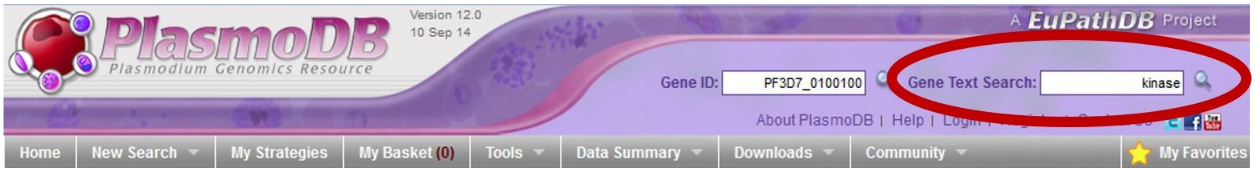
FINDING GENES, BUILDING SEARCH STRATEGIES AND VISITING A GENE PAGE

1. Finding a gene using text search.

For this exercise use http://www.plasmodb.org

1a. Find all possible kinases in *Plasmodium*.

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

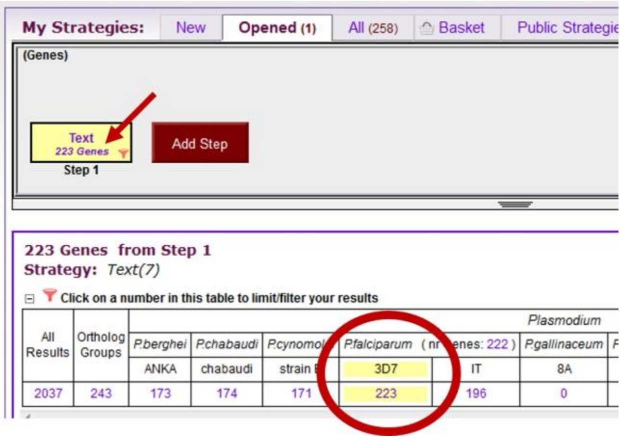


Gene ID: PF3D7_0100100

Gene Text Search: kinase

- How many genes did you get?
- Look closely at the sections of the result page. How many of those are in *P. falciparum*? 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 display on that species’ portion of the results.



223 Genes from Step 1

Strategy: *Text(7)*

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

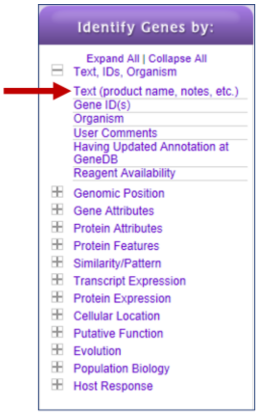
		Plasmodium							
All Results	Ortholog Groups	*P.berghei*	*P.chabaudi*	*P.cynomol*	*P.falciparum* (nr genes: 222)			*P.gallinaceum*	*F*
		ANKA	chabaudi	strain	3D7		IT	8A	
2037	243	173	174	171	223		196	0	

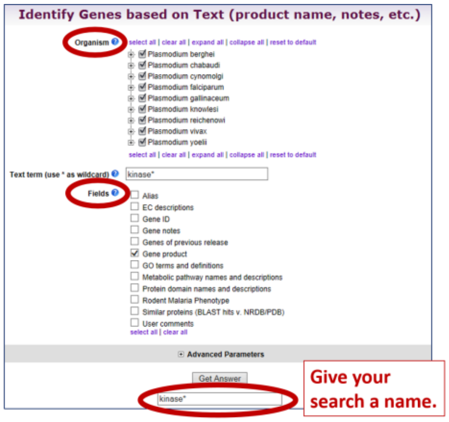
- What happens if you search using the term **kinases** in the Gene Text Search box? How many results are returned?

1b. 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 full text search form - **Identify Genes based on Text**, allows you to configure the search yourself, choosing parameters that best meet your needs. Use the search form to search for genes that have the word kinase in their **gene product** name/description.

- There are several ways to navigate to the **Identify Genes based on Text** page. 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.





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

Organism: select all | clear all | expand all | collapse all | reset to default
- ☑ Plasmodium berghei
- ☑ Plasmodium chabaudi
- ☑ Plasmodium cynomolgi
- ☑ Plasmodium falciparum
- ☑ Plasmodium gallinaceum
- ☑ Plasmodium knowlesi
- ☑ Plasmodium reichenowi
- ☑ Plasmodium vivax
- ☑ Plasmodium yoelii

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

Text term (use * as wildcard): kinase*

Fields:
- ☐ Alias
- ☐ EC descriptions
- ☐ Gene ID
- ☐ Gene notes
- ☐ Genes of previous release
- ☑ Gene product
- ☐ GO terms and definitions
- ☐ Metabolic pathway names and descriptions
- ☐ Protein domain names and descriptions
- ☐ Rodent Malaria Phenotype
- ☐ Similar proteins (BLAST hits v. NRDB/PDB)
- ☐ User comments

select all | clear all

Advanced Parameters

Get Answer

kinase*

Give your search a name.

- 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 and means any character) in your search term will broaden your search. Use the full text search, the specific page where you can define the fields to be searched (Fields = Gene Product).

