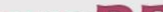


Finding Genes, Building Search Strategies and Visiting a Gene Page

1. Finding a gene using text search.
For this exercise use <http://ToxoDB.org>

- a. Find all possible kinases in *Toxoplasma*.

Hint: use the keyword “kinase” (without quotations) in the “Gene Text Search” box.


ToxoDB
 Toxoplasma Genomics Resource

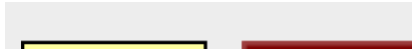
Release 43
 25 Apr 2019

A **EuPathDB** Project

Gene ID:

About ToxoDB | Help | Login | Register | Contact Us

[Home](#)
[New Search](#)
[My Strategies](#)
[My Basket \(0\)](#)
[My Data Sets](#)
[Tools](#)
[Data Summary](#)
[Downloads](#)
[Community](#)
[Analyze My Experiment](#)
[My Favorites](#)

- How many genes did you get?
 - Look closely at the sections of the result page. How many of those are in *T. gondii* ME49? How did you find this out?
- (Hint – the filter table is located between the strategy panel and the result table and shows the distribution of results across the organisms that you searched. Click on a number to ‘filter’ the result and display results from a specific species or strain, note that yellow highlighting indicates what you are currently viewing on the webpage).
- 
- The screenshot shows a portion of a web application. On the left, there is a yellow rectangular button with a black border. Inside the button, the text 'Text' is in black, '7137 Genes' is in blue, and 'Step 1' is in black. To the right of this button is a red rectangular button with the text 'Add Step' in white.

My Strategies:

New

Opened (1)

All (1)

Basket

Public Strategies (14)

Help

Hide search strategy panel

Genes

Text

313 Genes

Step 1

Add Step

Strategy: Text *

Rename

Duplicate

Save As

Share

Delete

313 Genes from Step 1

Revise

Strategy: Text

Click on a number in this table to limit/filter your results

| All Results | Ortholog Groups | Cyclospora | Cystoisospora | Eimeria | | | | | | | | | | Hammondia | Neospora | Sarcocystis | Toxoplasma | | | | | | | | | |
|-------------|-----------------|------------------|---------------|--------------|------------|----------------------|-----------|----------|------------|-----------|-----------------|---------------|-----------|-----------------|-----------------|-------------|------------|--------------------|-----|-----|------|----|-----|----|--|--|
| | | C.cayentanensis | C.suis | E.acervulina | E.brunetti | E.falciformis | E.maxima | E.mitis | E.necatrix | E.praecox | E.tenella | H.hammondi | N.caninum | S.neurona (416) | T.gondii (3869) | | | | | | | | | | | |
| | | strain CHN_HEN01 | strain Wien I | Houghton | Houghton | Bayer Haberkorn 1970 | Weybridge | Houghton | Houghton | Houghton | strain Houghton | strain H.H.34 | Liverpool | SN3 | SO SN1 | ARI | FOU | GAB2-2007-GAL-DOM2 | GT1 | MAS | ME49 | RH | RUB | Tg | | |
| 7137 | 891 | 246 | 315 | 239 | 209 | 222 | 187 | 203 | 243 | 161 | 257 | 293 | 277 | 202 | 214 | 387 | 381 | 329 | 311 | 373 | 313 | 0 | 383 | | | |

Gene Results

Genome View

Analyze Results

1

2

3

...

16

Rows per page: 20

Download

Add to Basket

Add Columns

| Gene ID | Transcript ID | Organism | Genomic Location (Gene) | Product Description | Found in | Score |
|---------------|---------------------|----------------|---------------------------------------|---|--|-------|
| MGME49_252360 | TGME49_252360-126_1 | T. gondii ME49 | TGME49_chrIII:512,377..515,416(+) | rhophy kinase family protein ROP24 (incomplete catalytic triad) | User Comments, GOTerms, InterPro, Product, Notes | 33 |
| MGME49_266100 | TGME49_266100-126_1 | T. gondii ME49 | TGME49_chrIX:1,304,347..1,307,417(-) | rhophy kinase family protein ROP41 | User Comments, Product, InterPro, Notes, GOTerms | 33 |
| MGME49_240090 | TGME49_240090-126_1 | T. gondii ME49 | TGME49_chrVI:1,057,698..1,061,688(+) | rhophy kinase family protein ROP34, putative | User Comments, GOTerms, InterPro, Notes, Product | 33 |
| MGME49_201130 | TGME49_201130 | T. gondii ME49 | TGME49_chrIII:1,010,044..1,000,000(+) | rhophy kinase family protein ROP20 | User Comments, InterPro, | 33 |

- Do you believe that these genes are kinases? Find the Product Description in the Gene Result tab. Can you presume the gene encodes a kinase just by looking at the name?
- What happens if you search using the term **kinases** (the plural word) in the Gene Text Search box? How many results are returned?

- b. Find only the kinases that specifically have the word “kinase” in the gene product name.

The search you ran in step 1a using the Gene Text Search box initiates a preconfigured search. Initiating the search from the Text Search form - **Identify Genes based on Text**, allows you to configure the search yourself, choosing parameters that best meet your needs. Use the Text Search form to search for genes that have the word kinase in their **gene product** name/description. Note that you can also revise the search from step 1a and configure the search parameters as described below.

