

Complex strategy: finding vaccines and drug targets (Exercise 13)

13.1 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://tinyurl.com/6gygmjm) (<http://tinyurl.com/6gygmjm>)

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 if you 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 in your strategy. After doing this, what are some of your top candidates? (hint: you can sort a column in your results 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):

[Weighted Vaccine Strategy](http://tinyurl.com/6dz4zbx) (<http://tinyurl.com/6dz4zbx>)

13.2 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?

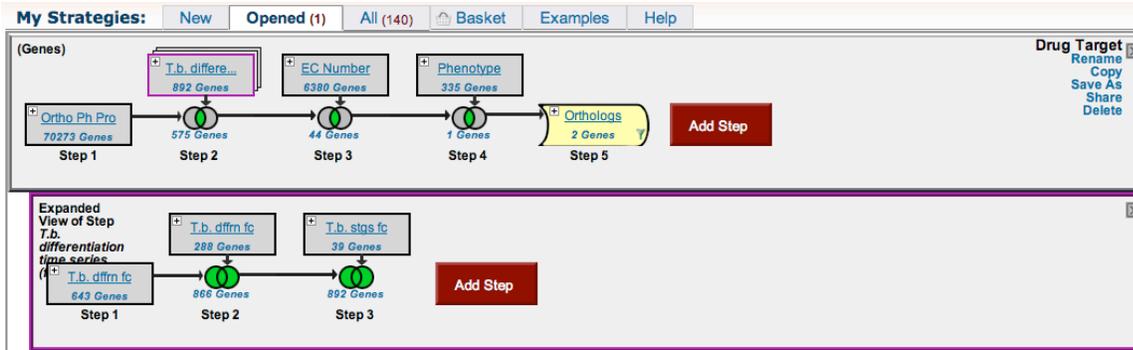
- What about the biology of the parasite you are working on? When would you prefer this protein to be expressed? For example, in *T. brucei* an enzyme may be more active during the slender replicative form. Alternatively, you may decide to concentrate on proteins that may be involved in differentiation 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 (for example) 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 limit 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?

The image shows a screenshot of the TriTrypDB 'Identify Genes by:' interface. On the left, a sidebar lists various search criteria: Text, IDs, Species; Genomic Position; Gene Attributes; Protein Attributes; Protein Features; Similarity/Pattern; Transcript Expression; Protein Expression; Cellular Location; Putative Function; Evolution; Orthology and Paralogy; and Orthology Phylogenetic Profile. The last two options are circled in red, with an arrow pointing from them to a larger red-bordered box on the right. This box contains a list of organisms under the heading 'All Organisms'. The list is categorized into Bacteria (BACT), Firmicutes (FIRM), and Proteobacteria (PROT). Under Bacteria, several species are listed with red 'X' marks, indicating they are excluded: *Bacillus anthracis* str. 'Ames Ancestor' (bant), *Clostridium botulinum* A3 str. Loch Maree (cbot), *Clostridium perfringens* str. 13 (cpez), *Listeria monocytogenes* EGD-e (Lmon), *Staphylococcus aureus* subsp. aureus Mu50 (saaz), and *Streptococcus pneumoniae* TIGR4 (spne). Under Proteobacteria, many species are listed with green checkmarks, indicating they are included: *alpha-Proteobacteria* (PROA) includes *Agrobacterium tumefaciens* str. C58 (atum), *Brucella suis* 1330 (bsui), *Rickettsia prowazekii* str. Madrid E (rpro), and *Rickettsia typhi* str. Wilmington (rtyt). *beta-Proteobacteria* (PROB) includes *Burkholderia mallei* ATCC 23344 (bma1), *Burkholderia pseudomallei* 1710b (bpae), and *Ralstonia solanacearum* GM1000 (rsa1). *delta-Proteobacteria* (PROD) includes *Geobacter sulfurreducens* PCA (gsu1). *gamma-Proteobacteria* (PROG) includes *Coxiella burnetii* RSA 493 (cbuz), *Escherichia coli* str. K-12 substr. W3110 (eco1), *Francisella tularensis* subsp. tularensis SCHU S4 (ftu1), *Salmonella enterica* subsp. enterica serovar Typhi str. CT18 (sent), *Shigella flexneri* 2a str. 301 (sf1e), *Vibrio cholerae* O1 biovar El Tor str. N16961 (vcho), *Yersinia enterocolitica* subsp. enterocolitica 8081 (yent), and *Yersinia pestis* CO92 (ypes). *epsilon-Proteobacteria* (PROE) includes *Campylobacter jejuni* subsp. jejuni NCTC 11168 (cjej) and *Wolbachia succinovorans* DSM 1740 (wsuc).