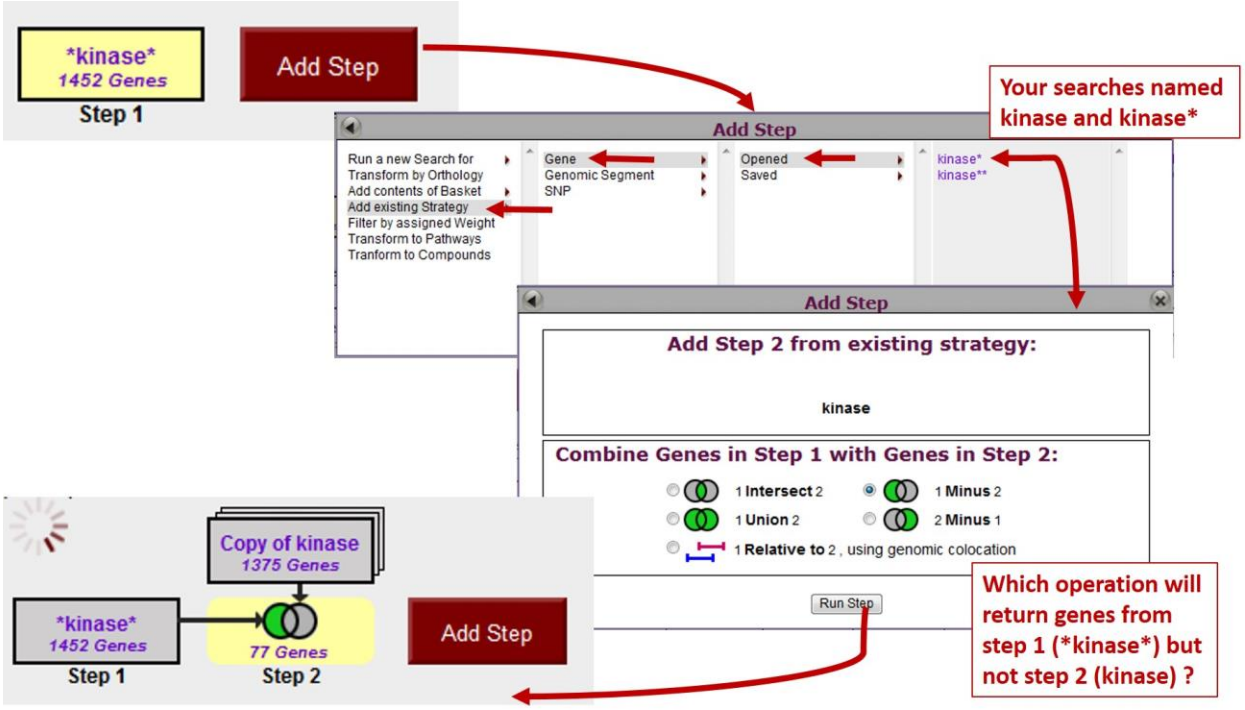
Try kinase *kinase *kinase*

- **Give each new search a name** to help you keep track of the searches.
- 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?

1 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 right 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.



Your searches named kinase and kinase*

Add Step 2 from existing strategy:

kinase

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

Which operation will return genes from step 1 (*kinase*) but not step 2 (kinase) ?

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

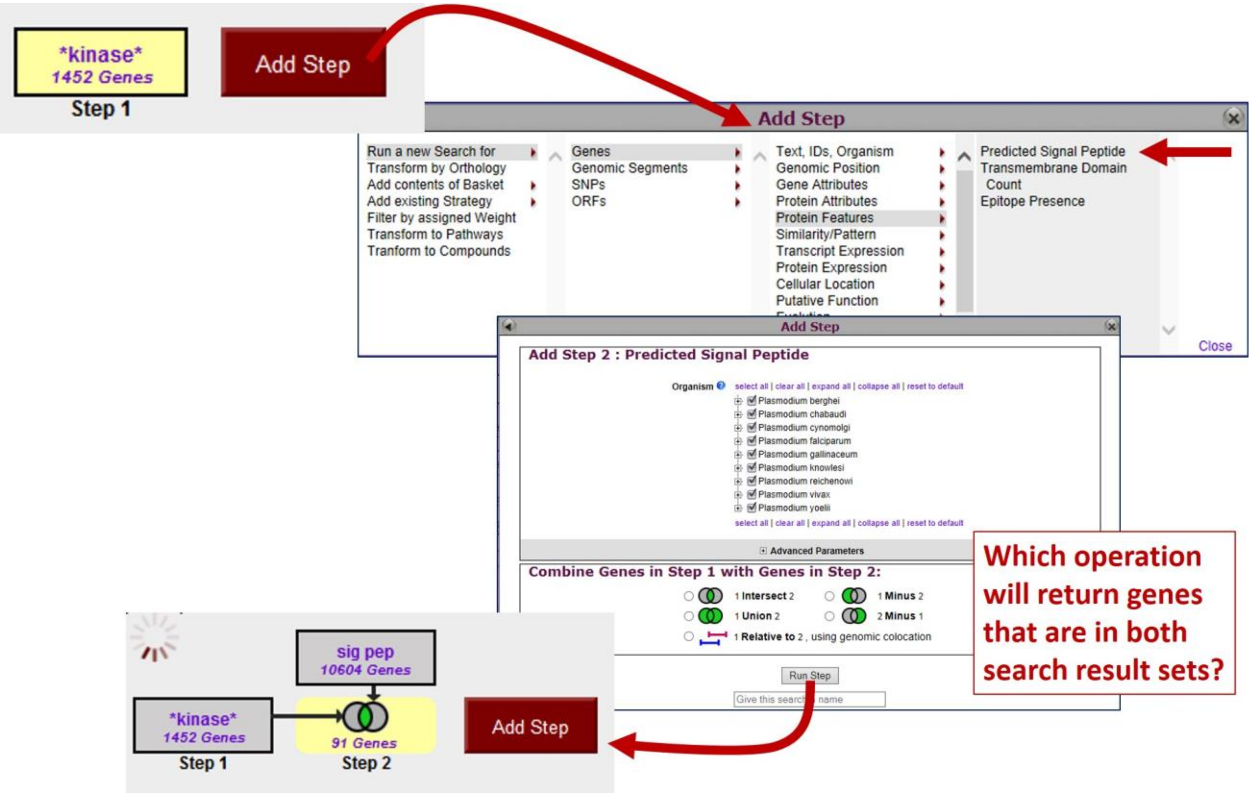
2. Combing text search results with results from other searches