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

Search for Genes

expand all | collapse all

Find a search...

Text

- Text (product name, notes, etc.)
- Gene models
- Annotation, curation and identifiers
- Genomic Location
- Taxonomy
- Orthology and synteny
- Phenotype
- Genetic variation
- Epigenomics
- Transcriptomics
- Sequence analysis
- Structure analysis
- Protein features and properties
- Protein targeting and localization
- Function prediction
- Pathways and interactions
- Proteomics
- Immunology

expand all | collapse all

Organism

26 selected, out of 26

Filter list below...

- ☒ Cyclospora
- ☒ Cystoisospora
- ☒ Eimeria
- ☒ Hammondia
- ☒ Neospora
- ☒ Sarcocystis
- ☒ Toxoplasma

select all | clear all | expand all | collapse all

Text term (use * as wildcard)

kinase

Fields

- ☐ Alias
- ☐ EC descriptions
- ☐ Gene ID
- ☐ Gene notes
- ☒ Gene product
- ☐ Gene name
- ☐ GO terms and definitions
- ☐ Metabolic pathway names and descriptions
- ☐ Protein domain names and descriptions
- ☐ PubMed
- ☐ Similar proteins (BLAST hits v. NRD/Protein)
- ☐ User comments

select all | clear all

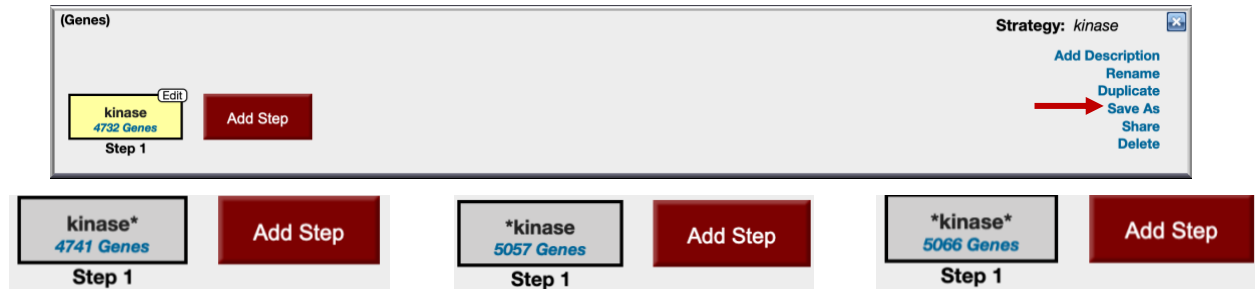
Advanced Parameters

Get Answer

kinase

Give this search a weight (optional)

- There are several ways to navigate to the **Identify Genes based on Text** page: home page ‘Search for Genes’ panel and the ‘New Search’ drop down menu. Notice the sections of the search page. At the top are parameters and the Get Answer button followed by a search description and a list of datasets used by the search.
- How can you make sure to find your text term in plural form or in compound words like “kinases” or “6-phosphofructokinase”? Adding a wild card (wildcard = asterisk * which means any character) in your search term will broaden your search. Go to the text search page, the specific page where you can define the fields to be searched and make sure to only select the field called Gene Product. Repeat this search with the words: **kinase *kinase kinase* *kinase***

