s=14627aa18a8052a5>

Note: copy and paste the URL into your browser if the link does not work.

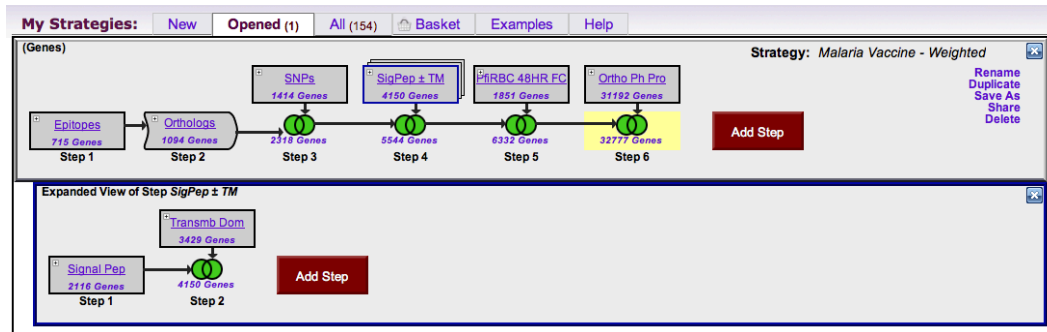


Try revising various components of this query to improve it still further to reflect your own insights, theories or experience. **Note that if you have logged in, you can save the results of your queries for future reference, or to share with others.**

- c. How would your results change if you used weighted searches? Revise each step of your strategy and assign a weight to it. The weight is arbitrary; you decide on the scale and the results are sorted based on the sum of the weights. Remember, for weighting to work you have to use the Union operation to join the steps. After doing this, what are some of your top candidates? (Hint: you can sort the results columns by clicking on the arrows next to the column name.) Here is the example from above with assigned weights (note that the weighted strategy may take a while to load):

Malaria Vaccine – Weighted - <http://plasmodb.org/plasmo/im.do?s=094f2aae5385bbd9>

Note: copy and paste the URL into your browser if the link does not work.



## 11.2 Defining possible drug targets in Trypanosomes.

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

a. Consider your ideal drug target – what may be some of its characteristics?

- Would it be useful if the drug target is an enzyme?
- Would it be better to identify something conserved between the parasite and the host or not?
- What about the biology of the parasite you are working on? When would you prefer this protein to be expressed? For example, in *T. brucei* an enzyme may be more active during the slender replicative form. Alternatively, you may decide to concentrate on proteins involved in differentiation of the parasite. For example, this may include proteins that are differentially expressed between slender and the stumpy (cell cycle arrested) forms of *T. brucei*.

- b. How would you build your strategy? One place to start is to ask for all proteins in TriTrypDB that do not have orthologs in mammals. You may wish to add a step for anything with an enzyme commission (EC) number. Using microarray evidence to specifically look for differentially regulated genes between slender and stumpy forms might be useful (hint: use a nested strategy to combine multiple microarray experiments). What about other data types, for example, phenotype data?

**Identify Genes by:**

Expand All | Collapse All

- Text, IDs, Organism
- Genomic Position
- Gene Attributes
- Protein Attributes
- Protein Features
- Similarity/Pattern
- Transcript Expression
- Protein Expression
- Cellular Location
- Putative Function
- Evolution
- Orthology Phylogenetic Profile
- Population Biology

**Identify Genes based on Orthology Phylogenetic Profile**

Show results from species: [select all](#) | [clear all](#) | [expand all](#) | [collapse all](#) | [reset to default](#)

- Leishmania
- Trypanosoma

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

Each gene in the result will belong to an orthology group with the chosen profile, below.  
Be aware that there could be "singletons" in the result: genes that are in a group alone. This will happen if the profile does not force the inclusion of