2 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 Features, Predicted Signal Peptide**

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



Add Step 2 : Predicted Signal Peptide

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

- ☑ Plasmodium berghei
- ☑ Plasmodium chabaudi
- ☑ Plasmodium cynomolgi
- ☑ Plasmodium falciparum
- ☑ Plasmodium gallinaceum
- ☑ Plasmodium knowlesi
- ☑ Plasmodium reichenowi
- ☑ Plasmodium vivax
- ☑ Plasmodium yoelii

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

Give this search a name

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

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

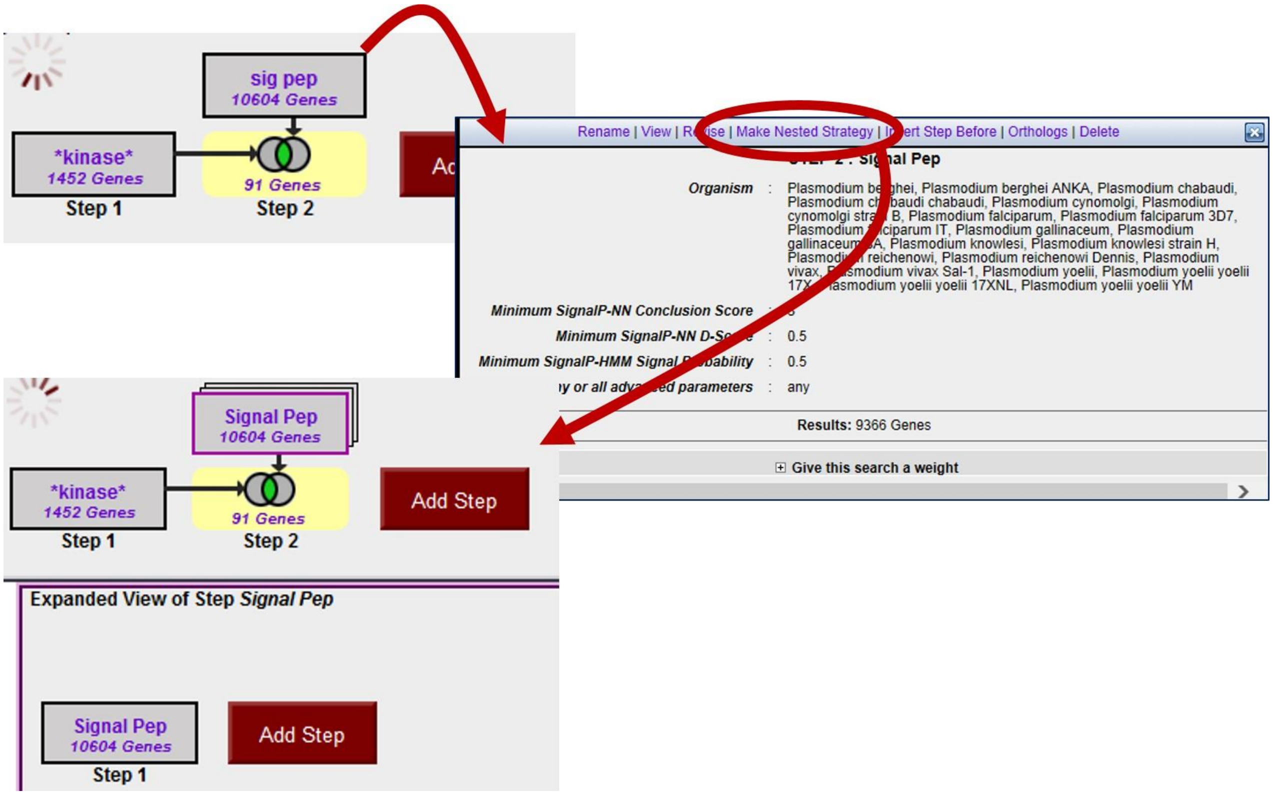
There is no wrong answer here!!

- 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?

2 c. In the above example, how can you define kinases that have either a secretory signal peptide, a transmembrane domain or both (a signal peptide and a transmembrane domain)?

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 x 3 + 5 = 11
Equation with nesting: 2 x (3 + 5) = 16



STEP 2 : Signal Pep

Organism : Plasmodium berghei, Plasmodium berghei ANKA, Plasmodium chabaudi, Plasmodium chabaudi chabaudi, Plasmodium cynomolgi, Plasmodium cynomolgi strain B, Plasmodium falciparum, Plasmodium falciparum 3D7, Plasmodium falciparum IT, Plasmodium gallinaceum, Plasmodium gallinaceum 8A, Plasmodium knowlesi, Plasmodium knowlesi strain H, Plasmodium reichenowi, Plasmodium reichenowi Dennis, Plasmodium vivax, Plasmodium vivax Sal-1, Plasmodium yoelii, Plasmodium yoelii yoelii 17X, Plasmodium yoelii yoelii 17XNL, Plasmodium yoelii yoelii YM