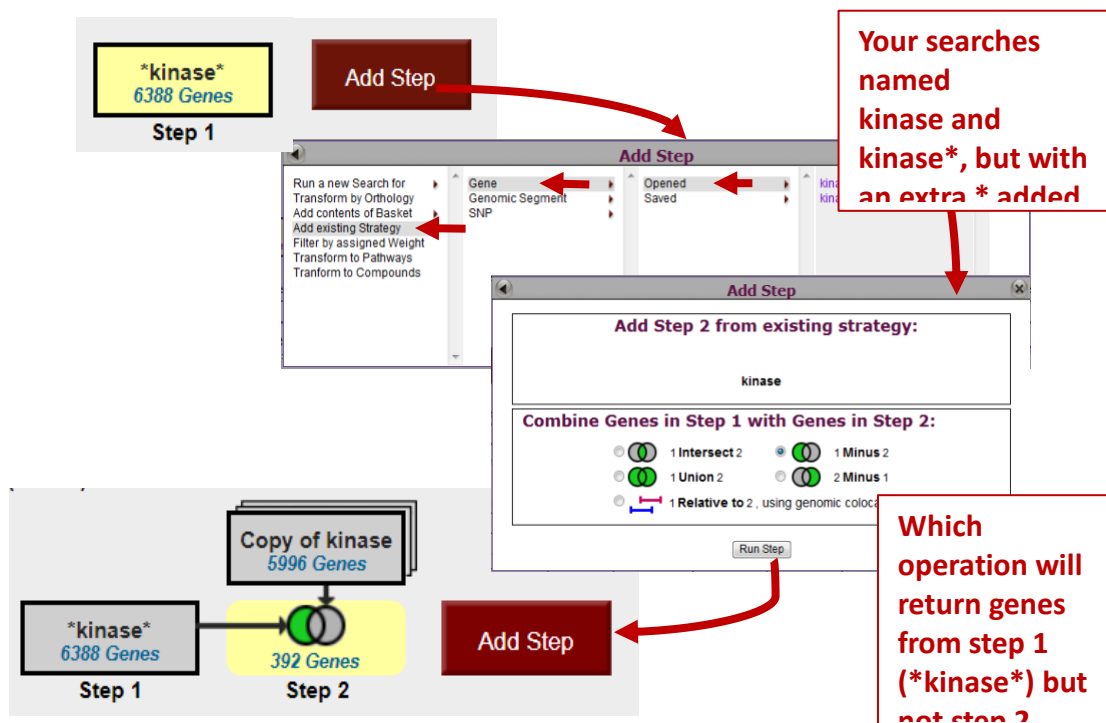


- Give each new search a name to help you keep track of the searches. Searches can also be saved using Save As in the strategy action items.
- How did you get to the Text Search page?
- How does limiting the number of fields searched affect your results?
- Did you remember to use the wild card?
- How many genes have the word kinase in their product names?

c. Combine the results of two text searches.

Find genes that were identified using the key word ***kinase*** but not the word **kinase**?

- Here we will build a search strategy that combines 2 of your searches. If you are not displaying the results of the ***kinase*** search (the strategy box will be highlighted in yellow), return to it by clicking on that step box in the strategy panel. To add your **kinase** search to this strategy, click on “Add Step” and select “existing strategy”.
- Select the correct strategy from your list of Gene Strategies and combine the strategies with the correct operation. Notice that there is an extra asterisk at the end of an unsaved strategy name. The list of available searches will have an * at the end of the name.



- Do the results make sense? Do all the product names contain the word kinase? From the result page look at the Gene Result Tab with the table of gene IDs returned by the search. The Product Description column contains the gene product name.

2. Combing text search results with results from other searches

a. Find kinase genes that are likely secreted.






In exercise 1b. you identified genes that have the word **kinase** somewhere in their gene product name (searching ***kinase*** in gene product field). Grow your search strategy by adding a step that returns genes whose protein products are predicted to have a signal peptide. In this search you are querying the results of our genome-wide analysis that used the SignalP program to predict the presence and location of signal peptide cleavage sites in amino acid sequences. <http://www.cbs.dtu.dk/services/SignalP/>

Focus your Strategies section on the ***kinase*** search and click Add Step. For the second search choose **Identify Genes based on Protein Targeting, Predicted Signal Peptide**

- How did you combine the search results?
- How many kinases are predicted to have a signal peptide?

The screenshot shows a workflow interface with two steps. Step 1 is a search for ***kinase*** resulting in 5066 Genes. An **Add Step** button is next to it. A red arrow points from this button to a modal window titled **Add Step**. This window lists various search categories. Under **Genes**, **Predicted Signal Peptide** is selected. A red arrow points from this selection to a second modal window titled **Add Step 2 : Predicted Signal Peptide**. This window shows 26 selected organisms out of 26, including Cyclospora, Cryptosporidium, Eimeria, Hammondia, Neospora, Sarcocystis, and Toxoplasma. Below this, the **Combine Genes in Step 1 with Genes in Step 2:** section shows several set operation icons. A red arrow points from the **Intersection** icon (two overlapping circles) to a third modal window. This window shows the combined workflow: Step 1 (*kinase* 5066 Genes) followed by Step 2 (Signal Pep 51001 Genes), resulting in 921 Genes. A red arrow points from this final result to a text box.

Which operation will return genes that are in both search result sets?

| Operator | : | Combined Result will contain: |
|---|---|--|
|  1 INTERSECT 2 | : | IDs in common between the two lists |
|  1 UNION 2 | : | IDs from list 1 and list 2 |
|  1 MINUS 2 | : | IDs unique to 1 |
|  2 MINUS 1 | : | IDs unique to 2 |
|  1 Relative to 2 | : | IDs whose features are near each other (colocated) in the genome |

b. Now that you have a list of possible secreted kinases, expand this strategy even further.

There is no wrong answer here!!