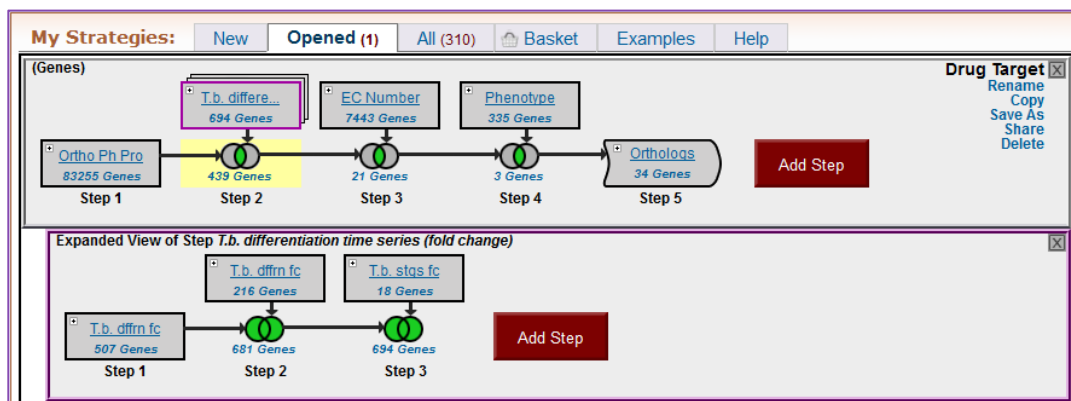
[Get Answer](#)

Select ortholog group profile: Click on the icon to specify which taxa or species to include or exclude in the profile.  
Key:  =no constraints |  =must be in group |  =must not be in group |  =mixture of constraints

- All Organisms
  - Bacteria (BACT)
    - Firmicutes (FIRM)
      - Bacillus anthracis str. Ames Ancestor (bant)
      - Clostridium botulinum A3 str. Loch Maree (cboc)
      - Clostridium perfringens str. 13 (cpez)
      - Listeria monocytogenes EGD-e (Lmon)
      - Staphylococcus aureus subsp. aureus Mu50 (saaz)
      - Streptococcus pneumoniae TIGR4 (spne)
    - Proteobacteria (PROT)
      - alpha-Proteobacteria (PROA)
        - Agrobacterium tumefaciens str. C58 (atum)
        - Brucella suis 1330 (bsu1)

- c. Here is a sample strategy that you may wish to consult. Copy and paste the URL into your browser if the hyperlink does not work.

<http://tritrypdb.org/tritrypdb/im.do?s=e1d4776be3f9b558>



- d. What would happen if you added weights to your strategy? Give it a try! Here is the above strategy with weights added. Copy and paste the link into your browser if the hyperlink does not work.

<http://tritypdb.org/tritypdb/im.do?s=ff340e3a1963dec9>

**My Strategies:** New Opened (1) All (311) Basket Examples Help

(Genes)

Drug Target -- Weighted

Ortho Ph Pro 83255 Genes Step 1

T.b. differe... 694 Genes Step 2

EC Number 7443 Genes Step 3

Phenotype 335 Genes Step 4

Orthologs 7064 Genes Step 5

Add Step

Expanded View of Step T.b. differentiation time series (fold change)

T.b. diffn fc 507 Genes Step 1

T.b. diffn fc 216 Genes Step 2

T.b. stas fc 18 Genes Step 3

Add Step

**My Step Result:**

Filter results by species (results removed by the filter will not be combined into the next step.)

Drug Target -- Weighted - step 5 - 7064 Genes

Add 7064 Genes to Basket | Download 7064 Genes

Genes Genome View (beta)

Gene ID	Organism	Genomic Location	Product Description	Annotated GO Process	Weight
Tb427.01.1930	<i>T. brucei</i> Lister strain 427	Tb427_01_v4: 501,680 - 510,391 (+)	phosphatidylinositol 3-kinase, putative	phosphorylation	180
Tb427.03.3340	<i>T. brucei</i> Lister strain 427	Tb427_03_v4: 856,087 - 858,117 (-)	3', 5'-cyclic nucleotide phosphodiesterase, putative (PDED)	signal transduction	180
Tb427.03.3070	<i>T. brucei</i> Lister strain 427	Tb427_03_v4: 788,084 - 790,813 (-)	3', 5'-cyclic nucleotide phosphodiesterase, putative (PDEC)	signal transduction	180
Tb427.08.2610	<i>T. brucei</i> Lister strain 427	Tb427_08_v4: 767,842 - 770,187 (+)	5-methyltetrahydropteroyltriglutamate-homocysteine S-methyltransferase, putative	methionine biosynthetic process	180
Tb427.07.1610	<i>T. brucei</i> Lister strain 427	Tb427_07_v4: 405,090 - 407,036 (+)	6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase, putative	fructose 2,6-bisphosphate metabolic process	180