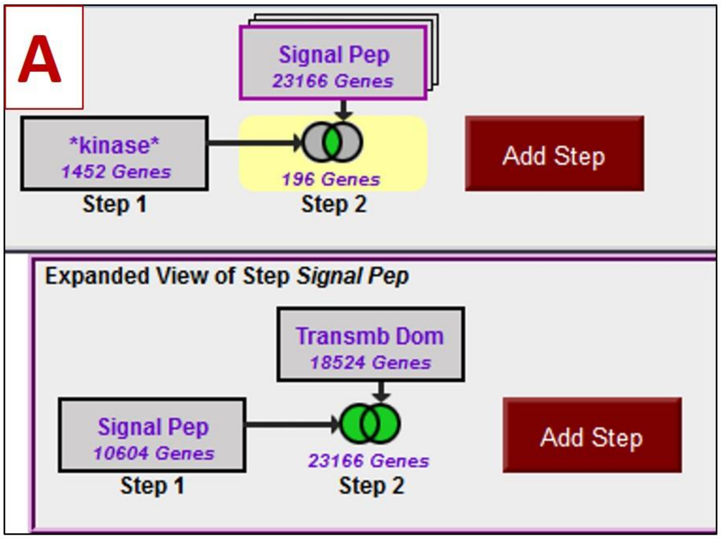
Minimum SignalP-NN Conclusion Score : 3

Minimum SignalP-NN D-Score : 0.5

Minimum SignalP-HMM Signal Probability : 0.5

Any or all advanced parameters : any

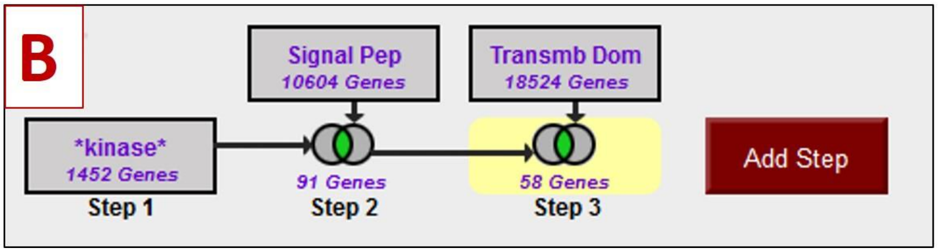
Results: 9366 Genes



Strategy Logic:

Strategy A returns kinases that have a signal peptide OR a TM domain OR both. (SP and/or TM) (either or both)

Strategy B returns kinases that have a signal peptide AND a TM domain



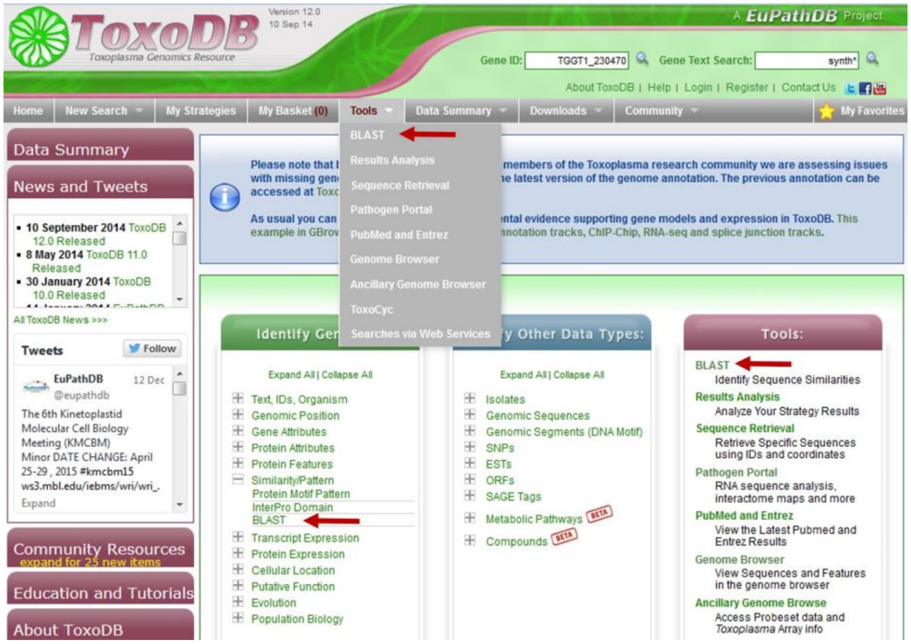
3. Finding a gene by BLAST Similarity.

Note: For this exercise start with http://www.toxodb.org

3 a. 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
  ttggattgccgtagcgttttatgagttgatagcttggctctaaaaaaacaaggctgaaaa
  atggaaaaaaatgtctccaat

- Sequence is also available from this URL:
  http://tinyurl.com/ex1blast
- Try using the BLAST search with this sequence (hint: you can get to the BLAST tool by clicking on the BLAST link under tools on the home page).



Please note that I … members of the Toxoplasma research community we are assessing issues with missing gen… e latest version of the genome annotation. The previous annotation can be accessed at Tox…

As usual you can … ntal evidence supporting gene models and expression in ToxoDB. This example in GBrov… nnotation tracks, ChIP-Chip, RNA-seq and splice junction tracks.