- c. Here is a sample strategy that you may wish to consult:

<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:

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

Gene	Genomic Location	Product Description	Weight
Tb09.160.5550	Tb927_09_v4: 1,188,157 - 1,190,727 (-)	calpain-like cysteine peptidase, putative,cysteine peptidase, Cian CA, family C2, putative	270
Tb11.01.8820	Tb927_11_01_v4: 4,491,738 - 4,495,499 (+)	expression site-associated gene (ESAG) protein, putative,expression site-associated gene 4 (ESAG4) p...	270
Tb11.01.7460	Tb927_11_01_v4: 4,125,493 - 4,126,407 (-)	NADH dehydrogenase subunit N12M, putative	270
Tb11.01.0930	Tb927_11_01_v4: 2,464,909 - 2,469,021 (+)	protein kinase, putative	270
Tb11.02.4150	Tb927_11_01_v4: 1,744,087 - 1,746,828 (-)	pyruvate phosphate dikinase	270
Tb11.02.3740	Tb927_11_01_v4: 1,661,760 - 1,665,494 (-)	receptor-type adenylyate cyclase GRESAG 4, putative	270
Tb11.02.1480	Tb927_11_01_v4: 1,144,335 - 1,145,891 (+)	mitochondrial processing peptidase alpha subunit, putative,metallo-peptidase, Cian ME, Family M16	270
Tb11.03.0990	Tb927_11_01_v4: 25,133 - 28,648 (-)	expression site-associated gene 4 (ESAG4) protein, putative,receptor-type adenylyate cyclase	270
Tb927.10.15410	Tb927_10_v5: 3,786,037 - 3,787,008 (+)	glycosomal malate dehydrogenase	270
Tb927.10.12500	Tb927_10_v5: 3,030,942 - 3,033,680 (-)	P-type H -ATPase, putative	270
Tb927.10.9760	Tb927_10_v5: 2,418,130 - 2,419,185 (+)	alternative oxidase	270
Tb927.10.6950	Tb927_10_v5: 1,752,628 - 1,753,707 (+)	sterol 24-c-methyltransferase, putative	270
Tb927.10.5940	Tb927_10_v5: 1,489,388 - 1,490,872 (-)	protein kinase, putative,(OTHER) NEK family, HsNEK1-like	270
Tb927.10.2350	Tb927_10_v5: 604,925 - 605,707 (+)	pyruvate dehydrogenase complex E3 binding protein, putative	270
Tb09.211.4880	Tb927_09_v4: 2,200,141 - 2,201,295 (-)	cyclophilin-like protein, putative	270
Tb09.160.4310	Tb927_09_v4: 944,348 - 947,326 (+)	glutamate dehydrogenase	270
Tb927.8.7170	Tb927_08_v4: 2,062,780 - 2,063,952 (+)	inositol polyphosphate 1-phosphatase, putative	270
Tb927.8.6930	Tb927_08_v4: 1,999,223 - 2,000,518 (+)	serine/threonine-protein kinase Nr1A	270
Tb927.8.6510	Tb927_08_v4: 1,883,490 - 1,884,086 (+)	ubiquitin-conjugating enzyme e2, putative,ubiquitin-protein ligase, putative,ubiquitin carrier prote...	270
Tb927.8.1420	Tb927_08_v4: 466,770 - 468,329 (+)	acyl-CoA dehydrogenase, mitochondrial precursor, putative	270
Tb927.7.7500	Tb927_07_v4: 2,173,833 - 2,174,795 (+)	iron/ascorbate oxidoreductase family protein, putative	270
Tb927.7.7470	Tb927_07_v4: 2,147,506 - 2,151,123 (-)	receptor-type adenylyate cyclase GRESAG 4, putative	270
Tb927.7.6850	Tb927_07_v4: 1,916,651 - 1,918,966 (-)	trans-sialidase	270