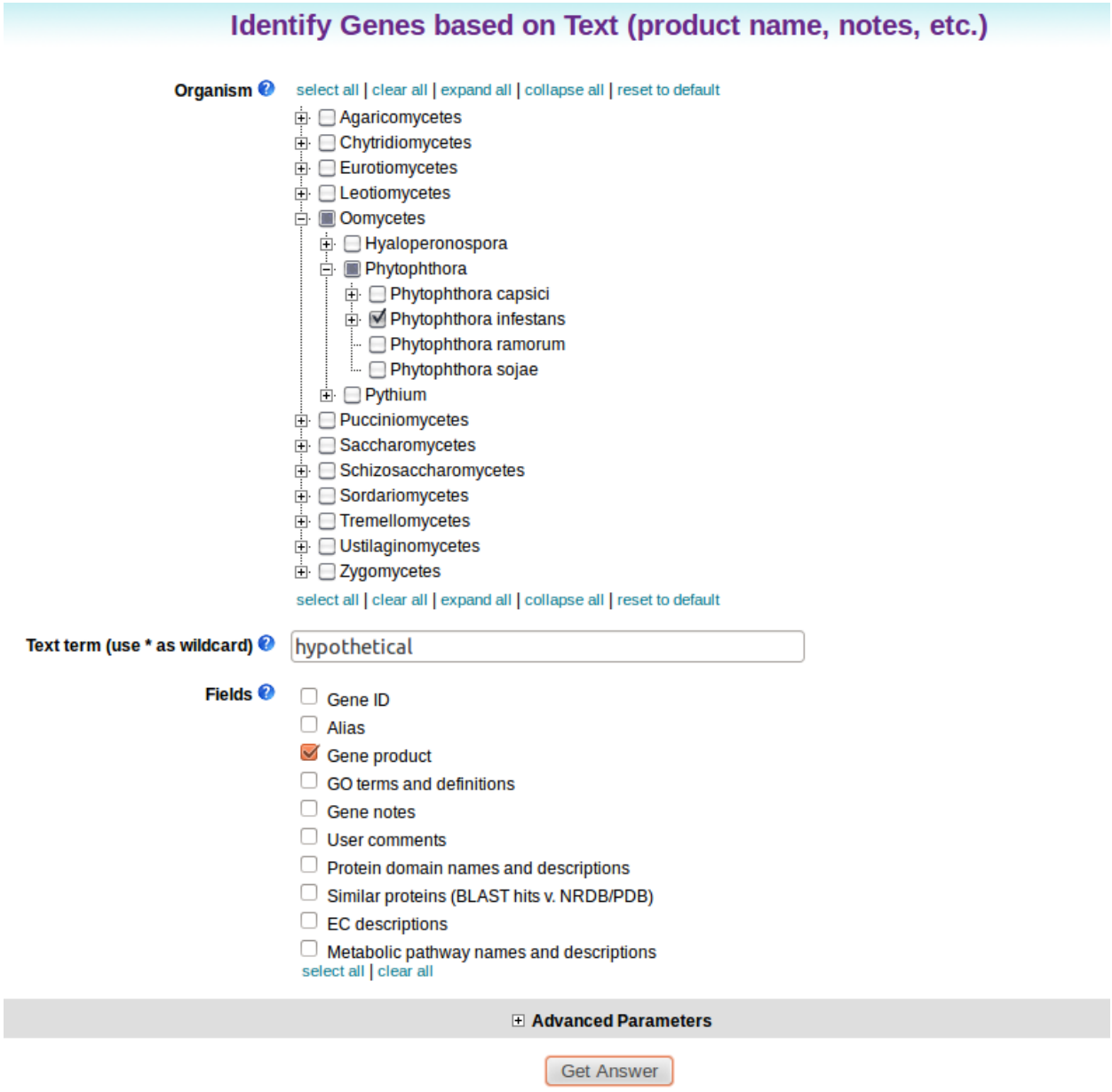
Protein Motif Searches and Regular Expressions
Exercise 6