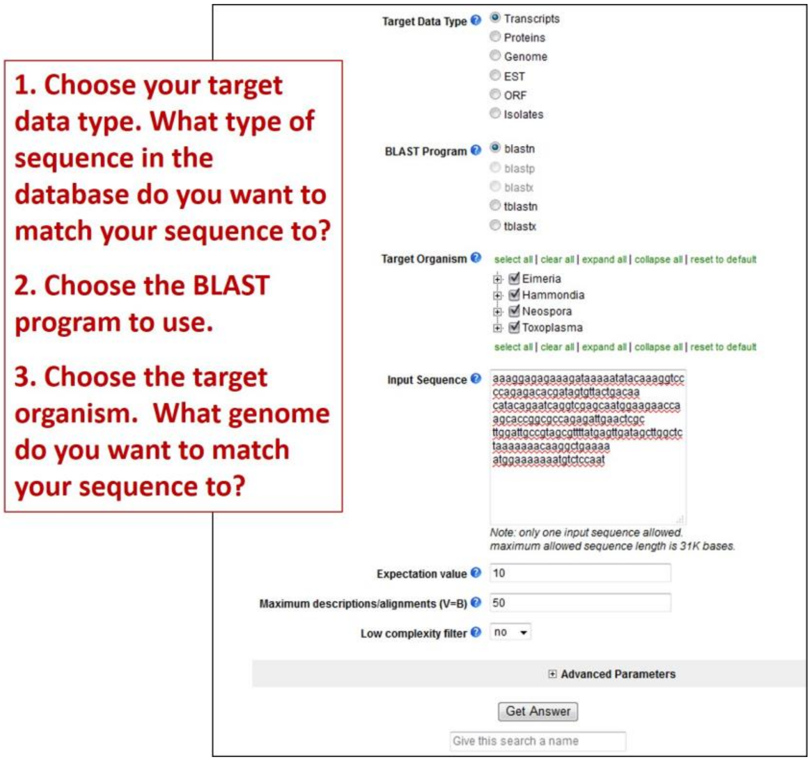
- 10 September 2014 ToxoDB 12.0 Released
- 8 May 2014 ToxoDB 11.0 Released
- 30 January 2014 ToxoDB 10.0 Released

EuPathDB @eupathdb 12 Dec
The 6th Kinetoplastid Molecular Cell Biology Meeting (KMCBM) Minor DATE CHANGE: April 25-29 , 2015 #kmcbm15 ws3.mbl.edu/iebms/wri/wri_.
Expand

- Which blast program should you use? (hint: try different combinations, just keep in mind that you have a nucleotide sequence so you have to 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.



1. Choose your target data type. What type of sequence in the database do you want to match your sequence to?

2. Choose the BLAST program to use.

3. Choose the target organism. What genome do you want to match your sequence to?

Target Data Type ● Transcripts ○ Proteins ○ Genome ○ EST ○ ORF ○ Isolates

BLAST Program ● blastn ○ blastp ○ blastx ○ tblastn ○ tblastx

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

☑ Eimeria
☑ Hammondia
☑ Neospora
☑ Toxoplasma

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

Input Sequence

```
aaaggagagaaagataaaaatatacaaaggtcc
ccagagacacgatagtgttactgacaa
catacagaatcaggtcgagcaatggaagaacca
agcaccggcgccagagattgaactcgc
ttggattgccgtagcgttttatgagttgatagcttggctc
taaaaaaacaaggctgaaaa
atggaaaaaaatgtctccaat
```

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

- 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).

4 More BLASTing in EuPathDB (OPTIONAL).

Note: for this exercise use http://www.eupathdb.org

4 a. Retrieve the DHFR-TS protein sequence and BLAST the sequence against all genomes.

- The first thing we will need to do is get a sequence to use for BLAST. Search for the keyword "dihydrofolate" (without quotations). (Hint: use the Gene Text Search on the upper right hand side of the EuPathDB home page).

- You should get multiple hits. Find the first one that is annotated as "**dihydrofolate reductase-thymidylate synthase**" (Look in the Product Description column).
- Once you find one, click on the gene ID and go to the gene page. It might be helpful to open the gene page in a new window or tab.
- Scroll down to the bottom of the page to the "Sequences" section.
- Copy the amino acid sequence and go back to EuPathDB (if you have not done so already, it might be helpful to open EuPathDB in a new window or tab).
- Go to the BLAST page from the EuPathDB home page. (hint: under Tools on the EuPathDB home page).
- Arrange the parameters and paste your sequence as input:
  - Target Data Type : Proteins
  - BLAST Program : blastp
  - Target Organism : Select All
  - Input Sequence : Your pasted sequence
  - Expectation value : 10
  - Maximum descriptions/alignments (V=B) : 100
  - Low complexity filter : no
- Click on "Get Answer".

4 b. Based on the results you should have identified excellent hits in almost all pathogens in EuPathDB but can you find good hits in *Giardia* or *Trichomonas*? *Let's try a different BLAST method:*

- Go back to the BLAST window. Change the target data type to Genome.
- Select the BLAST Program. Notice you cannot select BLASTP anymore. Try the other options. Notice how your input sequence type has to change when you select a different program. (Hint: TBLASTN is the one you need).
- Select all target organisms. Click on "Get Answer".

Note that the results are still missing a dihydrofolate from *Giardia* and *Trichomonas*. *Let's try a different BLAST method.*

- Go to your gene page window (in CryptoDB) and copy the **nucleotide coding sequence**.
- Go to the BLAST window (in EuPathDB) and paste the nucleotide sequence into the input window.
- Select the target data type (try different ones) and the BLAST program. Notice you can only select TBLASTX or BLASTN when your input sequence is nucleotide. (Hint: select TBLASTX).
- Select the target organisms. This time let's specifically only select *Giardia* and *Trichomonas*. Click on "Get Answer".

Getting frustrated? Not getting a hit for *Giardia* in this case is actually the correct answer! This organism does not have dihydrofolate reductase or thymidylate synthetase activity.