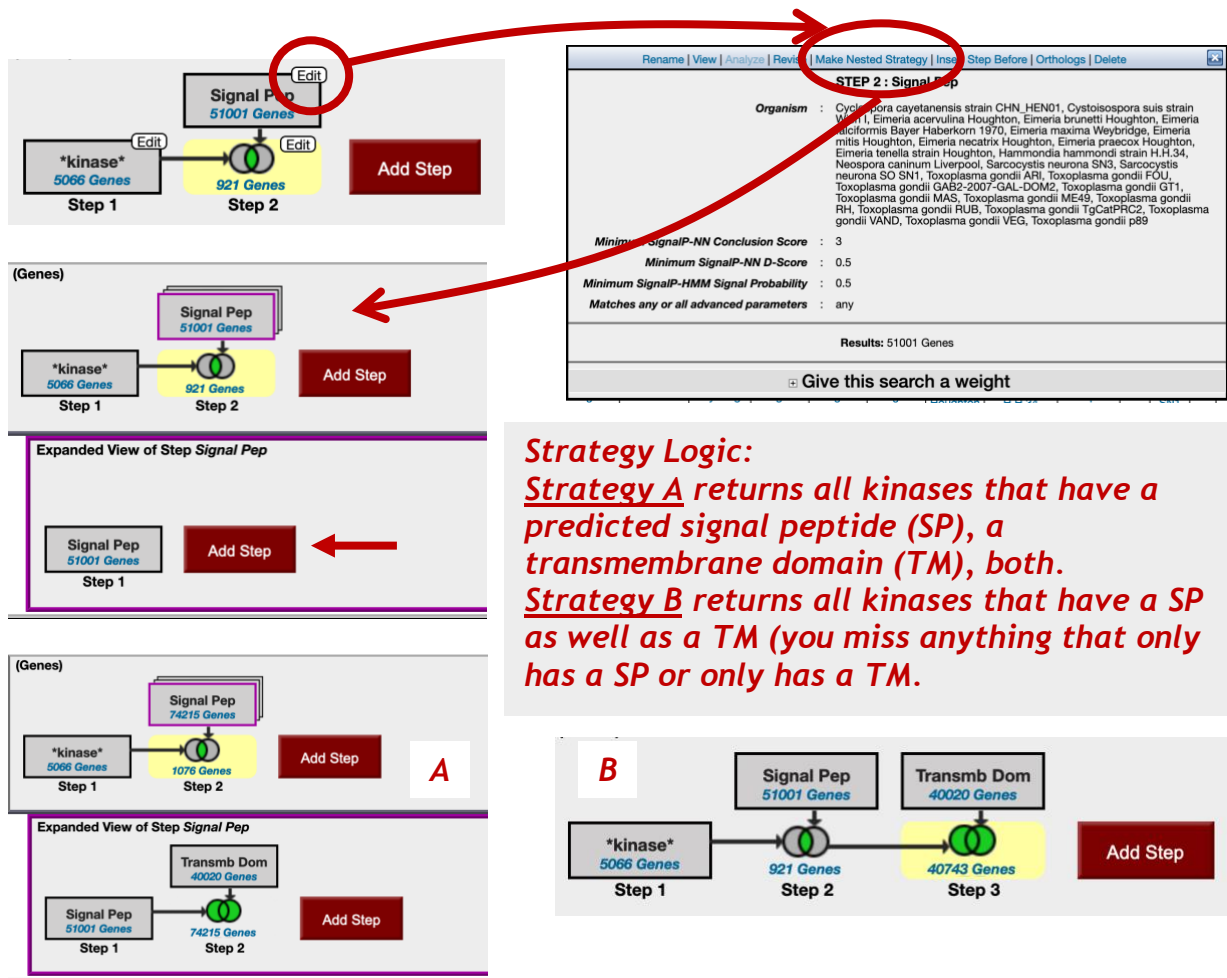
- From a biological standpoint what else would be interesting to know about these kinases? Add more searches to grow this strategy. Open the categories under Identify Genes By on the home page and explore the types of searches that are available. You can reduce (or expand) your result set by adding searches that are based on many types of data.
- For example, how many of the secreted kinases also have transmembrane domains?

c. In the above example, how can you define kinases that have either a secretory signal peptide AND/OR a transmembrane domain(s)?

Hint: to do this properly you will have to employ the “Nested Strategy” feature. Nesting a strategy allows you to control the order in which your result sets are combined. Think about the difference between two mathematical equations.

Equation without nesting: $2 \times 3 + 5 = 11$

Equation with nesting: $2 \times (3 + 5) = 16$



3. Finding a gene by BLAST Similarity.

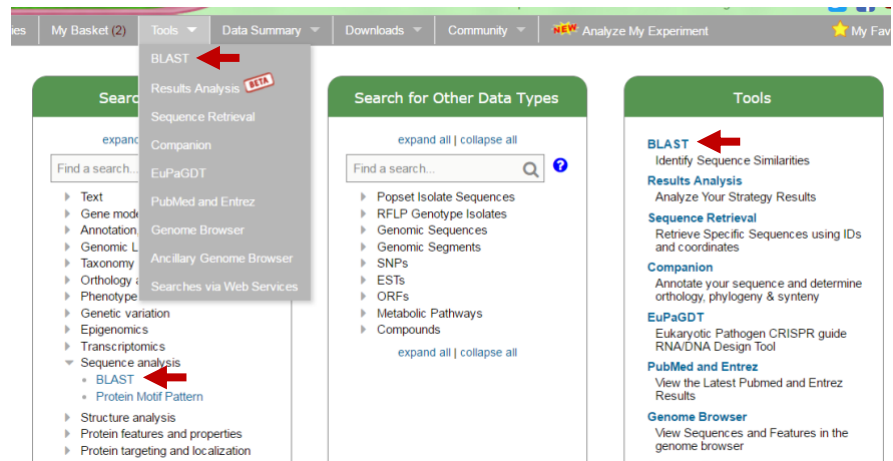
Note: For this exercise start with <http://toxodb.org/toxo/>

Imagine that you generated an insertion mutant in *Toxoplasma* that is providing you with some of the most interesting results in your career! You sequence the flanking region and you are only able to get sequence from one side of the insertion (the sequence shown below). You immediately go to ToxoDB to find any information about this sequence. What do you do?

```
aaaggagagaaagataaaaatatacaaaggtccccagagacacgatagtgttactgacaa
catacagaatcaggtcgagcaatggaagaaccaagcaccggcgccagagattgaactcgc
ttggattccgtagcggtttatgagttgatagcttggtcttaaaaaaacaaggctgaaaa
atggaaaaaatgtctccaat
```

- Sequence is also available from this URL: <http://tinyurl.com/ex1blast>
- Navigate to the BLAST search and run the search with this sequence. The BLAST search will return records for sequences that are similar to your input sequence.