6.1 Using InterPro domain searches to identify unannotated kinesin motor proteins.

For this exercise use http://fungidb.org

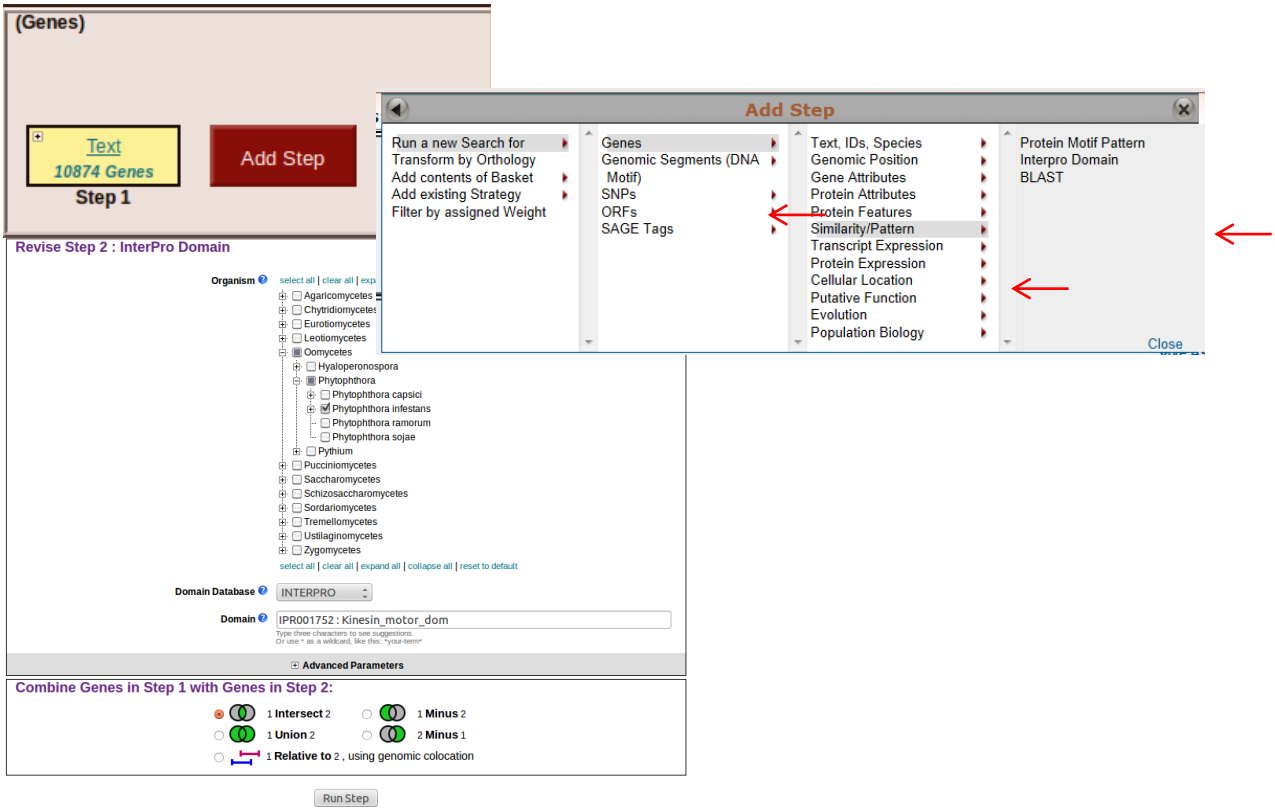
a. Identify all genes annotated as hypothetical in *Phytophthora infestans.*

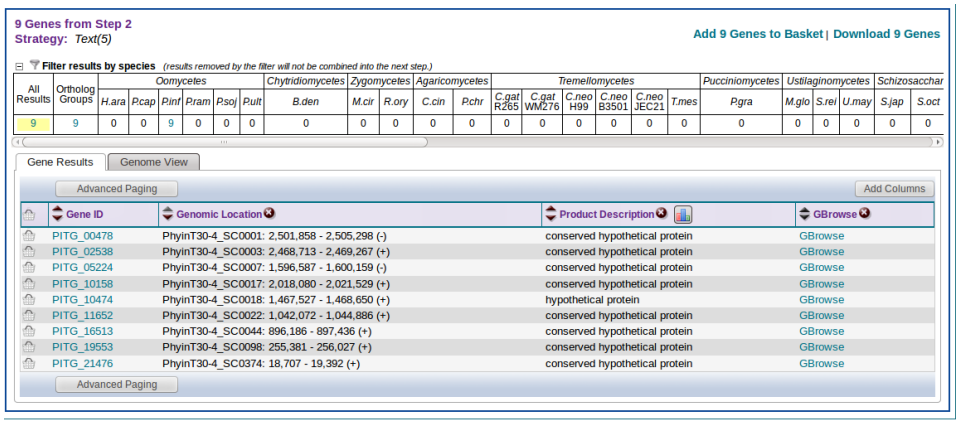
Hint: use the full text search and look for genes with the word “hypothetical” in their product names.