5. Viewing data on a gene page.
Note: For this exercise use http://plasmodb.org/

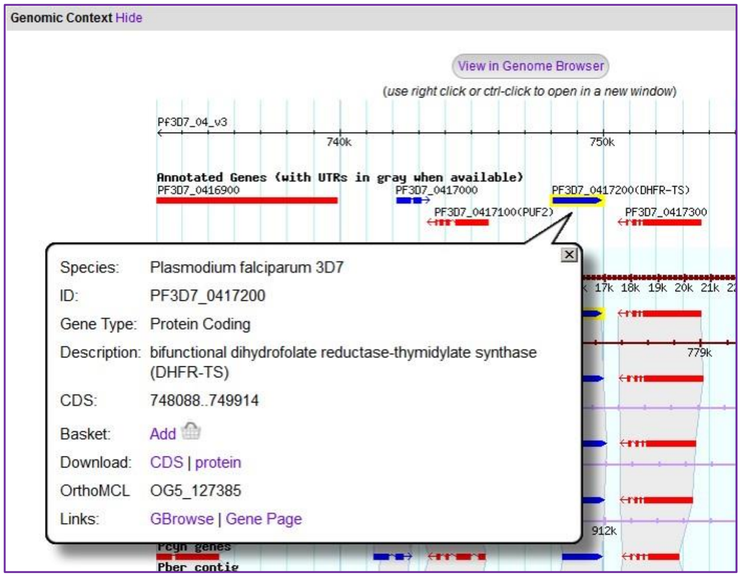
5 a. Find the gene page for one of the following *P. falciparum* genes and explore the information there to answer these questions.

1. bifunctional dihydrofolate reductase-thymidylate synthase (DHFR-TS, PF3D7_0417200)
2. apical membrane antigen 1 gene (AMA1, PF3D7_1133400)

- How did you navigate to this gene? What other ways could you get there? I can think of 4 ways to reach the gene page.

Look at the information on the gene page.

- What chromosome is this gene on?
- How many exons does this gene have? Hint: look at the graphic in the Genomic Context data track and mouse over the glyph representing the gene.
- What direction is the gene relative to the chromosome?
- How many nucleotides of coding sequence?
- Do you see a way to quickly download the coding and protein sequences?
- Does this gene have a user comment?

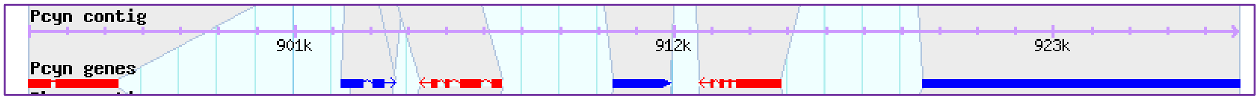


Species: Plasmodium falciparum 3D7
ID: PF3D7_0417200
Gene Type: Protein Coding
Description: bifunctional dihydrofolate reductase-thymidylate synthase (DHFR-TS)
CDS: 748088..749914
Basket: Add
Download: CDS | protein
OrthoMCL OG5_127385
Links: GBrowse | Gene Page

5 b. What genes are located upstream & downstream of DHFR-TS (AMA1) in P. falciparum?

- Is synteny (chromosome organization) in this region maintained in other species? Hint: look in the genomic context section of the gene page – what does the shading mean?

- How complete is the genome assembly for other species? Each genome is displayed as two tracks – the genomic sequence (chromosome or contig) on top and the gene models underneath. Do the contigs contain gaps or truncations?

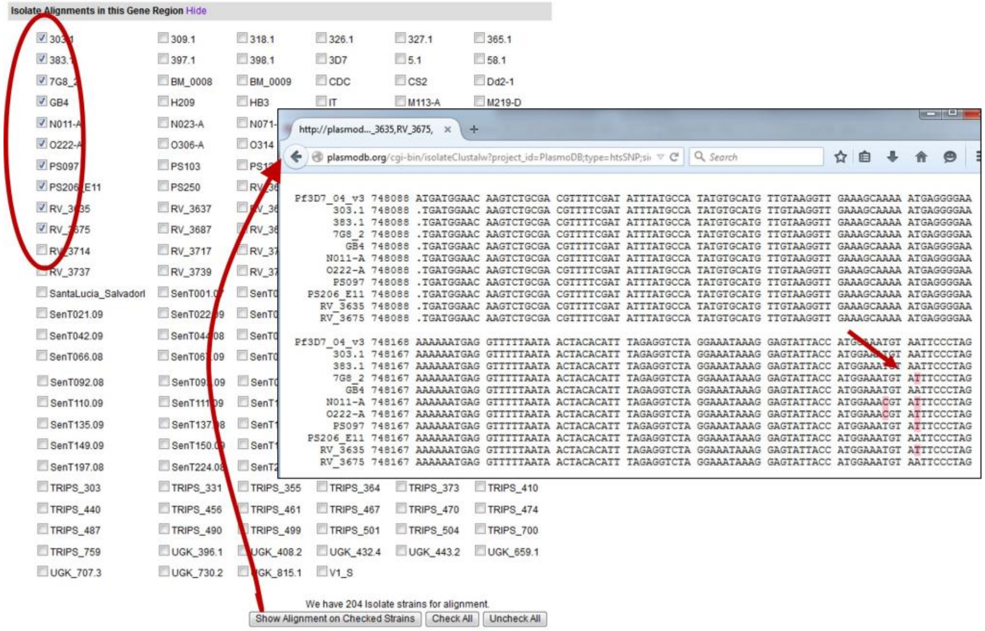


- What does synteny look like across the entire chromosome? To do this:
  - Click on the **"View in GBrowse"** button in the genomic context section.
  - Zoom out to the entire chromosome. There are a few ways to do this. For example, drag your cursor across the entire chromosome in the Overview panel and then select "zoom" from the popup menu.
  - Click on the tab called "Select tracks". Select the track called "Syntenic Sequences and Genes (Shaded by Orthology)". Go back to the Browser tab (this may take a minute to load).
  - 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?

5 c. Does the P. falciparum DHFR-TS (or AMA1) gene contain Single Nucleotide Polymorphisms (SNPs)?

SNPs are represented in a table called "SNP Overview" and using the "Isolate Alignments in this Gene Region" track you can view an alignment showing SNPs between specific strains/isolates.