- Which BLAST program should you use? (hint: try different BLAST programs, just keep in mind that you have a nucleotide sequence so you must use an appropriate BLAST program).



Note on BLAST programs:

- **blastp** compares an amino acid sequence against a protein sequence database;
- **blastn** compares a nucleotide sequence against a nucleotide sequence database;
- **blastx** compares the six-frame conceptual translation products of a nucleotide sequence (both strands) against a protein sequence database;
- **tblastn** compares a protein sequence against a nucleotide sequence database dynamically translated in all six reading frames (both strands);
- **tblastx** compares the six-frame translations of a nucleotide sequence against the six-frame translations of a nucleotide sequence database.
- Are you getting any results from blastx? tblastn? What about blastn?
- What is your gene? (hint: after running a blastn against *Toxoplasma* ME49 (Target organism) genomic sequence (Target Data Type), click on the "link to the genome browser". In the genome browser zoom out to see what gene is in the area).

Target Data Type

☒ Transcripts
☐ Proteins
☐ Genome
☐ EST
☐ ORF
☐ PopSet

BLAST Program

☒ blastn
☐ blastp
☐ blastx
☐ tblastn
☐ tblastx

Target Organism

1 selected, out of 25

Filter list below...

☐ Cyclospora
☐ Cystoisospora
☐ Eimeria
☐ Hammondia
☐ Neospora
☐ Sarcocystis
☒ Toxoplasma

select all | clear all | expand all | collapse all

Input Sequence

ttggattgcctagcgtttatgagttgatagcttgcctctaaaaaacaa
 gcctgaaaa
 atggaaaaaatgtctcaat

Note: only one input sequence allowed.
maximum allowed sequence length is 31K bases.

Expectation value

10

Maximum descriptions/alignments (V=B)

50

Low complexity filter

no

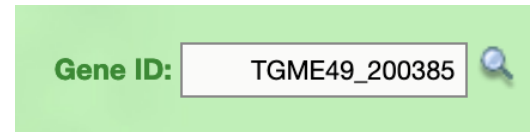
Get Answer

4. Viewing data on a gene page.

Note: For this exercise use <http://toxodb.org>

a. Find the gene page for TGME49_200385

- The easiest way to get to a gene page when you know the gene ID is to use the Gene ID search.
- Examine the information at the top of the gene page:
 - What is the gene name?
 - What chromosome is this gene on?
 - Explore the “shortcuts” section at the top of the gene page – try clicking on the magnifying glass. This option opens a preview of various sections of the gene page for quick access. Clicking the image itself will take you to that section of the gene page.



A screenshot of the ToxoDB website showing the gene page for TGME49_200385. The page title is "TGME49_200385 Myb family DNA-binding domain-containing protein". Below the title, there is a "Shortcuts" section with icons for Synteny, BLAT Alignments, Phenotype, SNPs, Transcriptomics, Protein Features, and Proteomics. A red arrow points from the "SNPs" icon to a zoomed-in view of the SNPs section. The zoomed-in view shows the gene's location on chromosome VIII (6832k to 6849k) and a list of SNPs by coding potential. The page also includes a "Contents" sidebar on the left and a "1 Gene models" section at the top.

- Examine the “Gene Models” section of the gene page.
 - How many exons does this gene have?
 - How many transcripts does this gene encode?
 - What direction are the transcripts relative to the chromosome?
 - What does the “RNA Evidence for introns” information mean?
 - From what type of data are the “introns” determined?
 - How many nucleotides is the largest transcript? (hint: examine the transcripts table underneath the gene models).
 - Try out the ‘View in GBrowse genome browser’ and ‘View in JBrowse genome browser’ buttons. Where do these links take you? What advantage do these genome browsers offer that the static images on the gene pages do not?

1 Gene models

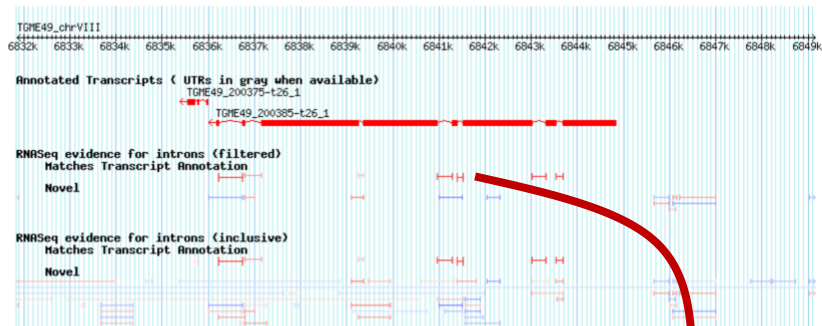
Exons in Gene 8

Transcripts 1

▼ Gene Models

[View in GBrowse genome browser](#)

[View in JBrowse genome browser](#) BETA



[View in GBrowse genome browser](#)

[View in JBrowse genome browser](#)