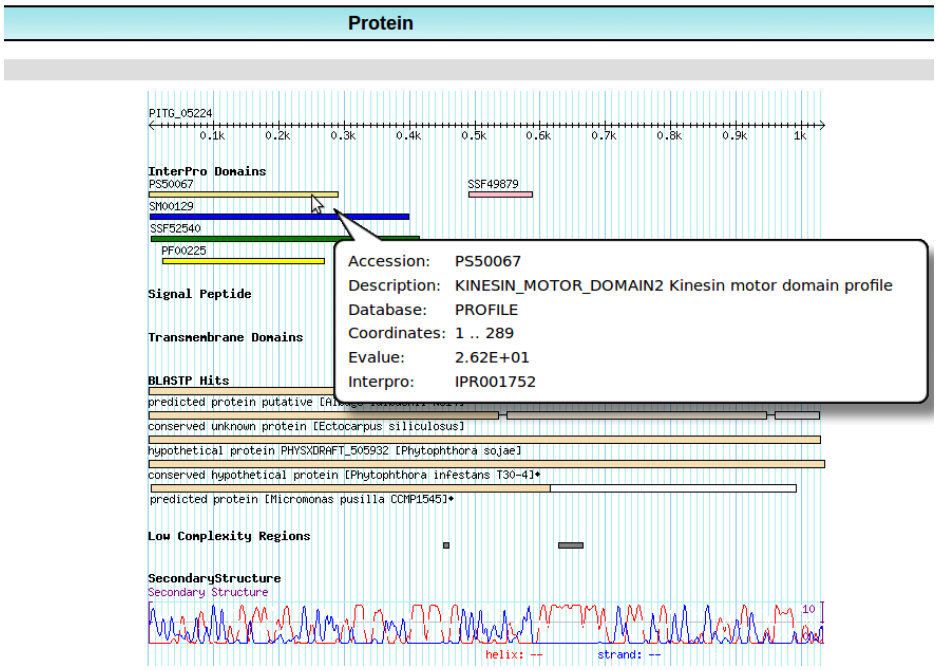
b. How many of these hypothetical genes have a kinesin-motor protein InterPro domain?

Hint: add a step to the strategy. Go to the “Interpro Domain” search under similarity/pattern, start typing the work kinesin and it should autocomplete.





c. Go to the gene page for PITG_05224 and look at the protein feature section. Does this look like a possible motor protein?
Hint: click on the ID for PITG_05224 in the result table to go to the gene page. Mouse over the glyphs in the Protein Features graphic.



6.2 Using regular expressions to find motifs in TriTypDB: finding active trans-sialidases in *T. cruzi.*

a. *T. cruzi* has an expanded family of trans-sialidases. In fact, if you run a text search for any gene with the word “trans-sialidase”, you return over 3500 genes among the strains in the database!!! Try this and see what you get.

b. However, not all of these are predicted to be active. It is known that active trans-sialidases have a signature tyrosine (Y) at position 342 in their amino acid sequence. Add a motif search step to the text search in ‘a’ to identify only the active trans-sialidases.
   - Hint: for your regular expression, remember that you want the first amino acid to be a methionine, followed by 340 of any amino acid, followed by a tyrosine ‘Y’. Refer to regular expression tutorial if you need to.

Add Step 2 : Protein Motif Pattern

Pattern: ^m.{340}y

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

- Leishmania
- Trypanosoma
  - Trypanosoma brucei
  - Trypanosoma congolense
  - Trypanosoma cruzi
  - Trypanosoma evansi
  - Trypanosoma vivax

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

Advanced Parameters

Combine Genes in Step 1 with Genes in Step 2:

1 **Intersect** 2 | 1 **Minus** 2

1 **Union** 2 | 2 **Minus** 1

1 **Relative to** 2 , using genomic colocation

Run Step