- Examine the SNP Overview table.
- What is the total number of SNPs in the gene?
- How many impact the predicted protein sequence?
- Is this likely to define the full spectrum of sequence variation in these particular strains?
- Compare the SNP characteristics of this gene to upstream and downstream genes. How do these results compare with SNP distribution in other genes?
- Open the Isolate Alignments in this Gene Region data track and run an alignment between several isolates: 303.1, 383.1, 7G8_2, GB4, N011-A, O222-A, PS097, PS206_E11, RV_3635, RV_3675



Isolate Alignments in this Gene Region Hide

☑ 303.1 □ 309.1 □ 318.1 □ 326.1 □ 327.1 □ 365.1
☑ 383.1 □ 397.1 □ 398.1 □ 3D7 □ 5.1 □ 58.1
☑ 7G8_2 □ BM_0008 □ BM_0009 □ CDC □ CS2 □ Dd2-1
☑ GB4 □ H209 □ HB3 □ IT □ M113-A □ M219-D
☑ N011-A □ N023-A □ N071-
☑ O222-A □ O306-A □ O314
☑ PS097 □ PS103 □ PS1
☑ PS206_E11 □ PS250 □ RV_36
☑ RV_3635 □ RV_3637 □ RV_36
☑ RV_3675 □ RV_3687 □ RV_36
□ RV_3714 □ RV_3717 □ RV_37
□ RV_3737 □ RV_3739 □ RV_37
□ SantaLucia_Salvador I □ SenT001.07 □ SenT0
□ SenT021.09 □ SenT022.09 □ SenT0
□ SenT042.09 □ SenT044.08 □ SenT0
□ SenT066.08 □ SenT067.09 □ SenT0
□ SenT092.08 □ SenT093.09 □ SenT0
□ SenT110.09 □ SenT111.09 □ SenT1
□ SenT135.09 □ SenT137.08 □ SenT1
□ SenT149.09 □ SenT150.09 □ SenT1
□ SenT197.08 □ SenT224.08 □ SenT2
□ TRIPS_303 □ TRIPS_331 □ TRIPS_355 □ TRIPS_364 □ TRIPS_373 □ TRIPS_410
□ TRIPS_440 □ TRIPS_456 □ TRIPS_461 □ TRIPS_467 □ TRIPS_470 □ TRIPS_474
□ TRIPS_487 □ TRIPS_490 □ TRIPS_499 □ TRIPS_501 □ TRIPS_504 □ TRIPS_700
□ TRIPS_759 □ UGK_396.1 □ UGK_408.2 □ UGK_432.4 □ UGK_443.2 □ UGK_659.1
□ UGK_707.3 □ UGK_730.2 □ UGK_815.1 □ V1_S

We have 204 Isolate strains for alignment.

Show Alignment on Checked Strains
Check All
Uncheck All
http://plasmod..._3635,RV_3675,
plasmodb.org/cgi-bin/isolateClustalw?project_id=PlasmoDB;type=htsSNP;si
Search

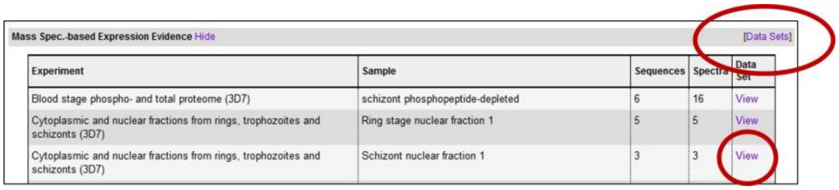
```
Pf3D7_04_v3 748088 ATGATGGAAC AAGTCTGCGA CGTTTTCGAT ATTTATGCCA TATGTGCATG TTGTAAGGTT GAAAGCAAAA ATGAGGGGAA
      303.1 748088 .TGATGGAAC AAGTCTGCGA CGTTTTCGAT ATTTATGCCA TATGTGCATG TTGTAAGGTT GAAAGCAAAA ATGAGGGGAA
      383.1 748088 .TGATGGAAC AAGTCTGCGA CGTTTTCGAT ATTTATGCCA TATGTGCATG TTGTAAGGTT GAAAGCAAAA ATGAGGGGAA
      7G8_2 748088 .TGATGGAAC AAGTCTGCGA CGTTTTCGAT ATTTATGCCA TATGTGCATG TTGTAAGGTT GAAAGCAAAA ATGAGGGGAA
        GB4 748088 .TGATGGAAC AAGTCTGCGA CGTTTTCGAT ATTTATGCCA TATGTGCATG TTGTAAGGTT GAAAGCAAAA ATGAGGGGAA
    N011-A 748088 .TGATGGAAC AAGTCTGCGA CGTTTTCGAT ATTTATGCCA TATGTGCATG TTGTAAGGTT GAAAGCAAAA ATGAGGGGAA
    O222-A 748088 .TGATGGAAC AAGTCTGCGA CGTTTTCGAT ATTTATGCCA TATGTGCATG TTGTAAGGTT GAAAGCAAAA ATGAGGGGAA
      PS097 748088 .TGATGGAAC AAGTCTGCGA CGTTTTCGAT ATTTATGCCA TATGTGCATG TTGTAAGGTT GAAAGCAAAA ATGAGGGGAA
  PS206_E11 748088 .TGATGGAAC AAGTCTGCGA CGTTTTCGAT ATTTATGCCA TATGTGCATG TTGTAAGGTT GAAAGCAAAA ATGAGGGGAA
    RV_3635 748088 .TGATGGAAC AAGTCTGCGA CGTTTTCGAT ATTTATGCCA TATGTGCATG TTGTAAGGTT GAAAGCAAAA ATGAGGGGAA
    RV_3675 748088 .TGATGGAAC AAGTCTGCGA CGTTTTCGAT ATTTATGCCA TATGTGCATG TTGTAAGGTT GAAAGCAAAA ATGAGGGGAA

Pf3D7_04_v3 748168 AAAAAATGAG GTTTTTAATA ACTACACATT TAGAGGTCTA GGAAATAAAG GAGTATTACC ATGGAAATGT AATTCCCTAG
      303.1 748167 AAAAAATGAG GTTTTTAATA ACTACACATT TAGAGGTCTA GGAAATAAAG GAGTATTACC ATGGAAATGT AATTCCCTAG
      383.1 748167 AAAAAATGAG GTTTTTAATA ACTACACATT TAGAGGTCTA GGAAATAAAG GAGTATTACC ATGGAAATGT AATTCCCTAG
      7G8_2 748167 AAAAAATGAG GTTTTTAATA ACTACACATT TAGAGGTCTA GGAAATAAAG GAGTATTACC ATGGAAATGT ATTTCCCTAG
        GB4 748167 AAAAAATGAG GTTTTTAATA ACTACACATT TAGAGGTCTA GGAAATAAAG GAGTATTACC ATGGAAATGT AATTCCCTAG
    N011-A 748167 AAAAAATGAG GTTTTTAATA ACTACACATT TAGAGGTCTA GGAAATAAAG GAGTATTACC ATGGAAACGT ATTTCCCTAG
    O222-A 748167 AAAAAATGAG GTTTTTAATA ACTACACATT TAGAGGTCTA GGAAATAAAG GAGTATTACC ATGGAAACGT ATTTCCCTAG
      PS097 748167 AAAAAATGAG GTTTTTAATA ACTACACATT TAGAGGTCTA GGAAATAAAG GAGTATTACC ATGGAAATGT ATTTCCCTAG
  PS206_E11 748167 AAAAAATGAG GTTTTTAATA ACTACACATT TAGAGGTCTA GGAAATAAAG GAGTATTACC ATGGAAATGT AATTCCCTAG
    RV_3635 748167 AAAAAATGAG GTTTTTAATA ACTACACATT TAGAGGTCTA GGAAATAAAG GAGTATTACC ATGGAAATGT ATTTCCCTAG
    RV_3675 748167 AAAAAATGAG GTTTTTAATA ACTACACATT TAGAGGTCTA GGAAATAAAG GAGTATTACC ATGGAAATGT AATTCCCTAG
```