▼ Transcripts [Download](#) [Data sets](#)

| Transcript | # exons | Transcript length | Protein length |
|---------------------|---------|-------------------|----------------|
| TGME49_200385-t26.1 | 8 | 6777 | 2258 |

| Track details | | | | | |
|--|--|--------|--------|---------|-------|
| Intron Location: | 6840968 - 6841282 (315 nt) | | | | |
| Intron Spanning Reads (ISR): | 2281 | | | | |
| ISR per million (ISRPM): | 1578.65 | | | | |
| Gene assignment: | TGME49_200385 - annotated intron | | | | |
| % of Most Abundant Intron (MAI): | 58.4 | | | | |
| Most abundant in: | Bradyzoite in vitro Transcriptome (ME49): bradyzoite | | | | |
| ISRPM (ISR / gene coverage) | 104.61 (.68) | | | | |
| Experiment | Sample | Unique | ISRPM | ISR/Cov | % MAI |
| Bradyzoite in vitro Transcriptome (ME49) | bradyzoite | 186 | 104.61 | .68 | 22.1 |
| Bradyzoite in vivo transcriptome (M4) | bradyzoite | 130 | 43.1 | .468 | 46.4 |
| Feline enterocyte, tachyzoite, bradyzoite stage transcriptome | EES5 | 1 | .05 | .0998 | 2.9 |
| Oocyst Time Series (M4) | Tachyzoites | 113 | 20.51 | .4697 | 49.1 |
| | Tissue cysts | 37 | 59.83 | .8722 | 42.5 |
| | day 4 | 3 | 2.41 | .7824 | 8.1 |
| Tachyzoite Transcriptome 3 and 4 days post-infection (VEG NeLiv) | day 10 | 6 | 5.06 | 1.0133 | 31.6 |
| | day 3 | 121 | 57.64 | .5232 | 68.4 |
| Tachyzoite Transcriptome Time Series (GT1) | day 4 | 230 | 89.79 | .6257 | 89.1 |
| Tachyzoite Transcriptome Time Series (ME49) | hour 16 | 84 | 27.9 | .5081 | 24.1 |
| | hour 2 | 4 | 17.99 | 1.6507 | 80 |

| RNASeq evidence for introns (filtered) | |
|---|--------------------|
| Click here to display in new window... | |
| Intron Spanning Reads (ISR): The total number of uniquely mapped reads (all samples) which map across the junction and are on the appropriate strand. GSNAP uses splice site consensus sequences to determine strand of the mapped read. | |
| ISR per million (ISRPM): Intron Spanning Reads Per Million intron spanning reads and thus represents a normalized count of unique reads. | |
| % of Most Abundant Intron (MAI): The percentage (ISRPM of this junction / ISRPM of maximum junction for this gene) of this junction over the maximum for this gene. | |
| Most abundant in: The experiment and sample that has the highest ISRPM for this gene. | |
| ISRPM, (ISR / coverage): ISRPM from sample with highest ISRPM and the ISR/coverage for that same sample. | |
| The table shows all experiments and samples that provide evidence for this intron junction. Note that the values for each row are based on each specific sample. The color of the glyph changes with the Score as follows: | |
| Reverse | Forward |
| 5-16 | 5-16 |
| 17-64 | 17-64 |
| 65-256 | 65-256 |
| 257-1024 | 257-1024 |
| 1025-4096 | 1025-4096 |
| 4097-16000 | 4097-16000 |
| greater than 16000 | greater than 16000 |

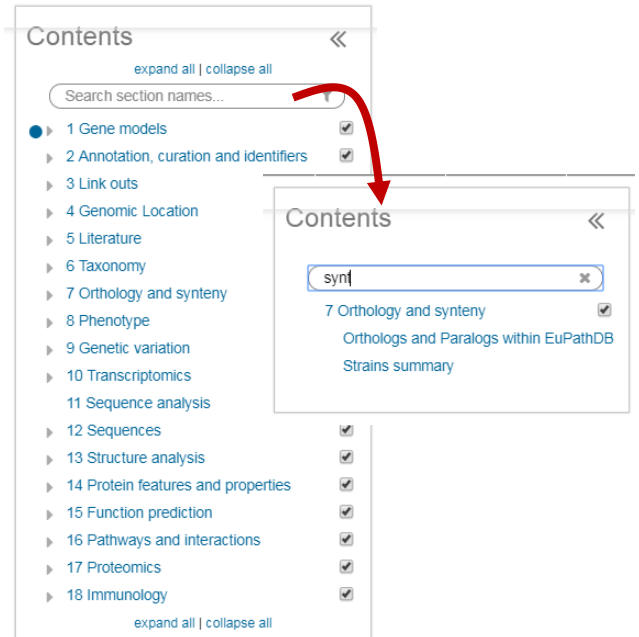
b. What does the synteny of this gene look like? How did you find/navigate to this section? (hint: you can use the “Contents” menu on the left side of the gene page to find/navigate to the different sections. You can also click on the images in the Shortcuts section to navigate to the image within the data section of the page).

