

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://tritypdb.org>

a. Identify all genes annotated as hypothetical in *L. braziliensis*.


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


Identify Genes based on Text (product name, notes, etc.)

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

- ☒ Leishmania
 - ☒ Leishmania braziliensis
 - ☐ Leishmania donovani
 - ☐ Leishmania infantum
 - ☐ Leishmania major
 - ☐ Leishmania mexicana
 - ☐ Leishmania tarentolae
- ☐ Trypanosoma

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

Text term (use * as wildcard) 

Fields 

- ☐ Gene ID
- ☐ Alias
- ☒ Gene product
- ☐ Phenotype
- ☐ GO terms and definitions
- ☐ Gene notes
- ☐ User comments
- ☐ Protein domain names and descriptions
- ☐ Similar proteins (BLAST hits v. NRDB/PDB)
- ☐ EC descriptions

[select all](#) | [clear all](#)

[Advanced Parameters](#)

[Get Answer](#)

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 word kinesin and it should autocomplete.

(Genes)

Text
5125 Genes
Step 1

Add Step

Add Step

Add Step 2 : Interpro Domain

Organism select all | clear all | expand all | collapse all

- ☒ Leishmania
 - ☒ Leishmania braziliensis
 - ☒ Leishmania infantum
 - ☒ Leishmania major
 - ☒ Leishmania mexicana
 - ☒ Leishmania tarentolae
- ☐ Trypanosoma

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

Domain Database INTERPRO

Domain kine

Combine Genes in:

- ☒ Union 2
- ☐ Relative to 2, using genomic colocation

Run Step

Add Step

Run a new Search for
Transform by Orthology
Add contents of Basket
Add existing Strategy
Filter by assigned Weight

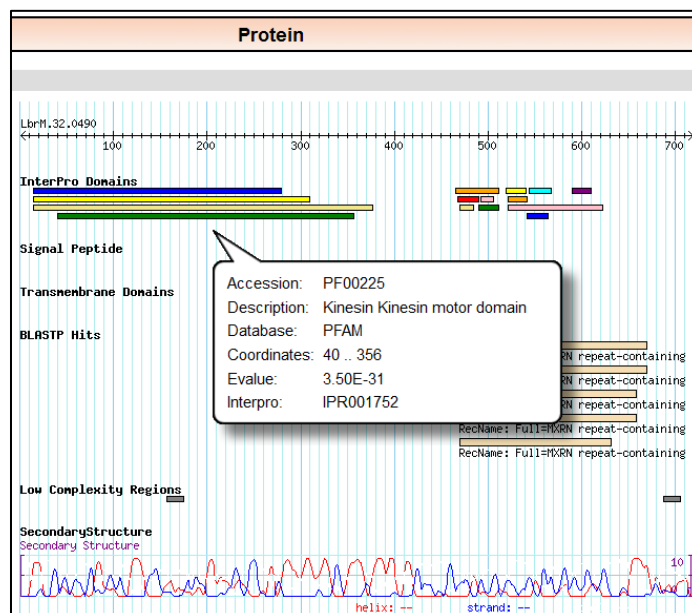
Genes
Genomic Segments (DNA
Motif)
SNPs
ORFs
SAGE Tags

Text, IDs, Species
Genomic Position
Gene Attributes
Protein Attributes
Protein Features
Similarity/Pattern
Transcript Expression
Protein Expression
Cellular Location
Putative Function
Evolution
Population Biology

Protein Motif Pattern
Interpro Domain
BLAST

Close

- c. Go to the gene page for LbrM.32.0490 and look at the protein feature section. Does this look like a possible motor protein?
Hint: click on the ID for LbrM.32.0490 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 TriTrypDB: finding active trans-sialidases in *T. cruzi*.

- 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.
- 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.

If you need help, you can go to this sample strategy below to see the answer:

<http://tritrypdb.org/tritrypdb/im.do?s=a905e36f634f7b42>

Add Step 2 : Protein Motif Pattern

Pattern ?

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 **Union** 2

☐  1 **Minus** 2

☐  1 **Minus** 2

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

Run Step