5 d. Is the DHFR-TS (or AMA1) gene expressed?

Look at the gene page sections entitled "Protein" and "Expression". You may have to click on the **show** link to reveal the data associated with that data track.

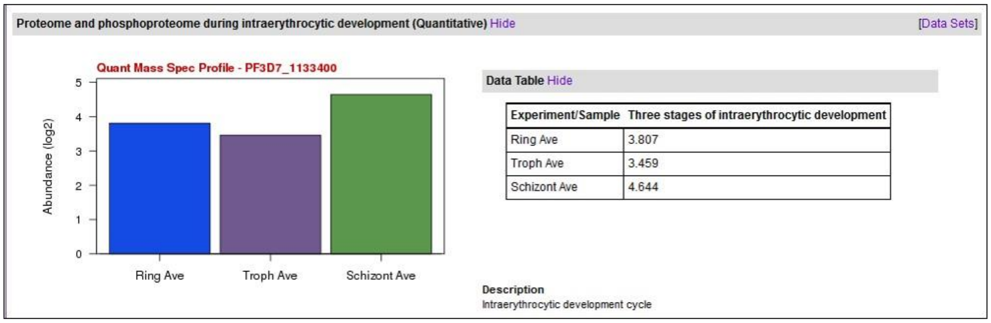
- What kinds of data in PlasmoDB provide evidence for expression? Hint: open the Protein Features graphic which is the first data track in the Protein section.
- Is this gene expressed at the protein level in salivary gland sporozoites? – in the blood stage phosphoproteome? Look at the Protein context graphic and the table of Mass Spec.-based Expression Evidence.
- Can you quickly link to the data set record for proteomics experiments?



Mass Spec.-based Expression Evidence Hide [Data Sets]

Experiment	Sample	Sequences	Spectra	Data Set
Blood stage phospho- and total proteome (3D7)	schizont phosphopeptide-depleted	6	16	View
Cytoplasmic and nuclear fractions from rings, trophozoites and schizonts (3D7)	Ring stage nuclear fraction 1	5	5	View
Cytoplasmic and nuclear fractions from rings, trophozoites and schizonts (3D7)	Schizont nuclear fraction 1	3	3	View

- How abundant is DHFR-TS (AMA1) protein? How confident are you of this analysis? Abundance can be estimated by counting the number of spectra supporting a peptide spectra that maps to the protein. Where do you find information about the number of spectra?
- Is the protein more abundant in the ring or schizont life cycle stage? Hint: open the quantitative proteomics track called **Proteome and phosphoproteome during intraerythrocytic development (Quantitative)**.



Proteome and phosphoproteome during intraerythrocytic development (Quantitative)

Data Table

Experiment/Sample	Three stages of intraerythrocytic development
Ring Ave	3.807
Troph Ave	3.459
Schizont Ave	4.644

Description
Intraerythrocytic development cycle

- Look at the Expression data track labeled **Life cycle expression data (3D7)**. Based on this data, at what life cycle stage is DHFR-TS (AMA1) most abundant? Does this make sense?
- Do the life cycle microarray expression profiles from different data tracks (and thus different experiments/data sets) give the same results? What tracks did you use?
- What about RNA-sequence data, does it agree with microarray data? See these two data tracks – **Strand specific transcriptomes of 4 life cycle stages**; **Transcriptomes of 7 sexual and asexual life stages**.