- Is synteny (chromosome organization) in this region maintained in other species? Hint: compare gene organization between the different species in the synteny section.
- What does the shading between genes indicate?
- What does synteny look like across the entire chromosome? To do this:

- Click on the “View in Genome Browser” button right under the synteny section

View in genome browser

- on the gene page.
- Zoom out to the entire chromosome. There are a few ways to do this. For example, drag your cursor across the entire chromosome in the Overview panel and then select “zoom” from the popup menu (this may take a minute to load).
- For each genome notice that there are two tracks: one called genes and the other contig. Which genome is composed of the most fragments? Are there any other interesting observations you can support by looking at synteny over large genomic regions?



c. Run a multiple sequence alignment comparing the protein sequence for this gene from *Hammondia hammondi*, *Neospora caninum* and *Toxoplasma gondii* GT1 (note the ME49 sequence is included by default since we are on the ME49 gene page).

- Scroll up to the Orthologs and Paralogs within EuPathDB table. This table also functions as a multiple sequence alignment tool. Use the first column to check the

strains you want to include in the alignment and then use the parameters at the bottom of the table to configure and run the alignment.

Ortholog Group [OG5_127030](#)

▼ Orthologs and Paralogs within EuPathDB [Data sets](#)

To run Clustal Omega, select genes from the table below. Then choose the sequence type and initiate the alignment with the 'Run Clustal Omega for selected genes' button.

Search this table... Showing 18 rows

| Clustal Omega | Gene | Organism | Product | is syntenic | has comments |
|-------------------------------------|-------------------------------|--------------------------------------|--|-------------|--------------------|
| <input checked="" type="checkbox"/> | HHA_200385 | Hammondia hammondi strain H.H.34 | Myb family DNA-binding domain-containing protein | yes | no |
| <input checked="" type="checkbox"/> | NCLIV_038230 | Neospora caninum Liverpool | Os10g0478300 protein, related | yes | no |
| <input type="checkbox"/> | SN3_02400135 | Sarcocystis neurona SN3 | myb family dna-binding domain-containing protein | yes | no |
| <input type="checkbox"/> | TGARI_200385 | Toxoplasma gondii ARI | Myb family DNA-binding domain-containing protein | yes | no |
| <input type="checkbox"/> | TGDOM2_200385 | Toxoplasma gondii GAB2-2007-GAL-DOM2 | Myb family DNA-binding domain-containing protein | yes | no |
| <input type="checkbox"/> | TGFOU_200385 | Toxoplasma gondii FOU | Myb family DNA-binding domain-containing protein | yes | no |
| <input checked="" type="checkbox"/> | TGGT1_200385 | Toxoplasma gondii GT1 | Myb family DNA-binding domain-containing protein | yes | no |
| <input type="checkbox"/> | TGMA5_200385 | Toxoplasma gondii MAS | Myb family DNA-binding domain-containing protein | yes | no |
| <input type="checkbox"/> | TGP89_200385 | Toxoplasma gondii p89 | Myb family DNA-binding domain-containing protein | yes | no |
| <input type="checkbox"/> | TGPRC2_200385 | Toxoplasma gondii TgCatPRC2 | Myb family DNA-binding domain-containing protein | yes | no |
| <input type="checkbox"/> | TGRUB_200385 | Toxoplasma gondii RUB | Myb family DNA-binding domain-containing protein | yes | no |

- Can you find areas of the protein that are more conserved between the aligned sequences?

d. Does this gene contain Single Nucleotide Polymorphisms (SNPs)?

In gene pages, SNPs are represented in a section called “Genetic variation”. This section includes an isolate alignment tool for displaying SNPs between chosen isolates and a DNA polymorphism browser with textual and graphical SNP representations.

Contents

genetic

- 9 Genetic variation
 - 9.1 DNA polymorphism
 - [Isolate Alignments in this Gene Region](#)

- Examine the DNA polymorphism section 9.1.
 - What is the total number of SNPs in the gene?
 - How many SNPs impact the predicted protein sequence?
 - Is this likely to define the full spectrum of sequence variation in this gene?
 - What do the different color diamonds in the browser view signify? (Hint: move your cursor over a diamond – without clicking - to get more information in a popup).

e. Is this gene expressed at the protein and/or transcript level?

Look at the gene page sections entitled “Proteomics” and “Transcriptomics”. You can use the contents panel to navigate to those sections. Or you can return to the top of the page with the ‘back to top button’ and then click on the ‘Shortcut’ image to navigate to that section of the page.



- What kinds of data in ToxoDB provide evidence for protein expression? (Hint, view the Mass Spec.-based Expression Evidence table).

17 Proteomics

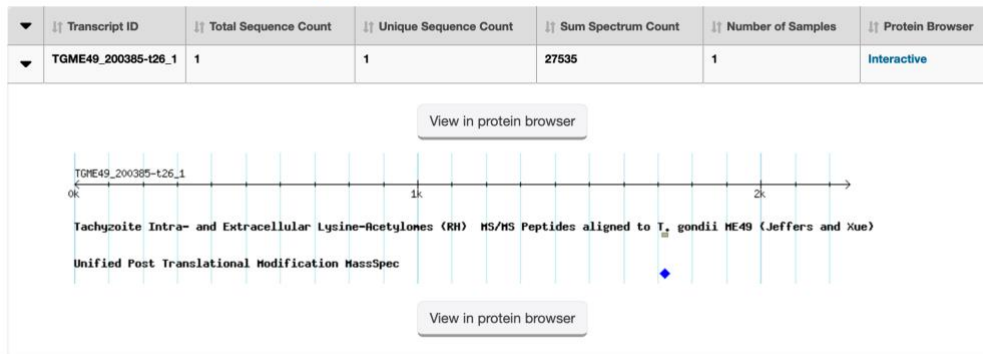
▼ Mass Spec.-based Expression Evidence

[Data sets](#)

| Transcript ID(s) | Experiment | Sample | Sequences | Spectra |
|---------------------|--|-------------------------|-----------|---------|
| TGME49_200385-t26_1 | Tachyzoite Intra- and Extracellular Lysine-Acetylomes (RH) | extracellular acetylome | 1 | 27535 |

▼ Mass Spec.-based Expression Evidence Graphic

[Data sets](#)



▼ Post Translational Modifications

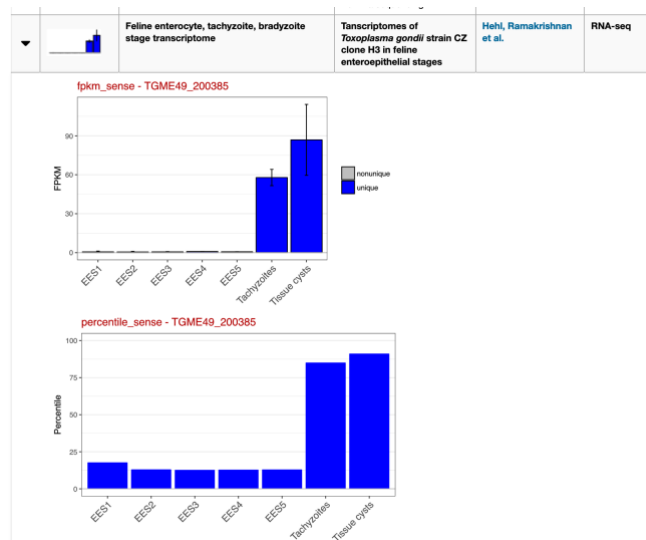
[Download](#)

[Data sets](#)

| Transcript ID(s) | Experiment | Sample | Modification Type | Modification Site | Residue |
|---------------------|--|-------------------------|-------------------|-------------------|---------|
| TGME49_200385-t26_1 | Tachyzoite Intra- and Extracellular Lysine-Acetylomes (RH) | extracellular acetylome | acetylation site | 1720 | K |

- Does it contain any post-translational modifications?
- Can you quickly link to the data set record for proteomics experiments?
- How abundant is this protein? How confident are you of this analysis? Abundance can be estimated by counting the number of spectra supporting a peptide. Where do you find information about the number of spectra?

- What about the transcriptomic data?
Navigate to this section of the gene page and explore the various experiments. Can you draw a conclusion about when this gene is expressed? For example, examine the experiment called “Feline enterocyte, tachyzoite, bradyzoite stage transcriptome”. Make sure to also turn on the percentile graph – why is this useful?

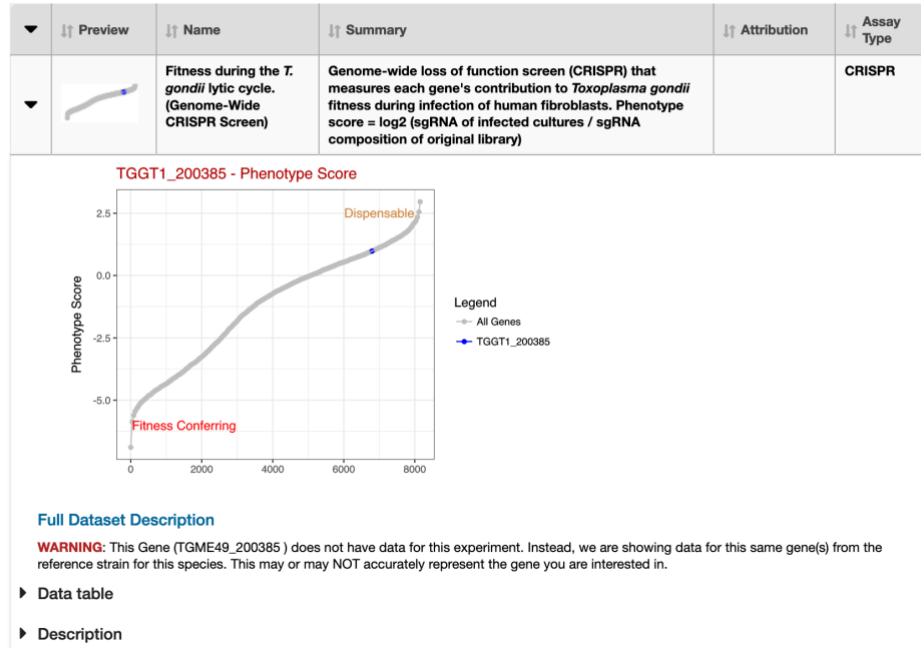


- How does the RNAseq data compare with the microarray data?

- f. Is this gene essential in *Toxoplasma*? Does mutating the gene reduce fitness?
- Navigate to the phenotype section and examine the CRISPR screen data

8 Phenotype

▼ CRISPR Phenotype Graphs [Data sets](#)



- What can you say about the essentiality of this gene? What other information do you need to think about when looking at these results?